ADVANCES IN INTELLIGENT AND SOFT COMPUTING 125

Tianbiao Zhang (Ed.)

Mechanical Engineering and Technology

Springer

Advances in Intelligent and Soft Computing 125

Editor-in-Chief: J. Kacprzyk

Advances in Intelligent and Soft Computing

Editor-in-Chief

Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Further volumes of this series can be found on our homepage: springer.com

Vol. 111. L. Jiang (Ed.)
Proceedings of the 2011 International Conference on Informatics, Cybernetics, and Computer Engineering (ICCE 2011) November 19-20, 2011, Melbourne, Australia, 2011
ISBN 978-3-642-25187-0

Vol. 112. L. Jiang (Ed.)
Proceedings of the 2011 International Conference on Informatics, Cybernetics, and Computer Engineering (ICCE 2011) November 19-20, 2011, Melbourne, Australia, 2011
ISBN 978-3-642-25193-1

Vol. 113. J. Altmann, U. Baumöl, and B.J. Krämer (Eds.)
Advances in Collective Intelligence 2011, 2011
ISBN 978-3-642-25320-1

Vol. 114. Y. Wu (Ed.)
Software Engineering and Knowledge Engineering: Theory and Practice, 2011
ISBN 978-3-642-03717-7

Vol. 115. Y. Wu (Ed.)
Software Engineering and Knowledge Engineering: Theory and Practice, 2011
ISBN 978-3-642-03717-7

Vol. 116. Yanwen Wu (Ed.)
Advanced Technology in Teaching - Proceedings of the 2009 3rd International Conference on Teaching and Computational Science (WTCS 2009), 2012
ISBN 978-3-642-11275-1

Vol. 117. Yanwen Wu (Ed.)
Advanced Technology in Teaching - Proceedings of the 2009 3rd International Conference on Teaching and Computational Science (WTCS 2009), 2012
ISBN 978-3-642-25436-9

Vol. 118. A. Kapczynski, E. Tkacz, and M. Rostanski (Eds.)
Internet - Technical Developments and Applications 2, 2011
ISBN 978-3-642-25354-6

Vol. 119. Tianbiao Zhang (Ed.)
Future Computer, Communication, Control and Automation, 2011
ISBN 978-3-642-25537-3

Vol. 120. Nicolas Loménie, Daniel Racoceanu, and Alexandre Gouaillard (Eds.)
Advances in Bio-Imaging: From Physics to Signal Understanding Issues, 2011
ISBN 978-3-642-25546-5

Vol. 121. Tomasz Traczyk and Mariusz Kaleta (Eds.)
Modeling Multi-commodity Trade: Information Exchange Methods, 2011
ISBN 978-3-642-25648-6

Vol. 122. Yinglin Wang and Tianrui Li (Eds.)
Foundations of Intelligent Systems, 2011
ISBN 978-3-642-25663-9

Vol. 123. Yinglin Wang and Tianrui Li (Eds.)
Knowledge Engineering and Management, 2011
ISBN 978-3-642-25660-8

Vol. 124. Yinglin Wang and Tianrui Li (Eds.)
Practical Applications of Intelligent Systems, 2011
ISBN 978-3-642-25657-8

Vol. 125. Tianbiao Zhang (Ed.)
Mechanical Engineering and Technology, 2012
ISBN 978-3-642-27328-5

Tianbiao Zhang (Ed.)

Mechanical Engineering and Technology

Selected and Revised Results of the 2011 International Conference on Mechanical Engineering and Technology, London, UK, November 24–25, 2011

Springer

Editor

Dr. Tianbiao Zhang
Central China Normal University
Lvting Yajing, Hongshan Qu
10-3-102
430079 Wuhan
China, People's Republic
E-mail: ztb@mails.ccnu.edu.cn

ISBN 978-3-642-27328-5 e-ISBN 978-3-642-27329-2

DOI 10.1007/978-3-642-27329-2

Advances in Intelligent and Soft Computing ISSN 1867-5662

Library of Congress Control Number: 2011943327

Typeset by Scientific Publishing Services Pvt. Ltd., Chennai, India

Printed on acid-free paper

5 4 3 2 1 0

springer.com

ICMET 2011 Preface

2011 International Conference on Mechanical Engineering and Technology (ICMET 2011) was held on London, UK, November 24--25, 2011. ICMET 2011 is a platform for researchers, engineers, academicians as well as industrial professionals from all over the world to present their research results and development activities in Mechanical Engineering and Technology, and related interdisciplinary subjects.

Mechanical engineering is a discipline of engineering that applies the principles of physics and materials science for analysis, design, manufacturing, and maintenance of mechanical systems. It is the branch of engineering that involves the production and usage of heat and mechanical power for the design, production, and operation of machines and tools. It is one of the oldest and broadest engineering disciplines.

The engineering field requires an understanding of core concepts including mechanics, kinematics, thermodynamics, materials science, and structural analysis. Mechanical engineers use these core principles along with tools like computer-aided engineering and product lifecycle management to design and analyze manufacturing plants, industrial equipment and machinery, heating and cooling systems, transport systems, aircraft, watercraft, robotics, medical devices and more.

Mechanical engineering technology is the application of physical principles and current technological developments to the creation of useful machinery and operation design. Technologies such as solid models may be used as the basis for finite element analysis (FEA) and / or computational fluid dynamics (CFD) of the design. Through the application of computer-aided manufacturing (CAM), the models may also be used directly by software to create "instructions" for the manufacture of objects represented by the models, through computer numerically controlled (CNC) machining or other automated processes, without the need for intermediate drawings.

Mechanical engineering technologists are also expected to understand and be able to apply concepts from the chemical and electrical engineering fields. Mechanical engineering technologists are expected to apply current technologies and principals to machine and product design, production, and manufacturing processes.

Persons who attended the conference were engineers, scientists, managers of various companies and professors of the universities from about 24 different countries and regions. This proceeding covers the subject areas of Mechanical Engineering and Technology, and also covers interdisciplinary subject areas of Computer Science and Technology. The subject areas covers: Mechanical Design and Manufacture & Automation (including Mechanical Dynamics and Its Application, Mechanical Transmission Theory and Application Technology, Mechanical Reliability Theory and Engineering, Detection Technique of Mechanical System Fault Diagnosis and Troubleshooting of Mechanical System, Theory and Application of Friction and Wear, Vibration, Noise Analysis and Control, Dynamic Analysis, etc), Advanced

Materials and Advanced Manufacturing Technology (including Surface Engineering/Coating, Modeling, Analysis, and Simulation of Manufacturing Process, Materials Forming, Materials Machining , Welding & Joining , Laser Processing Technology, Casting and Solidification, Waste Engineering and Management, etc), Electronic & Mechanical Engineering (including Mechanical Engineering Testing , Mechanical System Engineering , Hydraulic Transmission and Control, Application of Hydraulic Technology , Fault Diagnosis and Troubleshooting of Hydraulic System , Design and Use of Hydraulic Servo System , Electromechanical Transmission Control , Robot, etc), Computer Software Engineering and Information System Design (including Software Architectures, Software Design and Development, Software Testing, Software Agents, Web-based Software Engineering, Project Management, Software Performance Engineering, Service Engineering, etc), Computer Simulation and Modeling (including Simulation Tools and Languages, Discrete Event Simulation, Object-Oriented Implementation, Web-based Simulation, Monte Carlo Simulation, Distributed Simulation, Simulation Optimization, Numerical Methods, Mathematical Modeling, etc), Computer Automation Control (including Micro-computer Embedded Control Applications, Process Control and Automation, Sensors and Applications, Industrial Process Control, Decision Support Systems, Fuzzy Control and Its Applications, etc), Artificial Intelligence (including Machine Learning, Pattern Recognition, Knowledge Discovery, Intelligent Data Analysis, Neural Networks, Genetic Algorithms, Medical Diagnostics, Data Mining, etc), Computer Science(including Numerical Algorithms and Analysis, Computational Simulation and Analysis, Data Visualization and Virtual Reality, Computational Mathematics, Computational Graphics, Computational Statistics, etc). We hope that colleagues can benefit from the communication among the interdisciplinary subjects achievements.

We would like to thank the organizing committee for their efforts in helping us compile this volume. Special thanks go to Springer Publisher.

We hope that ICMET 2011 will be successful and enjoyable to all participants. We look forward to seeing all of you next year at the ICMET 2012.

Tianbiao Zhang
Publication Chair

Contents

Robot Enhanced Inspection and Restoration System

Yang Wang, Jin Huang, Xiang Wu, Ying Hu, and Zi Ma

Automation Research Center,
Dalian Maritime University, Ddlian,
Liaoning, P.R. China, 116026
xiangw.mail@gmail.com

Abstract. As to achieve the goal of automation remanufacture, a six degree-of-freedom (DOF) robot and a three dimensional (3D) laser scanning system are combined to construct a novel remanufacture system(RS). In the presented RS, the restoring path is generated automatically. It follows that the work efficiency and restore accuracy of RS is improved. Some key techniques such as the system calibration and path planning are discussed. The effectiveness of the presented approach is validated via some real experiments.

Keywords: Remanufacture, robotic inspection system, workpiece restoration.

1 Introduction

It is well known that the damaged region information and robot restoration path planning are two main factors in remanufacture system (RS). Structured light scanner is used to sample data, and cooperate with an additional movement mechanism to reconstruct the three dimensional (3D) surface of the scanned object [1][2] in the system. This design brings in the six degree-of-freedom (DOF) robot to make it more appropriate to remanufacture object. Once the scanning data was obtained, the robot path could be generated according to welding technology and robot work range.

2 System Calibration

The whole structure of RS is shown in Figure 1. The system consists of two robots and tow executors. The left robot which is in charge of the inspection task is named as measuring robot, the right one which is in charge of the restoration task is named as restoration robot. There are three groups of the device parameters need to be determined before RS works. One is the relationship between the two robots; the others are the parameters of the two key executors, namely the structured light vision sensor and the welding torch. In this section, a method for calibrating the welding torch and the two robots simultaneously is proposed. After that, a robust method to calibrate the light plane of the structured light vision sensor is given.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 1–8.
springerlink.com

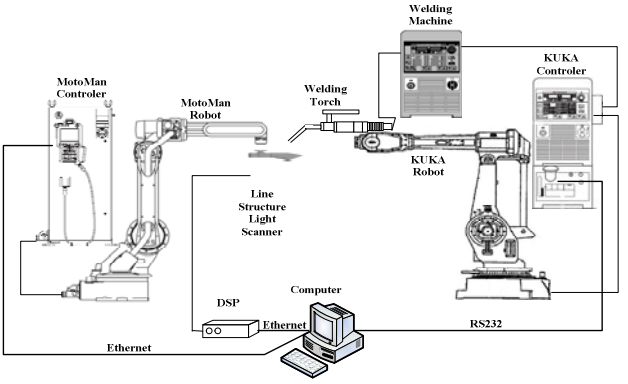


Fig. 1. The structure of the system

2.1 Calibration of Welding Torch and Robots' Relationship

A stereoscopic nummular target shown in Figure 2 is designed to calibrate the welding torch and the relationships of the two robots.

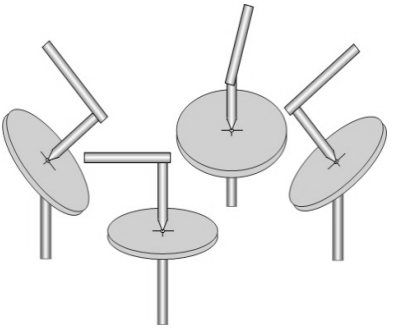


Fig. 2. The stereoscopic nummular Target

Some stereoscopic nummular targets, more than four, are placed in the workspace in different posture. The centers of the targets measured by the measuring robot are marked as $p_i(p_{ix}, p_{iy}, p_{iz})$. Then make the end of the welding torch moving to the centers as close as possible, so the coordinates of the centers in the restoration robot coordinate system are express as $q_i(q_{ix}, q_{iy}, q_{iz})$. The rotation matrix of the flange coordinate system to the base coordinate system of the restoration robot is ${}^{0}_{6}R_i$.

Assume the rotation matrix of the welding torch coordinate system to the flange coordinate system ${}^{6}_{T}R$ and the translation matrix ${}^{6}_{T}t$, and then we can get the equation

$$q_i = {}^0_6R_i\,{}^6_Tt + {}^0_6t_i \tag{1}$$

If there are four targets, the simultaneous equations could be written as

$$\begin{cases} q_2 - q_1 = ({}^0_6R_2 - {}^0_6R_1)\,{}^6_Tt + ({}^0_6t_2 - {}^0_6t_1) \\ q_3 - q_1 = ({}^0_6R_3 - {}^0_6R_1)\,{}^6_Tt + ({}^0_6t_3 - {}^0_6t_1) \\ q_4 - q_1 = ({}^0_6R_4 - {}^0_6R_1)\,{}^6_Tt + ({}^0_6t_4 - {}^0_6t_1) \\ q_3 - q_2 = ({}^0_6R_3 - {}^0_6R_2)\,{}^6_Tt + ({}^0_6t_3 - {}^0_6t_2) \\ q_4 - q_2 = ({}^0_6R_4 - {}^0_6R_2)\,{}^6_Tt + ({}^0_6t_4 - {}^0_6t_2) \\ q_4 - q_3 = ({}^0_6R_4 - {}^0_6R_3)\,{}^6_Tt + ({}^0_6t_4 - {}^0_6t_{31}) \end{cases} \tag{2}$$

when the vector q_{mn} represents $q_m - q_n$, matrix R_{mn} represents the rotation matrix ${}^0_6R_m - {}^0_6R_n$, and t_{mn} represents ${}^0_6t_m - {}^0_6t_n$, the equation(2) is expressed as

$$\begin{cases} q_{21} = R_{21}\,{}^6_Tt + t_{21} \\ q_{31} = R_{31}\,{}^6_Tt + t_{31} \\ q_{41} = R_{41}\,{}^6_Tt + t_{41} \\ q_{32} = R_{32}\,{}^6_Tt + t_{32} \\ q_{42} = R_{42}\,{}^6_Tt + t_{42} \\ q_{43} = R_{43}\,{}^6_Tt + t_{43} \end{cases} \tag{3}$$

Using p_{mn} represents the vector $p_m - p_n$, the modulus of the vector q_{mn} and p_{mn} are equal although they are in different coordinate system:

$$\|p_{mn}\| = \|q_{mn}\| \tag{4}$$

If ${}^6_Tt = \begin{bmatrix} t_x & t_y & t_z \end{bmatrix}^T$, we can get the equation as below:

$$\left\| R_{mn} \begin{bmatrix} t_x & t_y & t_z \end{bmatrix}^T + t_{mn} \right\| - \|q_{mn}\| = 0 \tag{5}$$

where (t_x, t_y, t_z) is the unknown that can be calculated by the nonlinear optimization techniques. So when $p_i(p_{ix}, p_{iy}, p_{iz})$ and $q_i(q_{ix}, q_{iy}, q_{iz})$ are obtained, the relationship of the two robots can be calculated finally.

2.2 Calibration of Structured Light Vision Sensor

The principle model of the structured light vision sensor is shown in Figure 3. The scanning laser line reflects the surface profile lies on the light plane, and only the points on the laser line are abstracted from the CCD. When the instinct parameter K of the camera in the sensor is calibrated [3], the coordinate of the point P in the normalized plane can be calculated via the equation

$$\begin{aligned} P'' &= K \cdot P' \\ z_c \cdot P' &= [I \;\; 0] \cdot P \end{aligned} \tag{6}$$

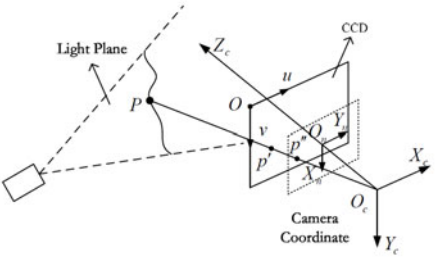


Fig. 3. The principle model of the structured light sensor

The light plane can be calibrated by a planar target. As shown in the Figure4, there are three collinear points A, B, C on the known target line l. The distance between A and B and the distance between B and C is known. M is the intersection of the line l and the laser plane, . a, b, c, m is the imaging point of A, B, C, M on the normalized plane that could be calculated by the equation 6. Based on the geometric projection principle, l , l' and the optic center O_c are on the same plane.

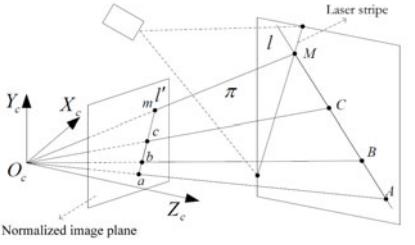


Fig. 4. Calibration principal for light plane

Assume the angles of the O_cA , O_cB and O_cC are θ_1 θ_2 ,according to the definition of the dot product of the vectors

$$\begin{cases} A = \|O_cA\| \cdot \dfrac{\tilde{a}}{\|\tilde{a}\|} \\ B = \|O_cB\| \cdot \dfrac{\tilde{b}}{\|\tilde{b}\|} \end{cases} \tag{7}$$

$\tilde{a}, \tilde{b}, \tilde{c}$ represent the homogeneous coordinates on the normalized plane, when the cross ratio γ is calculated by:

$$\begin{aligned} \gamma &= \frac{\sin\angle aO_cc}{\sin\angle bO_cc} : \frac{\sin\angle aO_cm}{\sin\angle bO_cm} \\ &= \frac{A-C}{B-C} : \frac{A-M}{B-M} \end{aligned} \tag{8}$$

The coordinate of point M is

$$M_i = \frac{\gamma \cdot (B_i - C_i) \cdot A_i - (A_i - C_i) \cdot B_i}{\gamma \cdot (B_i - C_i) - (A_i - C_i)} \quad i = x, y, z \tag{9}$$

So the 3D points on the light plane which are used for fitting the light plane can be obtained by moving the planar target arbitrarily. The objective function is the quadratic sum of the distance between the point and the plane.

$$\min f(a_1, a_2, a_3, a_4) = \min \sum_{i=1}^{n} D_i^2 \tag{10}$$

where

$$D_i = \frac{\left| a_1 x_i + a_2 y_i + a_3 z_i + a_4 \right|}{\sqrt{a_1^2 + a_2^2 + a_3^2}} \tag{11}$$

By using the LM algorithm [4], the parameter of the light plane can be estimated.

3 Data Processing

The substantive scan points obtained by the measuring robot form a point cloud data that represents the inspected workpiece. It contains both the surface and breakage information. The scanning data and the CAD model are composited to generate path for the restoration robot. Therefore, a scenario is proposed as follows.

3.1 Allocate the Damaged Region

When the scanning data of the workpiece was obtained, the information of the damaged region could be obtained. Firstly, register the CAD model to the scanning data by the ICP algorithm [7]. The key process of the algorithm is to calculate the transformation matrix by two kinds of points, one of them called source point which extracted from the scanning data and the other called corresponding point which generated by the target approach. Since the CAD model in this application is represented as STL file format that includes lots of triangles, so we could get the distance between source point and each triangles then take the one with the shortest distance as the corresponding point. Since the broken region reduces the effective of the ICP algorithm for registering the scan data with the CAD data, the points with an error that higher than a threshold need to be ignored.

The steps are as follows

- get the distance of a point to a plane expressed by the triangle.
- determine whether the point in the triangle, if so, go to step 4.ohterwise, go to step3
- calculate the distance between the point to the three sides of the triangle. That is to get a distance of a point to a line. If the intersaction out of the side of the triangle, get the distance of the two vertex.

- remove the points with the distance represents the error value higher than a threshold
- traverse all the triangles and take the one with the shortest distance as the corresponding point.

Once the scanning data was registered to the CAD model the damaged region can be easily allocated and generated by comparison with a threshold. Generally, the threshold is determined by the deposition rate that is the volume sprayed by the welding torch per second.

3.2 Path Planning for the Restoration Robot

The last procedure is to generate the path that guides the restoration robot. Firstly, the Principle Component Analysis (PCA) [8] was taken as a powerful tool for analysis of the damaged region. Then, one of the three directions formed by PCA that has the least eigenvalue was used as the direction along which the damaged region was layered. And thirdly, for each layer section, a convex hull was found.

For each convex hull, some lines that perpendicular to a specula direction were used to slice the border with a specific interval, and the intersections were taken as the origin and end of a sub path. There are two important issues: how to determine the specific direction and how to determine the interval of the lines.

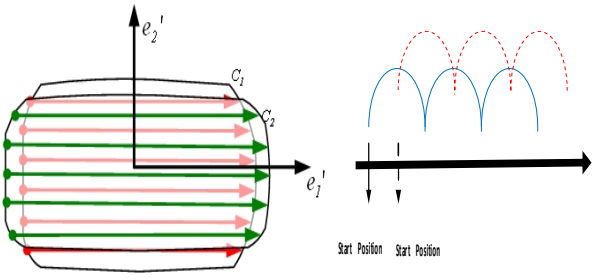


Fig. 5. Welding robot path generation

One principle of the restoration with arc welding is to make less path for more area, so we take the eigenvector which calculated by the PCA with largest eigenvalue as the direction. For the second issue, the base interval value is the width of a welding seam. The method of welding restoration process adopts multiple layers welding and multistage welding. We design a path generation method that is making next layer's interval same as the previous one and making the fist position equal the last one plus the interval (Figure 5) [7]. The restored surface is more even after several overlaps between contiguous torch passes by the method.

4 Experiment

An A3 steel plate with two testing slots is used to do experiment. The original plate and the scanning data are shown in Figure 6. After obtaining the broken region information, the restoration path can be generated as Figure 7.

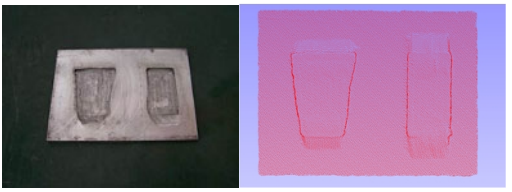


Fig. 6. Testing Steel Plat and Scanned Data

Once the restoration path is generated, the slots on the steel plate can be repaired by the welding parameters shown in table 1. The flatness of the welded plate (Figure9) is less than 0.5 mm.

Table 1. Processing parameters of Repair by pulsed PAW

Parameter	*Value*
Lon flow L/min	0.5
Protection gas flow L/min	12
Crest current IP(A)	230
Base current Ib(A)	50%IP
Speed of welding m/min	0.2
Speed of wire feed m/min	1.1
Interval L(mm)	4
Height h(mm)	1.2
Lap rate	0.33

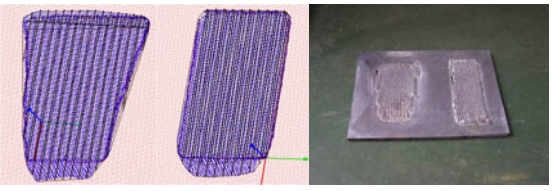


Fig. 7. Restoration path and Welded restoration effect

5 Conclusion

The proposed RS makes the structured light sensor inspection and the welding robot restoration into a closed loop system which can execute variety of restoration tasks automatically or semi- automatically.

It should be pointed out that, in this study, the system movement model adopts the uniform linear motion model, and the strategy of path generation employed is based

on a simplified welding deposition model. However, the actual welding process is probably very complicated. The current method can not solve most of complicated restoration problems. Therefore, to meet the various needs of the real restoration works, we will emphasize our future work on how to generate work path base on not only the scanning data but also the welding technology, and improving the current movement model to make it more precise and more practical.

References

1. Lavoie, P., Ionescu, D., Petriu, E.M.: 3-D Object Model Recovery From 2-D Images Using Structured Light. IEEE Transactions on Instrumentation and Measurement 53, 437–443 (2004)
2. Zhao, J., Ma, Z., Zhu, Q.M., Lin, N.: A novel adaptive laser scanning sensor for reverse engineering measurement. Chinese Journal of Scientific Instrument 28(7), 1164–1169 (2007)
3. Zhang, Z.: A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(11), 1330–1334 (2000)
4. Levenberg, K.: A Method for the Solution of Certain Non-Linear Problems in Least Squares. The Quarterly of Applied Mathematics, 164–168 (1944)
5. Gelfand, N., Rusinkiewicz, S., Ikemoto, L., Levoy, M.: Geometrically Stable Sampling for the ICP Algorithm. In: Proc. International Conference on 3D Digital Imaging and Modeling, pp. 77–93 (2003)
6. Li, B., Schnabel, R., Klein, R., Cheng, Z.Q., Dang, G., Jin, S.Y.: Robust normal estimation for point clouds with sharp features. Computers & Graphics 34, 94–106 (2010)
7. Zhu, S., Liang, Y.Y.: Path planning for MIG surfacing of robot-based remanufacturing system. China Welding 15, 59–62 (2006)

Optimal Tool Replacement Decision Method Based on Cost and Process Capability

Yu-hao Deng, Hai-ping Zhu, Guo-jun Zhang, and Hui Yin

State Key Laboratory of Digital Manufacturing Equipment & Technology,
Huazhong University of Science & Technology, Wuhan 430074
yuhaodeng@163.com, haipzhu@263.net,
zgj@mail.hust.edu.cn, 57496271@qq.com

Abstract. This paper distinguished "good" tools and defective tools considering the high tool failure rate in the initial cutting process, so the distribution function of tool life was given combined with the lognormal distribution. The paper optimized the tool replacement interval under the restriction of process capability, to minimize the tool using cost in every unit time. In addition, the effects tool grinding have on the cost was analyzed, to seek for the best combination of tool grinding interval and grinding times, and further decrease the using cost.

Keywords: Tool Replacement Strategy, Process Capability, Tool Grind.

1 Introduction

There are generally two different kinds of tool replacement method in the actual production process. One is replacement after failure until the tool cutting life ends, based on which the current tool replacement method is put forward usually. The other is preventive replacement, which happens when the precision of the product size is strictly required and the operator should change the tool as soon as it cannot satisfy the quality requirement, and the tool using life here is called quality life.

The research of the tool cutting life has experienced a change from summarized empirical formula to the distribution function [2]. Liu Zhanqiang[3] proved that the wear failure cutting life submits to lognormal distribution, through the analysis of field data collected from the NC machine tool life; Weibull distribution can also be used to describe tool wear process because of its good adaptability [4].

The quality life of the tool is available by measuring the process capability index C_p [6], based on which, this paper optimized the tool replacement interval to minimize the tool using cost in every unit time; then further considered the case of regular tool maintenance, to get the optimal combination of grinding interval and grinding times.

2 Tool Replacement Interval Decision

Two primary tool failure forms in the cutting process are wear and tear. Wear failure can be described by log-normal distribution or Weibull distribution; tear failure can

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 9–14.
springerlink.com

be reflected by the defective rate with "bad" tools which share the same distribution function with the "good" ones, but has smaller cutting life parameter than "good" tools. In addition, failure rate λ_S is introduced to present external impact.

Assuming that the defective rate is p, the wear life of “good" tools submit to the log-normal distribution with a logarithmic mean μ_1 and a logarithm standard deviation σ_1, while “bad" tools do with μ_2 ($\mu_1<\mu_2$) and σ_2. Taking the above three factors into account, the reliability function of the tool at time t is:

$$R(t)=\left[(1-p)(1-\int_0^t f_1(t)dt)+p(1-\int_0^t f_2(t)dt)\right]\exp(-\lambda_s t),$$
$$f_m(t)=\frac{1}{\sqrt{2\pi}\sigma_m t}\exp\{-\frac{[In(t)-\mu_m]^2}{2\sigma_m^{\ 2}}\} \qquad (m=1,2) \tag{1}$$

2.1 Tool Using Cost in a Unit Time

If the planed tool replacement interval is T, the actual tool life is t, the tool reliability function at moment t is $R(t)$, there will be two situations during the machining process:

(1) $t \geq T$, the tool is working until the planned replacement time, and only replacement cost arises. Suppose there are N times of tool replacement in the continuous machining processing, and ignore the tool replacement time. The cost of this case is $C_1 = N \bullet R(T) C_t$, and the total processing time is $T1 = N \bullet R(T) T$.

(2) $t<T$, the tool wear out before the planned time, which leads to the disqualification of the product and product loss cost arises additionally. So the cost is $C_2 = N \bullet [1 - R(T)] (C_b + C_t)$, due to the randomness of failure, t may be any value between 0 to T, the total machining time of this part is the integration of the product of t and the failure probability at the t moment, so

$$T_2 = N \cdot [1-R(T)] \int_0^T t \frac{f(t)}{1-R(T)} dt \tag{2}$$

$\frac{f(t)}{1-R(T)}$ is the failure probability at any time t during a replacement cycle T.

The expected cost in a unit time is the ratio of the total cost and total time, and the result is shown as equation (3).

$$E_t(C) = \frac{\sum C}{\sum T} = \frac{[1-R(T)]C_b + C_t}{R(T)T + \int_0^T tf(t)dt} \tag{3}$$

2.2 Tool Cutting Process Capability Index

Because of the general tool wear and tear, the machining dimension has internal increasing or decreasing tendency. What’s more, auto-correlation exists in the standard deviation of the machining error. At the t moment before cyclical tool adjustment, the machining error of the product is built with a linear trend:

$$x_t = a + bt + \varepsilon_t, t \le T$$

Where, a and b are linear parameters, ε_t is residual term, T is replacement interval.

For the process with trend terms, the deviation comes from two different sources; Xie [7] expressed the process variance as:

$$\sigma_p^2 = \sigma_T^2 + \sigma^2$$

Where, σ_T^2 is non-random variance caused by the trend terms, σ^2 represents the identical variance after removal of the trend terms.

It is obvious that the mean size is $\mu = a + b \times T/2$ in the cyclical tool replacement process, as a result,

$$\sigma_T^2 = \frac{\sum_{t=1}^{T}(a+bt-\mu)^2}{T} = \frac{b^2}{6}\left(\frac{T}{2}+1\right)(T+1) \tag{4}$$

When the replacement interval T is relatively big,

$$\sigma_T^2 \approx \frac{b^2 T(T+3)}{12}, \quad \sigma_p^2 \approx \frac{b^2 T(T+3)}{12} + \sigma^2 \tag{5}$$

Then the modified process capability index is:

$$C_p^* = \frac{USL-LSL}{6\sqrt{b^2T(T+3)/12+\sigma^2}}, C_{pk}^* = \frac{\min(USL-\mu, \mu-LSL)}{3\sqrt{b^2T(T+3)/12+\sigma^2}} \tag{6}$$

2.3 Tool Replacement Interval Decision Model

If customer appoint the smallest process capability C_{pl} or C_{pkl} , the mathematic optimization model for tool replacement interval can be established as below:

$$\begin{aligned} &\min\ E_t(C) = \frac{[1-R(T)]C_b + C_t}{R(T)T + \int_0^T tf(t)dt} \\ &s.t.\ \frac{USL-LSL}{6\sqrt{b^2T(T+3)/12+\sigma^2}} \ge C_{pl}, or \frac{\min(USL-\mu, \mu-LSL)}{3\sqrt{b^2T(T+3)/12+\sigma^2}} \ge C_{pkl} \end{aligned} \tag{7}$$

3 Regular Tool Maintenance Decision

The model established in the last section has not take tool maintenance into consideration, which is in fact common during the usage of the tool.

To develop a regular tool grinding plan, some assumptions are given that the regularly tool grinding cycle is T_M, the grinding cost is C_M, and the tool is grinded after a grinding cycle, or replaced if it is discard as useless during the grinding cycle. The tool must be replaced after grinding M times for standardized management as

well as convenient calculation. Tool grinding makes the tool "as good as old", while tool replacement makes the tool "as good as new".

Case (1): The tool is replaced after M successful tool grinding cycle, the cost and machining time are respectively calculated.

$$C_1 = N \cdot R(M \cdot T_M)[(M-1)C_M + C_t]$$
$$T_1 = N \cdot R(M \cdot T_M)M \cdot T_M$$

Case (2): After the mth grinding, the tool breaks down in the $m+1$th grinding cycle and is then replaced, where both product loss cost and replacement cost happen.

$$C_2 = \sum_{m=0}^{M-1} N \cdot P[T < (m+1)T_M \,|\, T \geq mT_M](C_b + mC_M + C_t)$$
$$T_2 = N\int_0^{MT_M} tf(t)dt$$

Where: $P[T < (m+1)T_M \,|\, T \geq mT_M] = 1 - \dfrac{P[T \geq (m+1)T_M, T \geq mT_M]}{P(T \geq mT_M)} = 1 - \dfrac{R[(m+1)T_M]}{R(mT_M)}$

As to the process capability, grinding can only make the tool “as good as old”, the residuals after removing the variance term trend will be increasing heteroscedasticity instead of identical, which need to be validated by Glejser test method.

If $|\varepsilon_t| = \alpha_0 + \alpha_1 t$, then $Var(\varepsilon_t) = E(\varepsilon_t)^2 = (\alpha_0 + \alpha_1 t)^2$

If $|\varepsilon_t| = \alpha_0 + \alpha_1 t^2$, then $Var(\varepsilon_t) = E(\varepsilon_t)^2 = (\alpha_0 + \alpha_1 t^2)^2$

Take $|\varepsilon_t| = \alpha_0 + \alpha_1 t$ for example, The mathematic optimization model for regular tool grinding interval is:

$$\min\ E_M(C) = \frac{R(MT_M)[(M-1)C_M + C_t] + \sum_{m=0}^{M-1}\left\{1 - \dfrac{R[(m+1)T_M]}{R(mT_M)}\right\}(C_b + mC_M + C_t)}{R(MT_M)MT_M + \int_0^{MT_M} tf(t)dt} \tag{8}$$

s.t.

$$\frac{USL - LSL}{6\sqrt{b^2 T_M (T_M + 3)/12 + (\alpha_0 + \alpha_1 T_M)^2}} \geq C_{pl}, or \frac{\min(USL - \mu, \mu - LSL)}{3\sqrt{b^2 T_M (T_M + 3)/12 + (\alpha_0 + \alpha_1 T_M)^2}} \geq C_{pkl}$$

4 Examples

A workshop uses XK715 CNC machine and cemented carbide end mills to accurate mill a planar, with a required thickness of $20^{0}_{-0.01} mm$ .Ignoring the occasional failure caused by external shocks, the cutting life distribution function of the tool is:

$$R(t) = 0.98\left\{1 - \int_0^t \frac{1}{\sqrt{2\pi} \times 0.26t} \exp\{-\frac{[In(t) - 3.09]^2}{2 \times 0.26^2}\}dt\right\}$$
$$+ 0.02\left\{1 - \int_0^t \frac{1}{\sqrt{2\pi} \times 0.18t} \exp\{-\frac{[In(t) - 2.08]^2}{2 \times 0.18^2}\}dt\right\}$$

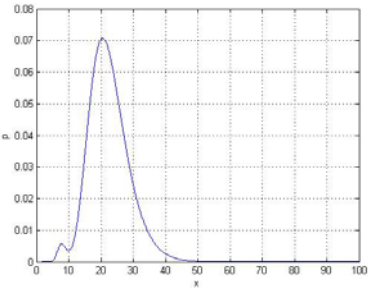


Fig. 1. The failure probability distribution of the cemented carbide end mills

Randomly select four end mills for experiment and measure the thickness of the work piece every hour until 25 points are collected for each tool. The machining error is the measured and shown in Fig. 2.

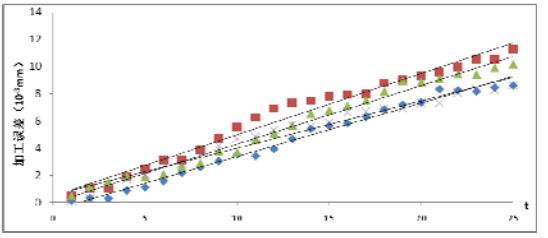


Fig. 2. Collected data of machining error

The machining process needs to be monitored by the requirement of customer, and C_{pl} cannot be less than 1.0. Apparently the machining error trends to increase linearly. The linear regression result is $x_t = (0.1006 + 0.4055t) \times 10^{-3} + \varepsilon_t, \varepsilon_t \sim N(0, 0.000292^2)$, so the process capability index restraint can be expressed as below:

$$\frac{0.01}{6\sqrt{0.0004055^2 T(T+3)/12 + 0.000292^2}} \geq 1$$

MATLAB is used to solve the model. When C_t=200RMB and C_b=100RMB, the minimum cost is $E_t(C)$ =14.89RMB / h which is obtained at the boundary T=12.06h.

When the product loss cost is much greater than the fixed tool replacement cost, the optimal tool replacement interval will be achieved within the time range restrained by process capability index. For example, when C_t=100RMB and C_b=450RMB, the optimal result is $E_t(C)$ =9.17RMB / h when T=13.49h.

When the case goes to regular maintenance, machining error during the grinding cycle is $x_t = (0.1006 + 0.4055t) \times 10^{-3} + \varepsilon_t$, and the result of the Glejser test of residual item is $|\varepsilon_t| = (0.312 + 0.004t) \times 10^{-3}$. When C_t=200RMB, C_b=100RMB, and C_M=20RMB, by setting M respectively with 1,2,3,… in formula (8), the optimal result is figured out to be $E_t(C)$ =13.84RMB / h when M=2, T_M =10.48h. 1.05RMB/h less than process without tool grinding, and the total machining time of the tool is 20.96h.

5 Conclusions

By the calculation and analysis of the model, conclusions can be made that the minimum cost is usually obtained at the boundary of the process capability index. Only when the product loss cost is much greater than the fixed tool replacement cost, the optimal tool replacement interval will be got within the quality life restrained by the process capability index; in addition, regularly tool grinding can greatly extend the quality life, and further reduce the tool using cost.

There are two places for improvement of the model:

1) The wear will continue to increase with the use of the tool, grinding intervals should be shortened time after time, so the tool grinding cycle will not be a constant value, but a decreasing Series group;

2) Decision can be made according to the usage and wear degree of the cutting tools, and the grinding probability and replacement probability at tool grinding time can be introduced to optimize the model.

Acknowledgements. The authors greatly acknowledge the financial supports from the National High Technology Research and Development Program of China (863 Program) with the Grant number 2009AA043301.

References

1. Zheng, D., Sun, L.: Study on tool wear and life based on dimensional inspection data. Modular Machine Tool and Automation Technology 8(4), 32–35 (2002)
2. Vagnorius, Z., et al.: Determining optimal replacement time for metal cutting tools. European Journal of Operational Research 206(2), 407–416 (2010)
3. Liu, Z., Li, Z., Ai, X.: Test on the spot for Scattered Pattern of Cutter Wear Life. Manufacturing Echnology & Machine Tool (12), 8–10 (1995)
4. Rausand, M., Hyland, A.: System Reliability Theory; Models, Statistical Methods, and Applications. Wiley, Hoboken (2004)
5. ElWardany, T.E., Elbestawi, M.A.: Prediction of tool failure rate in turning hardened steels. International Journal of Advanced Manufacturing Technology 13(1), 1–16 (1997)
6. He, Z., Zhang, M., Dong, Y.: A Study on Tool Life Based on Process Capability Indices. China Mechanical Engineering 17(3), 295–297 (2006)
7. Xie, M.: Process Capability Indices for A Regularly Adjusted Process. International Journal of Production Research 40(10), 2367–2377 (2002)

A Study on the Influence of the Computer-Aided Translation (CAT) Technology on the Quality of the Translated Text

Ying Wu[1, 2], Xiuhua Zhuang[1], and Qi Pan[1]

[1] Foreign Language Department, Zhenjiang Watercraft College of the PLA, Zhenjiang 212003, Jiangsu, P.R. China
[2] English Department, School of Foreign Studies, Nanjing University, Nanjing 210093, Jiangsu, P.R. China
wycherry@sina.com

Abstract. The present thesis reports a study on the quality of the translated text in computer-aided translation practice. The author assumes that the quality of the translated text may be greatly influenced by the quality of the translation memory (TM) database. That is to say, the poor quality of the TM database may lead to the poor quality of the translated text. Then a translation experiment is designed to test whether the problem proposed by the author do exist. After the experiment and the data analysis process, the results prove the existence of the problem and support the assumption of the author.

Keywords: Computer-Aided Translation, Translation Memory, Quality of Translated Text.

1 Introduction

With rapid development of the Computer-Aided Translation (CAT) technology, the Translation Memory System (TMS) has gradually become one of the most popular tools among professional translators. It is generally accepted that the TMS is helpful in improving the efficiency of the translation procedure and the consistency of the translated text as it is designed for reusing the previously translated materials to save time and energy of repetitive work [1]. However, it cannot be avoided that the TMS may also have some negative effects on the translation activity, such as the quality of the translated text [2]. Just as Mogensen points out that the application of the TMS may make the resulting text less readable because of the relative absence of cohesive devices between sentences. She also explains that using the TMS may lead to an oversimplified translation because the sentences are often short [3]. In this thesis, the author attempts to make some preliminary research on the influence of the quality of the Translation Memory (TM) and the attitudes of translators to the TM on the quality of the translated text, hoping that the results and findings of this research may provide some ideas for future studies and some references for the translation practices with the TMS.

This thesis attempts to make an empirical study on the influence of the quality of the TM to the quality of the translated text, especially on the existence of the repetition of

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 15–20.
springerlink.com

mistranslations and other types of mistakes. Meanwhile, this research is designed to make a survey on translators' attitudes towards the application of the TMS, trying to find out how translators' attitudes may affect the quality of the translated text.

2 Methodology

2.1 General Conception

The author thought that if translators failed to recognize the preset errors in the TM database and adopted the false translation suggestions in the translated text, the result would be that the translated text also contained the mistakes like those in the TM database. Therefore, by checking these mistakes in the translated text, we could find out whether the quality of the translated text was impacted by the poor quality of the TM database. Bearing this in mind, the author decided to build up a TM database with some mistakes preset in it. Then the translators were required to translate a text with the help of this database. Finally, the translated texts would be checked to see whether these mistakes were repeated in them. At the same time, an evaluation method would be used to assess the quality of the translated text. The result of this evaluation would be the evidence of whether the quality of the translate text was influenced by the quality of the TM database.

2.2 Evaluation Criteria

The evaluation of the quality of the translated text in the experiment would be taken out according to the Target Text Quality Requirements for Translation Services (GB/T 19682－2005) (hereafter abbreviated as the TTQRTS) [4] formulated by the General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China and the Standardization Administration of the People's Republic of China. The TTQRTS was a national standard issued in 2005 for regulating the translation service in the People's Republic of China. It provides a set of complete methodology for the evaluation of translation quality.

The quality of a translation is judged by a concept called Comprehensive Error Rate (hereafter abbreviated as the CER) in the TTQRTS. Generally speaking, the CER should be no more than 1.5‰. Otherwise, the translation is considered unqualified. The value of the CER is calculated from the following formula:

$$KC_A = \frac{c_I D_I + c_{II} D_{II} + c_{III} D_{III} + c_{IV} D_{IV}}{W} \times 100\% \quad (1)$$

In this formula, K is a difficulty coefficient with a suggested value from 0.5 to 1.0. C_A is a coefficient for the four categories of translation purposes. Accordingly, C is 1 for the first group, 0.75 for the second group, 0.5 for the third group and 0.25 for the fourth group. W stands for the number of the total words. C_I、C_{II}、C_{III} and C_{IV} are coefficients for the four groups of mistakes: C_I =3, C_{II} =1, C_{III} =0.5, and C_{IV} =0.25. Finally, D_I、DII、DIII、D_{IV} stand for the number of times that mistakes of the four groups appear in the translated text.

2.3 Construction of the TM Database

Among numerous TMS tools available in the market, Trados was selected as the instrument of the present experiment. Trados was founded in 1984 to provide translation services to IBM. In 1997, Microsoft decided to use Trados software as its internal translation memory system [5]. From then on, Trados has become one of the most widely used TMS tools around the world. Then, the author extracted about 170 words from an English text on the Eating Guideline for Lacto-Ovo Vegetarians as the source text of the experiment. Another text on the similar subject and the same type, namely, the Eating Guidelines for Vegans, was selected as the source material of the TM database. Based on some basic language processing steps, the text was first divided into small segments before it was added into the TM database. Generally speaking, the segmentation was based on a sentence or a sentence-like structure. The source material of the TM database was divided into 12 segments at the sentence level. Then, in the alignment process, each segment was matched with its Chinese translation, which contained some kinds of mistakes. Altogether 14 mistakes belonging to the two mistake groups were set up in the material. Finally, the Translators' Workbench of the Trados software was started, and the aligned segments were input to a TM database in the Trados.

3 Results and Discussion

3.1 CER Comparison

Among 30 participants of the experiment, 29 are considered valid in the final calculation since the material of one subject (No. 15) was missing in the collection. The number of mistakes in the translated texts, the comprehensive error rate, and their respective mean values are all calculated. According to the results, among 14 mistakes preset in the TM database, the highest number of mistakes found in the translated text is 10 while the lowest is 0, with a mean value being 4.2. The highest Rate of Correctness is 100% and the lowest is 28.6%, with a mean value of 70%. The highest Comprehensive Error Rate is 14.0‰ and the lowest is 0, the mean value being 6.4‰.

Furthermore, Table 1 shows a comparison of the CERs with the quality standard of the TTQRTS. According to Table 1, among 29 valid participants, the CERs of 4 subjects are no more than 1.5‰, and the CERs of 25 subjects are more than 1.5‰. At the same time, the mean value of the CER of all 29 participants is more than 1.5‰.

Table 1. CER Comparison

	CER ≤ 1.5‰	CER > 1.5‰	Total
Number of Subjects	4	25	29
Mean Value of CER > 1.5‰			
Qualification Rate = 13.8%			

It means that among the 29 valid subjects, only 4 reach the quality standard of the TTQRTS (i.e. CER≤1.5‰), and the other 25 participants fail to do that. In other words, it means that only 4 translated texts are in good quality, while the other 25 texts are in poor quality. As for the mean value, the CER of the 29 participants is 6.4‰, which also fails to reach the requirement of the TTQRTS. As a result, the total qualification rate is only 13.8%. That is to say, in a whole, the 29 translated texts are poor in quality. Therefore, considering the purpose of this experiment, it is safe to make a conclusion that the poor quality of the TM database is able to lead to the poor quality of the translated text.

Such a result is in consistence with the idea of Frank Austermuhl [6], who claims that the recycling of past translations may lead to the repetition and reproduction of mistakes. There may be several reasons that account for this problem discovered in the experiment, for example, the design of the TM software, the attitude of the translator, the requirement of the client etc,. After all, the TMS is designed to help translators reduce the time and energy spent on the repetitive work in translation, so it is natural that the translators trust the suggestion of the software and rely on the operation of the system to improve their working efficiency. What's more, some clients may require that translators should only work in accordance with the suggestion of the TM database provided, and some even put forward that no pay would be given for the repetitive part of the article. In such a case, translators would be unwilling to check the translation suggestions of the TM database and intend to accept whatever was provided by the system.

3.2 RC of Mistakes in Different Groups

In addition, the present research also provides a perspective to see what kind of mistakes has the highest rate of repetition in the translated text. Table 2 shows the Rate of Correctness (RC) of each category of mistakes.

Table 2. RC of Mistakes in Different Groups

Mistake Categories	Absence		Mistranslation	
Matching Types	Exact Match	Fuzzy Match	Exact Match	Fuzzy Match
No. of Mistakes	1, 2	5, 6, 7, 8, 10, 11, 12	9, 13, 14	3, 4
No. of Group	A	B	C	D
Rate of Correctness (RC)	62.1%	87.7%	43.7%	46.6%

According to Table 2, the RC of mistakes in Group A (i.e. the absence of words, phrases or sentences in the fuzzy match) is 62.1%, of Group B is 87.7%, of Group C is 43.7% and of Group D is 46.6%. That is to say, the mistakes classified as the absence of words, phrases or sentences in the fuzzy match (i.e. Group B) have the highest rate of correctness, while mistakes classified as the mistranslation of words, phrases or sentences in the exact match (i.e. Group C) have the lowest rate of correctness. In other words, the mistakes in Group B have the lowest rate of repetition in the translated text while those in Group C have the highest rate of repetition. It means that it is easier for

translators to repeat the mistakes of Group C in the translated text while not so likely to repeat the mistakes of Group B in the same condition. Therefore, it can be further concluded that mistakes belonging to the category of mistranslations and presented in the exact match of the translation suggestions are most likely to be repeated in the translated text.

In addition, Table 2 also shows that Group B has a higher rate of correctness than Group A, and Group D has a higher rate of correctness than Group C. That is to say, in the same mistake category (i.e. Group A and B, Group C and D), mistakes presented in the fuzzy match (i.e. Group B and D) have a higher rate of correctness than those in the exact match (i.e. Group A and C). In other words, mistakes in the exact match are more likely to be repeated in the translated text than those in the fuzzy match.

Such a result indicates that mistakes in the exact match are more invisible to translators than those in the fuzzy match, and that translators tend to believe in suggestions provided by the system in the exact match than in the fuzzy match. The explanation to this phenomenon can be very complicated. It is possible that translators have a tendency to equal the exactness with the correctness, and to think that the 100% match means the 100% correct. Therefore, they are more likely to check the translation suggestions in the fuzzy match than those in the exact match.

The problems revealed in the findings of the experiment are important not only to the quality improvement of the translation service providers, but also to the healthy development of the computer-aided translation technology. Therefore, it is necessary to pay enough attention to these problems and take up some measures to minimize their influences.

4 Conclusion

The results of the experiment show that the poor quality of the TM database will lead to the poor quality of the translated text. What's more, it is further concluded that mistakes belonging to the category of mistranslations and presented in the exact match of the translation suggestions are most likely to be repeated in the translated text, and that mistakes in the exact match are more likely to be repeated in the translated text than those in the fuzzy match.

The problems revealed in the findings of this research may be of great importance either to the cultivation of translators, the quality improvement of translation service, or to the healthy development of the computer-aided translation technology. Therefore, some measures must be taken to minimize the negative influences of these problems in these fields. In the author's view, at least two aspects should be paid attention to in this process. One is the management of the TM database, and the other is translators' awareness of these problems.

On the one hand, the management of the TM database is an objective way to solve the problem. If no mistakes existed in the TM database, then the problems described above would be impossible to happen at all. On the other hand, the translators' awareness is a subjective factor in solving the problem. After all, the translated text is an output of translator's conceptual work. It is the translator who has the final decision on what should be accepted and what should not. If translators were aware of these problems and were able to overcome their negative influences intentionally, it would be possible that these problems were significantly reduced.

References

1. Bowker, L.: Computer-Aided Translation Technology: A Practical Introduction. University of Ottawa Press, Ottawa (2002)
2. Macklovitch, E., Russell, G.: What's Been Forgotten in Translation Memory. In: White, J.S. (ed.) AMTA 2000. LNCS (LNAI), vol. 1934, pp. 137–146. Springer, Heidelberg (2000)
3. Mogensen, E.: Orwellian Linguistics: How using computer-aided translation tools impacts the target language. Language International 12, 28–31 (2000)
4. Target Text Quality Requirements for Translation Services (GB/T 19682 — 2005), http://www.sunyu.com/wenzhanglist.asp?id=65
5. The History of Trados Development, http://www.oktranslation.com/Forums/ShowPost.aspx?PostID=2823
6. Austermuhl, F.: Electronic Tools for Translators. Foreign Language Teaching and Research Press, Beijing (2001)

Energy Saving System for Home Energy Measurement and Efficient Power Control

Kwang-Soon Choi[1], Young-Choong Park[1], Yang-Keun Ahn[1], Kwang-Mo Jung[1], Sukil Hong[2], and Ha-Bong Chung[3]

[1] #1599, Sangam-Dong, Mapo-Gu, Seoul, 121-835, Korea
[2] Soong-Sil University, 511, Sangdo-Dong, Dongjak-Gu, Seoul, 156-743, Korea
[3] Hong-Ik University, 72-1, Sangsu-Dong, Mapo-Gu, Seoul, 121-791, Korea
{lenon,ycpark,ykahn,jungkm}@keti.re.kr,
hongsukil@gmail.com,
habchung@wow.hongik.ac.kr

Abstract. Energy Saving System (ESS) is designed and implemented for real-time home energy measurement and the efficient power control at homes or buildings. For energy saving functions and the future extension to the Smart Grid, architectural and functional requirements of ESS server and ESS client are defined. In addition, prototypes of ESS components such as ESS server, ESS client and portable ESS terminal are implemented for the practical usage.

Keywords: Energy Saving System, ESS, ESS server, ESS client, portable ESS terminal.

1 Introduction

Due to the global warming and climate change issues caused by reckless use of energy, the efforts to develop new technologies for efficient energy generation and consumption are in progress. To enable low carbon and green growth especially in this green IT generation, Information and Communication Technologies (ICT) are being widely used all over the industrial fields. The Smart Grid technologies that enable the eco-friendly intelligent power grid for efficient harvesting, distribution, and consumption of electric energy, are evolving with technology development and global competition for international standards. At present, the global demand of electric energy and the emission of greenhouse gases are assumed to be increased up to 82% and 59% until 2030, respectively [1]. However, by deploying the Smart Grid, we can get energy saving effects through the reduction of the electric energy loss caused during transmission and distribution, integration of renewable energies, feedback of energy consumption information, and implementation of demand response. We can also reduce the emission of 20.3 tons of CO2 that amount approximately 124 billion dollars [2].

In this paper, we implemented ESS that can interoperate with Smart Grid and is applicable in residential area and office buildings. ESS basically provides users with information on electric energy consumed by each home appliance so that the users

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 21–26.
springerlink.com

can voluntarily reduce the energy consumption. In addition, ESS provides various methods to control power consumption of home appliances from the remote site as well as intelligent energy saving functions.

2 Design of ESS

The ESS network basically consists of one ESS server and one or more ESS clients as shown in Fig. 1 [3]. For user's convenience, various types of portable ESS terminals that communicate with the ESS server are provided.

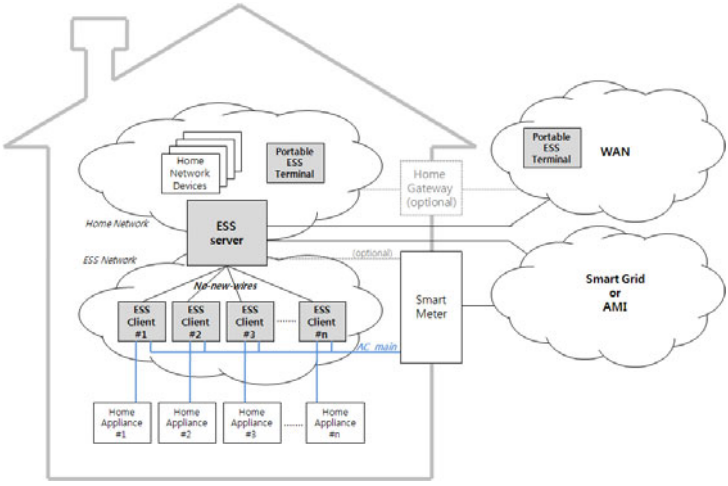


Fig. 1. Overview of ESS

The ESS server is connected with ESS clients through no-new-wires such as ZigBee, Wi-Fi, PLC, etc. The ESS server can be implemented either as a home gateway (or home server) type by adding the ESS server functions in it or as a standalone server type. In the first case, the ESS server can support both the existing home network and new ESS network within the Home Area Network (HAN) environment. It also supports connection to Wide Area Network (WAN) and Smart Grid. In the second case, the server can support the ESS network only within the Home Area Network (HAN) environment. It also supports connection to Wide Area Network (WAN) and Smart Grid. The ESS server must provide the connection to Smart Grid or Advanced Metering Infrastructure (AMI) in both cases. The Smart Meter enables the connection between the ESS server and AMI. The ESS server can connect ESS clients and portable terminals through the ESS network. Additionally, the ESS server plays a gateway role to connect between the inner part and the outer part as shown in Figure 1. The functional requirements of ESS server are listed in Table 1.

Table 1. Designed functional requirements of ESS server

Connectivity (in HAN)	• Wire/wireless interfaces to ESS clients • Wireless interfaces to portable ESS terminals • Wire/wireless interfaces to the home network (optional)
Connectivity (in WAN)	• Connection to the Internet (connection to the outdoor portable ESS terminals) • Connection to Smart Grid (optional) • Connection to AMI through Smart Meter (optional)
Functions	• Support of EPCM protocol • Real-time electric energy consumption monitoring for individual appliance • Power control for each appliance group • Home ID setting • Remote power control • User interface functions • Security functions to protect private information • Standby mode learning and automatic standby power cutoff functions for appliances (optional) • Automatic power control for individual appliance by registering user events (optional) • Monthly electricity bill estimation (optional)

The ESS server can be implemented with five different types: Hand-held terminal such as smart phone, display device (wall-pad or TV) with the ESS server function, home gateway or home server with the ESS server function, PC with the ESS server function, and always-on home appliance (e.g. refrigerator) with the ESS server function.

An ESS client basically supplies or blocks AC power to a home appliance connected to it. The ESS client can also measure the power consumed by the home appliance and deliver the power consumption information to the ESS server. To measure the power consumption and communicate with the ESS server, DC power components including power metering module, processing unit, low-power communication module (LPCM), and AC power switch, are required in the ESS client. For this purpose, a small-size of AC-DC converter is especially designed. The processing unit controls the power metering module to measure the power consumption and to deliver the measured data to the ESS server through LPCM. It also controls the AC power switch according to the power control commands from the ESS server. LPCM communicates with the ESS server. In this study, the IEEE802.15.4 based Ubiquitous Sensor Network (USN) module was used as an LPCM. A power metering module calculates the current, voltage, and energy consumed by home appliances. Then the processing unit periodically reads the calculated results from the power metering module. The power metering module uses the resistive divider circuitry and shunt resistor to measure the consumed voltage and current. LPF is used to internally remove high frequency and the digital filter is used for phase compensation between current and voltage. It also adjusts the measured values for DC offset, AC offset, power offset, and gain. The functional requirements of ESS client are listed in Table 2.

Table 2. Functional requirements of ESS client

Connectivity	• Wire/wireless interfaces to ESS server
Functions	• Support of EPCM protocol • Real-time electric energy consumption measurement • Home ID setting • AC output control for appliances • User interface functions • Support of active, standby, and sleep modes • DC output control for low power operation (optional) • Standby mode learning and automatic standby power cutoff functions for appliances (optional)

A portable ESS terminal is connected with the ESS server through wireless communication such as Wi-Fi, ZigBee, etc. and receives necessary data from the server. The users can query the power consumption status through the portable ESS terminal in real time from a local or remote site, and control the power of home appliances connected to ESS clients.

3 Implementation of ESS

The ESS server was implemented with three different types as shown in Table 3. The ESS client was implemented as an adaptor type shown in Fig. 2. For the low-power design of ESS client, a switch has been implemented with an apparatus in the AC inlet part. Through this implementation, the AC input power is not supplied to both the home appliance and the internal AC-DC conversion module unless any home appliance is plugged in the ESS client. In this case, the ESS client consumes no power (0 Watt). The ESS client supports two types of operation modes, which can be set by an external DIP switch. If it is set to “0” (/ESS_MODE = ‘0’), the function specified in Table 2 is performed. If it is set to “1” (/ESS_MODE = ‘1’), the ESS clients acts like a normal electric outlet. No metering and communication functions are enabled. As shown in Fig. 2 (b), the IR receiver that is separately located outside of the ESS client can be connected with the ESS client. The IR receiver can learn the signals of the remote control of the home appliance connected with the ESS client. While the AC power output is off by a power-off command from the ESS server, when the signal learned by the remote control is received, the ESS client switches on the AC power output to the home appliance so that the home appliance can operate. This function is the most effective and intuitive method to supply power again to the inactive home appliance whose standby power has been automatically cut off.

The portable ESS terminal supports the IP communication with the ESS server. Using a portable ESS terminal, a user can connect with the ESS server to monitor the power consumption status for a specific home appliance and control the power. The designed portable ESS terminals have two different types as shown in Table 4.

Table 3. Prototypes of ESS server

Type of ESS server	Descriptions	Prototype
Hand-held device type	hand-held H/W platform with embedded Linux	
Display device type	ESS server functions added in a commercialized wall-pad system	
PC type	ESS server application for Windows and Linux PC	

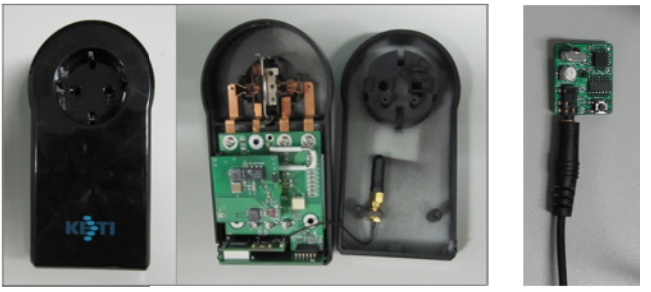


Fig. 2. Prototype of ESS client

Table 4. Prototypes of portable ESS terminal

Type of portable ESS terminal	**Descriptions**	**Prototype**
Hand-held device type (only for indoor usage)	hand-held H/W platform with LCD module and 8-bit CPU	
Smart Phone type	ESS terminal application for Android smart phone (Samsung's Galaxy Tab)	

4 Conclusion

In this paper, we defined and implemented the ESS network that consists of ESS server, ESS clients, and portable ESS terminals, as well as the functional requirements for the network. The ESS server has three different types: had-held device type, display device type, and PC type. The ESS client has been implemented with the electric outlet type. The portable ESS terminals have been implemented with two different types: hand-held device type and smart phone type.

For interoperation among them, a test bed has been implemented. The basic functions that have been defined in Table 1 and Table 2 have been successfully verified (optional functions excluded). If the interfaces and functional requirements for Smart Grid are specified as standards in the future, we will additionally develop the ESS server that supports them.

Acknowledgement. This research was supported by the "Seoul R&BD Program (PA090893)".

References

1. EIA, DOE, U.S.: International Energy Outlook (2007)
2. Global e-Sustainability Initiative: SMART 2020 – Enabling the low carbon economy in the information age (2008)
3. Choi, K., Ahn, Y., Park, Y., Park, W., Seo, H., Jung, K., Seo, K.: Architectural Design of Home Energy Saving System Based on Realtime Energy-Awareness. In: The 4th International Conference on Ubiquitous Information Technologies & Applications (2009)

A 6-Bit, 12.5 GS/s Comparator for High-Speed A/D Conversion in 0.35 μm SiGe BiCMOS Technology*

Kuai Yin, Qiao Meng**, Haitao Liu, and Kai Tang

Institute of RF- & OE-ICs, Southeast University, Nanjing, China
mengqiao@seu.edu.cn

Abstract. This paper presents a monolithic comparator implemented in CHRT 0.35 μm SiGe BiCMOS technology for high-speed medium-resolution flash analog-to-digital conversion. The SiGe BiCMOS process provides HBT transistors with maximum f_T as 55 GHz. The comparator occupies a die area of 240×65 μm^2 with a power dissipation of 54.16 mW from a 3.3-V power supply. Operating with an input frequency of 1 GHz, the circuit can oversample up to 12.5 GS/s with 7 bits of resolution; while operating at Nyquist, the comparator can sample up to 12.5 GS/s with 6 bits resolution. This design achieves a considerable trade-off between power, speed and resolution.

Keywords: Analog-to-digital conversion, comparator, high-speed, SiGe BiCMOS.

1 Introduction

As a bridge between the "real world" and the "digital world", the analog-to-digital converter (ADC) is essential for modern digital signal processing (DSP) systems. With the increase in performance of DSP circuits, substantial area savings and power reduction can be obtained to maintain high efficiency. To realize this goal, ADCs with high sampling rate, modest bit resolutions, and wide input bandwidths are required. As the comparator plays a key role in determining the overall performance of the ADC, the research of high sampling rate and medium resolution comparator certainly becomes the hot point.

In general, high-f_T HBT technology can bring much higher speed at the cost of power and area[2][3][12]; while the power dissipation of CMOS comparators is much less than HBT comparator but with poor sampling rate [4][5]. As reported, comparators in BiCMOS technology used for this work could be exploited to achieve the trade-off between operating speed and power dissipation [1][6][10][11].

In this paper, the design and implementation of a high-speed comparator working at a sampling rate of 12.5 GHz with a resolution of 6 bit is presented. The comparator consists of a pre-amplifier stage, a core latching comparator and an output buffer. The pre-amplifier is designed with high-f_T (~55 GHz) HBT to lessen the overdrive

* Sponsored by the Ph.D. Programs Foundation of Ministry of Education of China under 20090092110011.
** Corresponding author.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 27–34.
springerlink.com

recovery time and offer sufficient gain for the latch (short for latching comparator). The latch, which consists of two stages, amplification and regenerative, is designed to compare two voltage levels and produce the digital decision speedily as possible. In order to achieve a higher speed and lower power consumption, the current bias and the clock-controlled switches are replaced by only one small size MOSFET instead of several HBTs. A detailed analysis of this kind of structure could be found in [1]. The output buffer made up of a differential amplifier followed by two CMOS inverters is designed to produce well defined digital data.

The details of the comparator design are discussed in Section 2. Section 3 describes the layout of the comparator. Simulation results of the comparator are presented in Section 4, followed by a conclusion.

2 Circuit Design

In this design, all the devices applied for amplifying are implemented in high performance HBTs to meet the high-speed requirement, whereas the current bias and switch devices is implemented in standard CMOS technology to keep the power level low.

2.1 Architecture Overview

The block diagram of the proposed voltage comparator is show in Fig. 1. It consists of a preamplifier, a core latching comparator and an output buffer. The input analog signals are fed to the pre-amplifier. The output of the pre-amplifier is then fed to the latching comparator which produces the digital decision. The *2-φ* clock controls the operation of the latch. After the latch, a differential amplifier followed with two CMOS inverters is placed as output buffer to make the output level approach digital level as possible. As a whole, the design objective was to achieve the fastest possible operating speed without requiring excessive amounts of power and chip area.

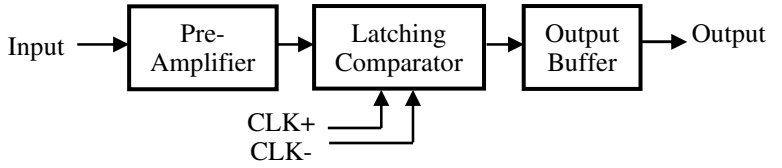


Fig. 1. Block diagram of the proposed comparator

The minimum gain expected of the pre-amplifier is determined by the offset voltage and noise of latch. As the same, the required performance feature of output buffer is also dependent on the latch's output voltage level. Thus the latch is the most crucial part in comparator design.

2.2 Latching Comparator

As illustrated in Fig. 2, the latching comparator is formed with two cross-coupled differential pairs (Q_1~Q_4) activated by *2-φ* current switching clocks. To improve the

operating speed of the latch and reduce the power consumption, two small MOSFETs (M_1-M_2) is used to act as both current bias and clock-controlled switches [1]. With the timing control signals from 2-φ clock, the latch alternates between the tracking mode and the latching mode by switching on *clk* and $\overline{clk}$ signals, respectively.

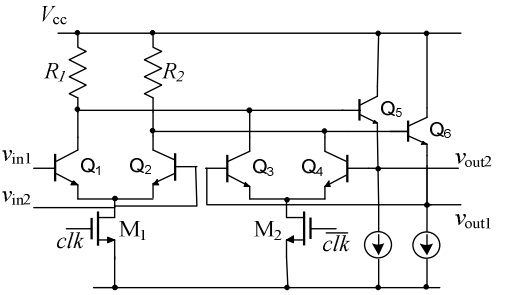


Fig. 2. Schematic of Latching Comparator

When *clk* is in high level namely the latch operates in tracking phase, M_1 is on and the latch tracks the amplified input signal available at pre-amplifier output. Meanwhile, M_2 is off and Q_3~Q_6 not work. When *clk* turns to low level, M_1 shuts off while M_2 turns on, the latch runs in regenerative mode and produces digital decision in high speed as possible. During this period the pre-amplifier also has enough time for overdrive recovery and re-setup before the next tracking phase.

It has been reported [7] that the maximum sampling rate of the latch is determined by the recovery time t_{rec}, which is defined as the time required by the gate differential pair's output voltage to change from a stable logic state to the midpoint of the logic state when the latch is switching from the latching mode to tracking mode. The recovery time can be expressed as [8]

$$t_{rec} = R_L C_{total}\left(1+\frac{1}{\tanh(\triangle V_{in}/2V_T)}\right) \tag{1}$$

where R_L is the load resistance, C_{total} is the total capacitance at collector of differential pair (Q_1-Q_2), ΔV_{in} is the differential input voltage of the latch, and V_T is the thermal voltage. As well, the expression of regeneration time constant τ_{reg}, another dominating factor, is [8]

$$\tau_{reg} = \frac{R_L C_{total}}{g_m R_L - 1} \tag{2}$$

where g_m is the transconductance of the HBTs in the second differential pair (Q_3-Q_4). According to the two expressions, a small R_L will shorten the recovery time but limit the output swing, thus degrade the latch's operation. On the contrary, a large R_L can bring a considerable output swing at the cost of increasing recovery time [9]. As the same, a small size of Q_3~Q_6 means small C_{total} and will reduce recovery time, whereas

results in small gain [1]. Thus a trade-off should be taken in choosing a suitable R_L and the size of Q_3~Q_6. The simulation result shows that, with an eclectic R_L and a NMOS switch, the recovery time is brought down to 17.9 ps, which is sufficient for operation at a 12.5 GHz sampling rate.

2.3 Pre-amplifier

Pre-amplifier is introduced to improve sensitivity to small input signals by suppressing the kick-back noise created by the switching of the latch and decreasing the offset voltage due to device mismatch of the following stages. The gain expected of the pre-amplifier is determined by the substantial offset voltage and noise accumulated by the latch. With higher differential voltage to the CMOS latch, the regeneration time of the latch can be reduced and accordingly the conversion speed can be improved.

The schematic of pre-amplifier is shown in Fig. 3. Two different amplifiers in cascade are used to produce a high GBW with a gain distribution of ~11dB and ~26dB. To give a faster response and shorter recovery time, the gain of the first stage is designed lower than the second one. With two stages, the pre-amplifier can achieve not only high gain, but also a suitable bandwidth and a small input capacitance which is expected at high operating speed. To mitigate the thermal mismatch of the input differential pair (*Q1*-*Q2* and *Q5*-*Q6*) that may cause undesirably large input offset, a large emitter size and low current density were desired for the input transistor. A larger emitter size also results in a better physical matching of the transistor pair. However, larger emitter size means lower conversion speed. Thus a compromise should be made with the size of *Q1*-*Q2* and *Q5*-*Q6*. Simulation results show the pre-amplifier delivering over ~36dB of gain and over ~20 GHz of -3dB bandwidth.

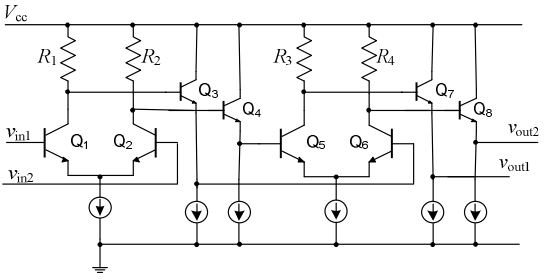


Fig. 3. Schematic of Two-stage Pre-Amplifier

2.4 Output Buffer

From the output of the latch to the digital encoder of the comparator an output buffer is required. As is shown in Fig. 4, a simple output buffer made up of a differential amplifier followed by two CMOS inverters is presented. To ensure the gain of the buffer and mitigate the device self-heating, a large tail current and a large emitter size are provided for the differential amplifier. With two inverters placed at the end, the output buffer can produce a well-defined digital data for other following stages of the ADC.

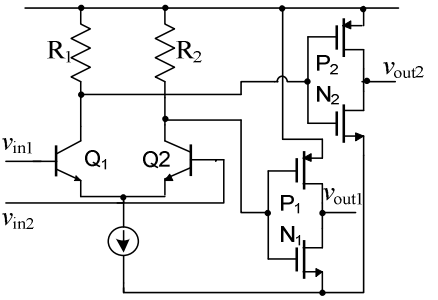


Fig. 4. Schematic of Two-stage Pre-Amplifier

3 Layout

The proposed comparator is laid out in double poly four metal 55 GHz, 0.35 µm BiCMOS technology. The use of differential signal in both the data and the clock paths leads to a fully symmetrical topology. To have better match between the devices, common centroid layout technique is adopted. Some dummy devices are placed around to provide an identical environment to reduce the random offsets caused by mismatch. The micrograph of the comparator is shown in Fig. 5 which takes an active area of 240×65µm^2.

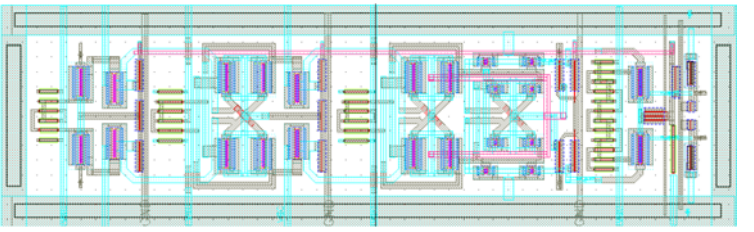


Fig. 5. Layout of the Comparator

4 Simulation Results

The simulation results, using SPECTRE (Cadence 5.1), show that, when operating of a 3.3V power supply, the comparator dissipates a total power of 54.16 mW. With sine wave input and clock, the comparator was simulated under two differential conditions: (1) a 1 GHz input oversampled up to 12.5 GS/s; and (2) full Nyquist with sampling rate up to 12.5 GS/s. The output waveforms of the first condition, shown in Fig. 6, measured a rise/fall time (20%~80%) of 30.2/32.7 ps. The output waveforms of the second condition, shown in Fig. 7, measured a rise/fall time of 90/79.2 ps.

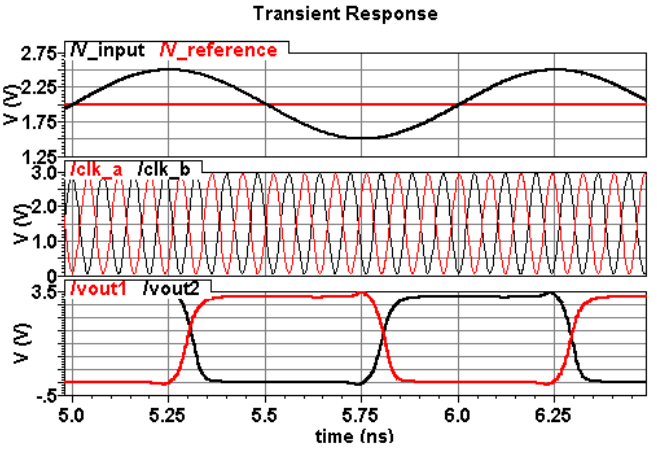


Fig. 6. Output waveforms of the Comparator with 1GHz@12.5GS/s

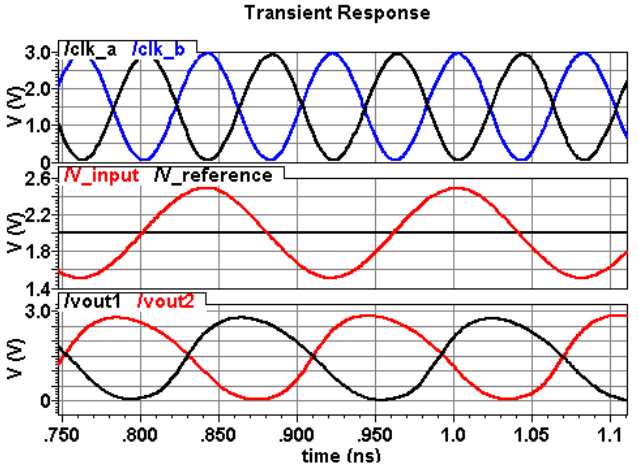
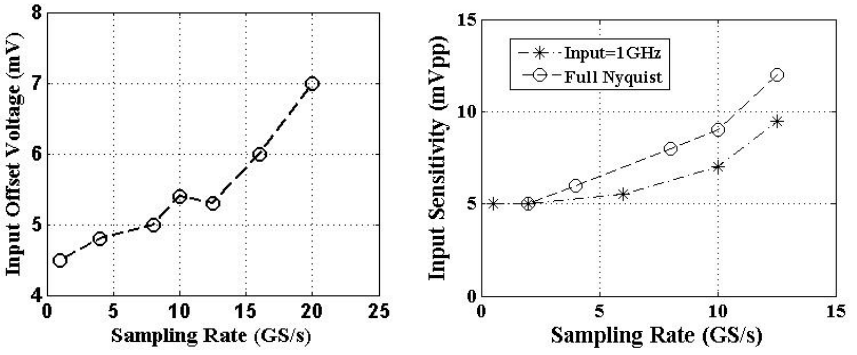


Fig. 7. Output waveforms of the Comparator with 6.25GHz@12.5GS/s

The measured input offset voltage of the comparator, caused by device mismatch, as a function of sampling rate with dc input is shown in Fig. 8 (a). The offset remains less than 5 mv when sampled below 8 GS/s, reaches a 6 mv at 16 GS/s.

Fig. 8 (b) shows the input sensitivity (the minimum differential peak-to-peak input voltage that can be detected) of the comparator as a function of sampling rate operating with a 1 GHz input and at Nyquist. With the 1 GHz input, the measured sensitivity ranges from 5.2 mV_{pp} at 5 GS/s to 9 mV_{pp} at 12.5 GS/s, equivalent to 7.0 bits and 6.0 bits of ADC resolution, respectively, for the full input range of 1 V_{pp} . At Nyquist, the sensitivity ranges from 6 mV_{pp} at 5 GS/s to 13 mV_{pp} at 12.5 GS/s, equivalent to 7.0 bits and 6.0 bits of ADC resolution, respectively. This demonstrates that the comparator is well-suited for high-speed medium-resolution flash ADCs. The performance of the comparator is summarized in Table 1. Comparisons of this work to other high-speed comparators are shown in Table 2.



(a) (b)

Fig. 8. (a)Input offset voltage of the comparator as a function of sampling rate; (b) input sensitivity of the comparator as a function of sampling rate

Table 1. Performance Summary of the Comparator

Parameters	Value
Supply voltage	3.3 V
Power Consumption	54.16 mW
Input Range	1 V_{pp} max
Sampling Rate	12.5 GS/s max
Input Offset Voltage	<5.3 mV up to12.5 GS/s
Input Sensitivity	<4 mv (in/clk=1 GHz / 12.5 GS/s)

Table 2. Performance Comparison with Published Comparators with Similar Sampling Rate

Refere -nce	Maximum Sampling rate [GS/s]	Sensitivity@ Input/Clock [@GHz/GS/s]	Supply Voltage [V]	Total Power Consumption [mW]	SiGe Process/fT [-/GHz]	Active Area [mm^2]
[2]	12.5	4mv@3.125/ 12.5	3.3	150	0.25µm HBT/190	0.0252
[3]	20	8.9mv@3/18	3.5	82	0.18µm HBT/120	0.0455
[6]	5	<4mv@1/4	3	89.04	0.5µm HBT/55	0.0582
[10]	3.8	<2mv	3/2.5	11.5	0.25µm BiCMOS/80	0.0167
[11]	12.5	N/A	2	1.09	0.12µm BiCMOS	N/A
[12]	16	N/A	N/A	80	0.5µm HBT/55	0.096
This work	12.5	<4mv@1/12.5	3.3	54.16	0.35µm BiCMOS/55	0.0156

5 Conclusion

In this paper, a high-speed comparator has been designed and implemented in a 55 GHz 0.35μm SiGe BiCMOS technology. The comparator can provides good performance when oversample a 1 GHz input up to 12.5 GS/s or sample at Nyquist up to 12.5 GS/s, while attaining 6.0 to 7.0 bits of ADC resolution. It can be opined that this design achieves a considerable trade-off between power, speed and resolution.

References

1. Liu, H., Wang, Z., Meng, Q., Tang, K.: A Bias-Switch-Multiplexing Comparator for the ADC used in SWR. In: WCSP, pp. 1–4 (2009)
2. Ghetmiri, S., Salama, C.A.T.: Track-and-hold and comparator for a 12.5-GS/s 8-bit ADC. In: MWSCA, pp. 353–356 (2009)
3. Li, X., Kuo, W.-M.L., Lu, Y., Krithivasan, R., Chen, T., Cressler, J.D., Joseph, A.J.: A 7-bit 18-GHz SiGe HBT comparator for medium resolution A/D conversion. In: BIPOL, pp. 144–147 (2005)
4. Banihashemi, M.: A high-speed high-resolution comparator. In: MWSCAS 2004, vol. 1, p. I-81-4 (2004)
5. Stefanou, N., Sonkusale, S.R.: An average low offset comparator for 1.25-GSample/s ADC in 0.18 μm CMOS. In: ICECS, pp. 246–249 (2004)
6. Gao, W., Snelgrove, W.M., Kovacic, S.J.: A 5-GHz SiGe HBT return-to-zero comparator for RF A/D conversion. In: JSSC, vol. 31, pp. 1502–1506 (1996)
7. Lim, P.J., Wooley, B.A.: An 8-bit 200-MHz BiCMOS comparator. In: JSSC, vol. 25, pp. 192–199 (1990)
8. Kuo, W.-M.L., Li, X., Krithivasan, R., Lu, Y., Cressler, J.D., Borokhovych, Y., Gustat, H., Tillack, B., Heinemann, B.: A 32-GSample/sec SiGe HBT comparator for ultra-high-speed analog-to-digital conversion. In: APMC, vol. 1, pp. 1–4 (2005)
9. Vessal, F., Salama, C.A.T.: An 8-bit 2-GSample/s analog-to-digital converter in 0.5μm SiGe technology. In: ISCAS, vol. (1), pp. 893–896 (2003)
10. Dey, S.K., Banerjee, S.: An 8-bit 3.8GHz dynamic BiCMOS comparator for high-performance ADC. In: VLSD (2006)
11. Marvast, M.J.T., Ali, M.A.M.: A 4.23-bit, 12.5-GS/s comparator for high speed flash ADC in BiCMOS technology. In: ICIAS, pp. 1–4 (2010)
12. Jensen, J.C., Larson, L.E.: A 16GHz ultra-high-speed Si/SiGe HBT comparator. In: BIPOL, pp. 120–123 (2002)

An Improved Fabric Simulation Algorithm*

Wenqing Huang, Wenjie Li, and Yuehong Sheng

Department of Zhejiang Sci-Tech University
{patternrecog,liwenjie0918,yue765}@163.com

Abstract. This paper puts forward a fabric simulation algorithm. The fabric simulation algorithm based on the improved spring-mass model. Firstly, the rigid rod is introduced to replace the structure spring, which is to improve the spring-mass model. Secondly, based on the improved spring-mass model, the fabric model is established .The fabric is composed of equally distributed particles. The particles in the model are force analyzed and the fabric mechanical equations are developed .Finally, the Display Euler equation method is used to solve the numerical values of the space displacements in the particle model. The hyper-elastic problem is perfected in the solving process. The experimental results show that the method in this paper can keep the fabric simulation stable and high efficiency, which realizes the efficient simulation to the fabric.

Keywords: spring-mass model, rigid rod, dynamic constraints, display Euler equation.

1 Introduction

Fabric is necessary in people's daily life as articles for daily use. It has always been the research focus that how to use the modern computer technology to simulate the fabric realistically. In recent ten or twenty years, scholars have been put forward many models, such as the geometry model[1], the physical model[2], the hybrid model[3]. The spring-mass model mentioned in this article is based on the physical model. Compared with other models the spring-mass model is wider used because of its structure which is simple to use, easy to realize optimization algorithm and high efficiency to calculate. But the model also has a defect: If the spring strength is too small, there is easy to appear the 'super elastic phenomenon'. Provot [4] and Bridson [5] implemented a series of correction algorithm in particle dormant. Because it is the local adjust and has influence to the length of around spring. The efficiency is low, and the constraints result is bad.

In order to achieve a better simulation result, the spring-mass model is improved in this paper. The rigid rod takes the place of the structure spring, which reduces the "super elastic phenomenon" and simplifies the calculation to some extent. The dynamic constraints are introduced into the spring shear and the spring bending. If the particle satisfied the restriction of the threshold in a step time is not found, we constrain the movement speed of the particle. This method can be a very good improvement to the "super flexible problem", and has a high efficiency.

* This project is the national natural science fund project support (funding number: 50875245) and Zhejiang natural science fund project support (funding number: Y107558) funded project.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 35–42.
springerlink.com

The content of this paper is as follows: the section 2 analyses the structure and stress of the improved spring-mass model, and the corresponding motion equations are established. The section 3 discusses how to use the explicit euler equation method to calculate the particle space displacement. The section 4 introduces the dynamic restraint method to improve the 'super elastic phenomena appear' in the fabric simulation. The section 5 gives the experiment results, and the results are cooperated and analyzed.

2 Improved Spring-Mass Model

2.1 Spring-Mass Model Structure

Breen, etc. think the fabric is the discrete body composed of particles and spring and put forward the model of particle fabric. Provot puts forward the particle discrete model, inheriting the assumption of the fabric discretization and linking up the particle with spring.

In the particle spring model, the fabric is taken as a $M \times N$ grid structure. The grid longitude intersection are shown with the particle, so that the quality of each grid unit is enriched into each particle, as figure1 (a) shown.

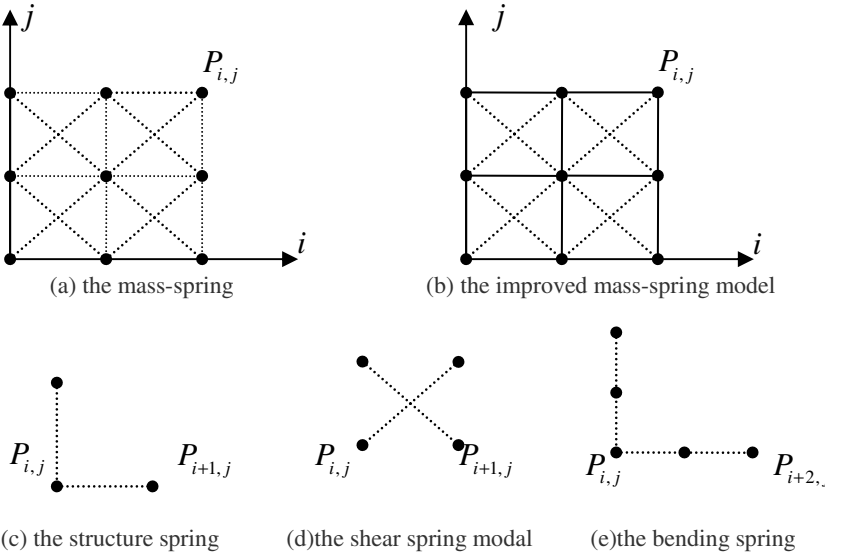


Fig. 1. The Spring-mass model structure of the fabric

The traditional spring-mass model is easy to emerge the 'super elastic phenomenon', and the calculation is great. Therefore, this paper introduces the rigid pole to replace the structure spring in the traditional particles spring, as figure1 (b) shows. The rigid pole is the pole that point does not deform when it is under the forces action .The rigid pole connects the particles through the smooth ball. When the fabric is impacted by the external force such as gravity, the tensile deformation is small.

2.2 Force Analysis

The fabric is a net composed of $M \times N$ particles. The particle motion is a dynamical system which changes with time. Based on Newton's mechanics formula:

$$F = m \cdot a \tag{1}$$

The dynamic equation of the particle *i* is got:

$$F(i) = F_e(i) + F_i(i) = ma(i) \tag{2}$$

in the formula: *m* is the mass of the particle *i*. $a(i)$ is the acceleration. $F_e(i)$ is the external force. $F_i(i)$ is the inner force, $F(i)$ is the sum of external force and inner force.

2.2.1 External Force

In the practical application, the external force of the fabric is diverse, such as gravity and wind-force, etc. In particular environment, the gravity can not be neglected, but others can be left out. Therefore, in this paper the motion is considered only forced by the gravity. This time, the external force is expressed as:

$$F_e(i) = F_{gravity}(i) = m(i) \cdot g \tag{3}$$

in the formula: *m* is the mass of the particle *i*; g *is the* acceleration of gravity, and its value is 9.8*N/kg*

2.2.2 Inner Force

The rigid pole role is introduced to replace the structure spring in the original model, because the structure spring produces a negligible interaction for the simulate effect. And in view of the acting force from the rigid pole is not obvious, the force between particles can be ignored.

The mainly internal force of the particle is the elastic deformation force. The elastic deformation force in this paper mainly refers to the shear spring deformation force and the bending deformation force spring.

The internal force of the particles is:

$$F_i(i) = \sum_{i=0}^{k} (f + F_{i,j}^d) \tag{4}$$

The spring elastic force in the ideal spring-mass model is calculated by the Hooke's Law.

$$f = \sum_{j \in R} k(|x_0 x_j|_t - |x_0 x_j|_0) \frac{x_0 x_j}{|x_0 x_j|} \tag{5}$$

in the formula: k is the elastic deformation coefficient. The parameters curve is established by the material performance of the fabric. $|x_0 x_j|_t$ is the distance between the particle x_0 and x_j at *t* . $|x_0 x_j|_0$ is the initial distance between the x_0 and x_j at the original.

The damping force between *i* and *j* is (the acting force to *i*):

$$F_{i,j}^{d} = K_{i,j}^{d}(V_j - V_i) \tag{6}$$

in the formula: $K_{i,j}^{d}$ is damping coefficient between the two particles. V_j is the speed of *j* and the V_i is the *i*'s.

3 Numerical Solve of the Dynamic Equation

The particle position in the next moment is determined in order to realize the fabric simulation. The particle position in the next moment is the key. Because the *a(i)* at the right of the formula (2) is the second order derivatives of the *i's* position vector, the unknown number *X* can be moved to the one side of the equation:

$$\ddot{X}_i = \frac{F_i}{m_i} \tag{7}$$

in the formula: m_i is a $3n*3n$ quality diagonal matrix(the diagonal elements is the single value of particle), F_i (it is column vector which includes $3n$ component vector)is the resultant force vector of the particle. n is the particle number system.

The display euler equation is employed to solve the formula (4). From mathematical knowledge, we can know $v = x'$, $a = v_0'$.For the initial value, the time step size is *h,* the position vector at time *t* is X_t . V_t is the speed vector, and the particle state at time $t+h$ is:

$$V_{t+h} = V_t + H\frac{F}{M} \tag{8}$$

$$X_{t+h} = X_t + HV_{t+h} \tag{9}$$

in the formula: *H* is a $3n*3n$ a time step size diagonal matrix(the diagonal elements is the single value of particle).when the time step is small enough and the overall particles are updated in every step, the simulation is realized.

Compared with the implicit integrating algorithm, the biggest strength of the display Euler integrating algorithm is that the calculation is small. The particle position vector can be got based on twice integration. But there are faults that the spring elastic coefficient can not be not more than a certain threshold. When the operator wants to do the simulation, he should select appropriate elastic coefficient of the spring.

4 Dynamic Constraint

If the length of the spring excesses the given constraint length it excesses, the dynamic constraint equations are used to amend, which can change the coefficient of elasticity. Orit goes on the next judgment.

When the spring endures more than its the biggest bear range, the elastic coefficient becomes into infinity. It can use a mathematical formula for this:

$$k = \begin{cases} \infty & L > L_c \\ K_0 & 0 \le L \le L_c \end{cases} \tag{10}$$

in the formula: k is the spring elasticity coefficient. κ_0 is the constant. L is the actual length of the spring. L_c is the constraint length of the spring.

When the spring length is out of the constraint length, the dynamic constraint as follows is taken. The time step is Δt. The current time is t_0.The position of the particle at t_0 and the next time $t_0 + \Delta t$ are $q_1(t_0)$ and $q_1(t_0 + \Delta t)$. The origin is at constant speed, and the speed can be described as:

$$v = \frac{\overline{q_1(t_0)q_1(t_0 + \Delta t)}}{\Delta t} \tag{11}$$

The dynamic constraint is: from the initial position, along the movement direction $q_1(t_0)q_1(t_0+\Delta t)$, the first particle unsatisfied the formula (11) is taken as the end position.and the speed v at this time will be operated.

As figure (2) shows, when the speed is out of the speed threshold, taking the particle q_1 as an example, the speed v_1 is resulted in the $q_1(t_0)q_1(t_0+\Delta t)$ direction and the direction normal to $q_1(t_0)q_1(t_0+\Delta t)$.The component along the $q_1(t_0)q_1(t_0+\Delta t)$ direction will make the spring excessive tensile deformation. The aim to updating the speed constrained is to decrease the separation speed of the two particles, which makes the two particles got closer to prevent the jitter. The real simulation of fabric is implemented.

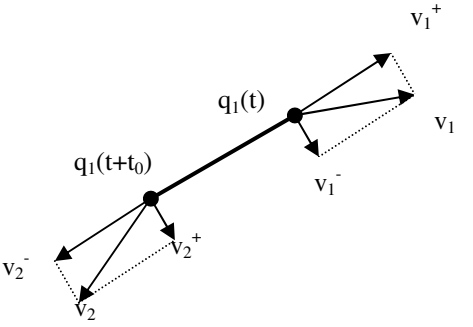


Fig. 2. The Speed constraint

5 Experimental Result

In order to realize the algorithm in this paper, the fabric simulation system is designed based on the VC++6.0, c++ and OpenGL. The algorithm process is as figure 3.

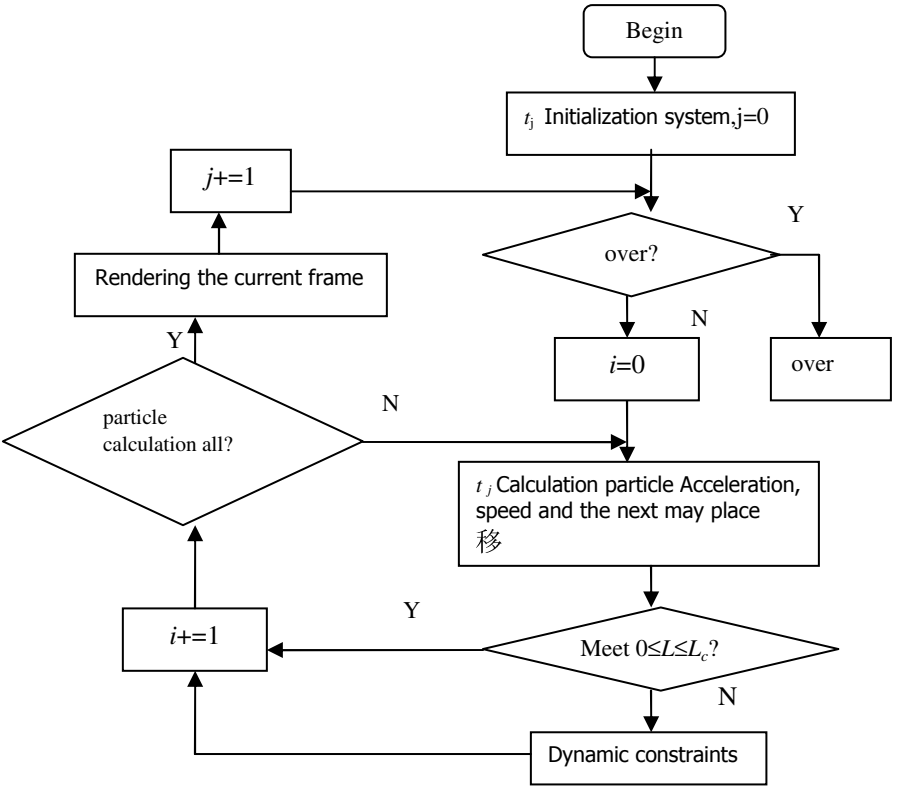
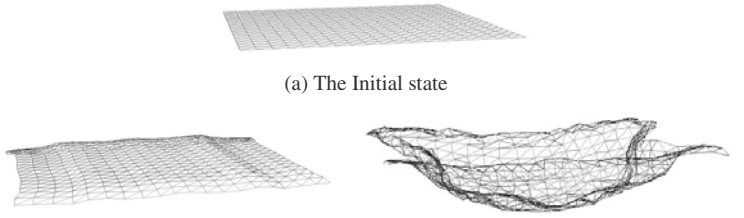


Fig. 3. Program realization process

In this paper, the fabric motion is simulated on the free falling under the gravity action and when two points suspended. Every effect is compared with the improvement mimic diagram and the strengths in this paper are conspicuous.figure4, figure5, figure6 give out the simulated comparison results of the fabric motion.

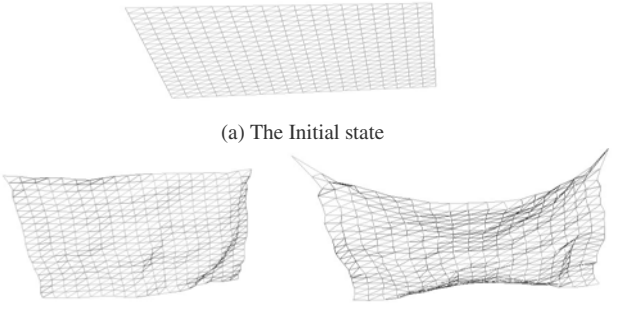
The 'hyper elastic phenomenon' exists in the improvement spring-mass model, which makes the simulation effect unsatisfactory. Based on improved algorithm in this paper, the 'hyper elastic phenomenon' has an obvious development. From these pictures, we can see that the algorithm in this paper has a better simulation results to the fabric motion.



(a) The Initial state

(b) Results of this Paper Method (c) The results without improvement

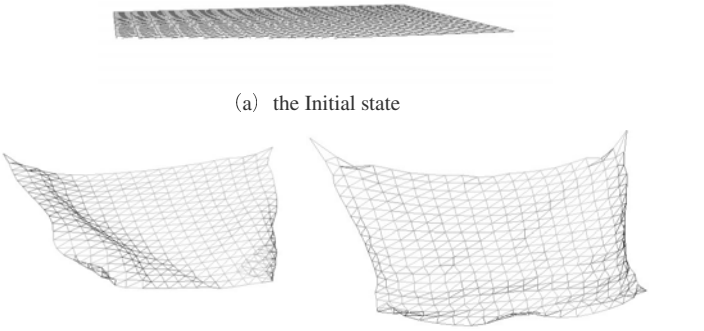
Fig. 4. The fabric sports simulated pictures in free fall



(a) The Initial state

(b) Results of this Paper Method (c) The results without improvement

Fig. 5. The fabric sports simulated pictures in case of two points hung vertically



(a) the Initial state

(b) Results of this Paper Method (c) The results without improvement

Fig. 6. The fabric sports simulated pictures in case of two points hung horizontally

6 Conclusion

In this paper, the classic spring-mass model is improved. The rigid pole is introduced to take the place of the structure spring in classic model, which simplifies the calculation and raises the program execution rate. Based on the force analysis of the

improved spring-mass model, mechanics equations are built. And the numerical solution of the model space through the display Euler equations. By using the dynamic constraint method, the 'hyperelastic phenomenon' appearing in the simulation has a good solve. The fabric simulation method in this paper will have a good application prospect in the virtual reality, the animation production areas, and the game production areas.

References

1. Feynman, C.R.: Modeling the Appearance of Fabric, M.Sc.Thesis, Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA (1986)
2. Baraff, D., Witkin, A.: Large Steps in Fabric Simulation. In: Computer Graphics (Proc. SIGGRAH), pp. 43–54 (1998)
3. Kunii, T.L., Gotoda, H.: Singularity Theoretical Modeling and Animation of Garment Wrinkle Formation Processes. The Visual Computer (6), 326–336 (1990)
4. Provot, X.: Deformation constraints in a mass-spring model to describe rigid fabric behavior. In: Proceedings of Graphics Interface, pp. 147–154 (1995)
5. Bridson, R., Fedkiw, R., Anderson, J.: Robust treatment of collisions, contact and friction for fabric animation. ACM Transactions on Graphics (ACM SIGGRAPH) (S0730-0301) 21(3), 594–603 (2002)
6. Baraff, D., Witkin, A.P.: Large steps in fabric simulation. In: Proceedings of SIGGRAPH 1998, Computer Graphics Proceedings, Annual Conference Series, pp. 43–54 (1998)
7. Xiu, Y., Li, C.: Fabric three dimensional simulation based on the particle-rigid pole model. Textile Journal 28(10), 113–116 (2007)
8. Lv, M., Li, F., Tang, Y., Bi, W.: Rapid lifelike fabric simulation based on the spring-mass model. System Simulation Journal 21(16), 236–5239 (2009)
9. Shen, Z., Pan, Z.: Simulation and collision of fabric processing method based on the spring-mass model. The Computer Simulation 23(3), 284–287 (2006)
10. Hao, L.: Mass-spring-based three-dimensional cloth simulation system. Shanghai Jiao Tong University, Shanghai (2007) (in Chinese)

Parameter Estimation for Dual-Rate Sampled Data Systems with Preload Nonlinearities

Chen Jing[1,2], Lv Lixing[1], and Ding Ruifeng[2,*]

[1] Wuxi Professional College of Science and Technology, Wuxi\ 214028, P.R. China
[2] School of Internet of Things Engineering, Jiangnan University, Wuxi\ 214122, P.R. China
rfding@yahoo.cn

Abstract. In this paper, we propose a novel estimation algorithm for a dual-rate system with preload nonlinearity. The input-output data are measured two different sampling rates. A switching function and a polynomial transformation technique are employed to derive a mathematical model for such a dual-rate and nonlinear system. Furthermore, two modified stochastic gradient algorithms are given to improve the convergence rate. Finally a simulation example is provided to verify the effectiveness of the proposed method.

Keywords: Parameter estimation, Stochastic gradient algorithm, Preload nonlinearity, Dual-rate system, Convergence rate.

1 Introduction

Dual-rate/multirate sampled-data systems, in which the input and the output signals have different sampled rates, are abundant in industrial processes [1-3]. Two or more than two operating frequencies appear in systems due to sensor and actuator speed constraints. For example, in polymer reactors and petroleum production, the composition and gasoline octane quality measurements are typically obtained after several minutes of analysis, but the manipulated variables can be adjusted at relatively fast rate [4,5].

Recently, dual-rate/multirate sampled-data system identification has received much attention and a large number of publications discuss the identification problems of the dual-rate/multirate sampled data systems [6-14]. For example, Ding proposed a modified stochastic algorithm for a class of dual-rate sampled data system by using a polynomial transformation technique [6] and an extended stochastic gradient algorithm for auto-regression times series models with missing observations [7]. Ding derived a mathematical model for the dual-rate systems by using a polynomial transformation technique and presented a recursive least squares algorithm to estimate all the parameters of the systems [10].

Hard-input nonlinearities severely limit the performance of control systems. The characteristic of the hard-input nonlinearity is that the parameters of the nonlinear part are coupled with the linear part. Moreover, the output of the hard nonlinear block may

* Corresponding author.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 43–50.
springerlink.com

not be written as an analytic function of the input. In the field of system identification, Bai used a deterministic correlation analysis method to estimate the parameters of systems with hard input nonlinearity [15]. Voros used an appropriate switching function to model and identify a Hammerstein system with multisegment piecewise-linear characteristics [16] and with backlash [17]. However, to the best of our knowledge, there exists only scattered work reported in the literature about the identification of the dual-rate sampled data systems with hard input nonlinearity, especially the dual-rate sampled data systems with the preload hard nonlinearity, which are the focus of this paper.

The objective of this paper is in the following ways:

1. By using a switching function and a polynomial transformation technique, derive a model suitable for dual-rate identification of the preload nonlinearity system.
2. Based on the derived model, propose a stochastic gradient (SG) algorithm and two modified SG algorithms.

Briefly, the paper is organized as follows. Section 2 describes the problem formulation of the dual-rate system with preload nonlinearity and derives a suitable model. Section 3 presents a SG algorithm and two modified SG algorithms for the suitable model. Section 4 provides an illustrative example. Finally, concluding remarks are given in Section 5.

2 Problem Formulation

Consider a system with a preload nonlinearity described by the following model

$$A(z)y(t) = B(z)f(u(t)), \tag{1}$$

where $y(t)$ is the system output and $f(u(t))$ is the system input, and $A(z)$ and $B(z)$ are polynomials in the unit backward shift operator [$z^{-1}y(t) = y(t-1)$] and

$$A(z) := 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_n z^{-n},$$
$$B(z) := b_1 z^{-1} + b_2 z^{-2} + b_3 z^{-3} + \cdots + b_n z^{-n}.$$

The preload nonlinearity in Figure 1 can be expressed as

$$f(u(t)) = \begin{cases} u(t) + m, & u(t) > 0, \\ 0, & u(t) = 0, \\ u(t) - m, & u(t) < 0, \end{cases}$$

where m and $-m$ are the preload points.

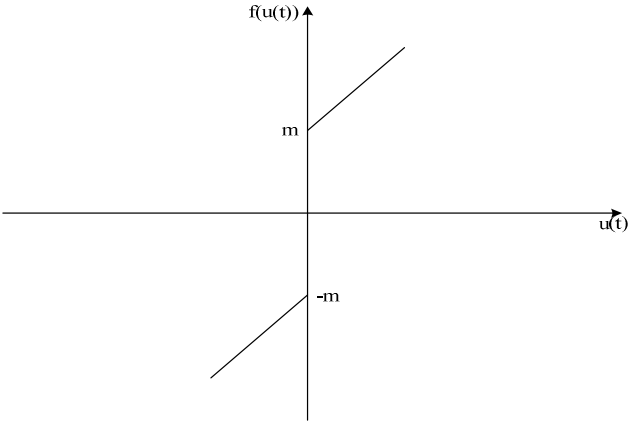


Fig. 1. The preload nonlinearity

For the dual-rate sampled data systems, we have all input data $u(t), t = 0,1,2,\cdots,$ and scarce output data, $y(tq),\ q \geq 2, t = 0,1,2,\cdots,$ the intersample output or missing outputs $y(tq+j),\ j = 1,2,\cdots,q-1$ are 2 unavailable. In this paper we use the polynomial transformation technique to derive a mathematical model of the dual-rate sampled data system.

Let the roots of $A(z)$ be $z_i, i = 1,2,\cdots,n$ to get

$$A(z) = (1 - z_1 z^{-1})(1 - z_2 z^{-1}) \cdots (1 - z_n z^{-1}).$$

Define a polynomial

$$\begin{aligned} r(z) &:= \prod_{i=1}^{n} (1 + z_i z^{-1} + z_i^2 z^{-2} + \cdots + z_i^{q-1} z^{-q+1}) \\ &= 1 + r_1 z^{-1} + r_2 z^{-2} + \cdots + r_m z^{-m}, \quad m := n(q-1). \end{aligned} \tag{2}$$

Multiplying both sides of (1) by $r(z)$ yields

$$\alpha(z) y(t) = \beta(z) f(u(t)), \tag{3}$$

where

$$\alpha(z) := r(z) A(z) = 1 + \alpha_1 z^{-q} + \alpha_2 z^{-2q} + \cdots + \alpha_n z^{-nq}, \tag{4}$$

$$\beta(z) := r(z) B(z) = \beta_1 z^{-1} + \beta_2 z^{-2} + \beta_3 z^{-3} + \cdots + \beta_{nq} z^{-nq}. \tag{5}$$

From (4) and (5), one can see that the model in (3) involves all input-output data $\{u(t), y(tq) : t = 0,1,2,\cdots\}$.

Without loss of generality, a noise term $v(t)$ with zero-mean is introduced to the model in (3):

$$\alpha(z)y(t)=\beta(z)f(u(t))+v(t). \tag{6}$$

To simply the preload nonlinearity, we introduce a sign function,

$$\operatorname{sgn}[u(t)]=\begin{cases}1, & if\ \ u(t)>0,\\ 0, & if\ \ u(t)=0,\\ -1, & if\ \ u(t)<0.\end{cases}$$

The preloa nonlinearity $f(u(t))$ can be written as

$$f(u(t))=u(t)+m\operatorname{sgn}(u(t)) \tag{7}$$

Then Equation (6) can be rewritten as

$$\alpha(z)y(t)=\beta(z)(u(t)+m\operatorname{sgn}(u(t)))+v(t). \tag{8}$$

This is an input-output representation of the input nonlinear systems.

3 The Dual-Rate Parameter Estimation Algorithm

Define the parameter vector θ and the information vector $\varphi(t)$ as

$$\theta:=[\beta_1,\beta_2,\cdots,\beta_{nq},\beta_1 m,\beta_2 m,\cdots,\beta_{nq}m,\alpha_1,\alpha_2\cdots,\alpha_n]^{\mathrm{T}}\in R^{2nq+n}, \tag{9}$$

$$\begin{aligned}\varphi(t):=[&u(t-1),u(t-2),\cdots,u(t-nq),\operatorname{sgn}(u(t-1)),\operatorname{sgn}(u(t-2),\cdots\\ &\operatorname{sgn}(u(t-nq)),y(t-q),y(t-2q),\cdots,y(t-nq)]^{\mathrm{T}}\in R^{2nq+n}.\end{aligned} \tag{10}$$

Equations (8) can be written as

$$y(t)=\varphi^T(t)\theta+v(t).$$

When t is an integer multiple of q, then $\varphi(t)$ contains only the available measurement outputs and inputs. Replacing t with qt gives

$$y(tq)=\varphi^T(tq)\theta+v(tq), \tag{11}$$

this is the identification model of the input nonlinear systems in (1).

Let $\hat{\theta}(tq)$ be the estimate of θ at time tq. The following SG algorithm can estimate the parameter vector θ in (9):

$$\hat{\theta}(tq)=\hat{\theta}(tq-q)+\frac{\varphi(tq)}{r(tq)}[y(tq)-\varphi^{\mathrm{T}}(tq)\hat{\theta}(tq-q)], \tag{12}$$

$$\hat{\theta}(tq+i)=\hat{\theta}(tq),\quad i=0,1,2,\cdots,q-1, \tag{13}$$

$$r(tq)=r(tq-q)+\|\varphi(tq)\|^2,\quad r(0)=1, \tag{14}$$

where $\frac{1}{r(tq)}$ is the step-size and the norm of matrix X is defined by $\|X\|^2=\mathrm{tr}[XX^{\mathrm{T}}]$.

The convergence of the SG algorithm is relatively slower compared with the recursive least squares algorithm. In order to improve the tracking performance of the SG algorithm, we introduce a forgetting factor λ in the SG algorithm to get the SG algorithm with a forgetting factor (the FFSG algorithm for short) as follows:

$$\hat{\theta}(tq)=\hat{\theta}(tq-q)+\frac{\varphi(tq)}{r(tq)}[y(tq)-\varphi^T(tq)\hat{\theta}(tq-q)], \tag{15}$$

$$\hat{\theta}(tq+i)=\hat{\theta}(tq),\quad i=0,1,2,\cdots,q-1, \tag{16}$$

$$r(tq)=\lambda r(tq-q)+\|\varphi(tq)\|^2,\quad 0<\lambda<1,\quad r(0)=1. \tag{17}$$

We can also introduce a convergence index ε in the SG algorithm to obtain a modified SG algorithm (the M-SG algorithm for short):

$$\hat{\theta}(tq)=\hat{\theta}(tq-q)+\frac{\varphi(tq)}{r^{\varepsilon}(tq)}[y(tq)-\varphi^{\mathrm{T}}(tq)\hat{\theta}(tq-q)],\quad \frac{1}{2}<\varepsilon\le 1, \tag{18}$$

$$\hat{\theta}(tq+i)=\hat{\theta}(tq),\quad i=0,1,2,\cdots,q-1, \tag{19}$$

$$r(tq)=r(tq-q)+\|\varphi(tq)\|^2,\quad r(0)=1. \tag{20}$$

The initial value is generally taken to be $\theta(0)=$ a small real vector.

4 Example

Consider the following system,

$$\begin{aligned}
&A(z^{-1})y(t)=B(z^{-1})f(u(t))+v(t),\\
&A(z^{-1})=1+0.60z^{-1}+0.80z^{-2},\\
&B(z^{-1})=0.60z^{-1}+0.40z^{-2},\\
&f(u(t))=u(t)+0.2\,\mathrm{sgn}(u(t)).
\end{aligned}$$

The input $\{u(t)\}$ is taken as a persistent excitation signal sequence with zero mean and unit variance, and $\{v(t)\}$ is taken as a white noise sequence with zero mean and variance $\sigma^2 = 0.10^2$. Here, we take $q = 2$ and $r(z) = 1 - 0.60q^{-1} + 0.80q^{-2}$ thus we have

$$\alpha(z) = r(z)A(z) = 1 + 1.24z^{-2} + 0.64z^{-4},$$
$$\beta(z) = r(z)B(z) = 0.60z^{-1} + 0.04z^{-2} + 0.24z^{-3} + 0.32z^{-4},$$
$$\theta = [1.24, 0.64, 0.60, 0.04, 0.24, 0.32, 0.12, 0.08, 0.048, 0.064]^{\mathrm{T}}.$$

Applying the SG algorithm, the M-SG algorithm and the FFSG algorithm to estimate the parameters of this system, the parameter estimates and their errors are shown in Tables 1-3, the parameter estimation errors $\delta := \| \hat{\theta} - \theta \| / \| \theta \|$ versus t are shown in Figure 2

Table 1. The SG estimates and errors

t	b_1	b_2	a_1	a_2	a_3	a_4	$a_1 m$	$a_2 m$	$a_3 m$	$a_4 m$	δ (%)
100	0.36091	-0.03768	0.39537	0.09481	-0.03688	0.08948	0.33171	-0.12272	-0.10070	0.27772	78.55713
200	0.38226	-0.02826	0.39742	0.10262	-0.05303	0.08934	0.33096	-0.11668	-0.11293	0.27449	77.53155
300	0.39261	-0.02359	0.39610	0.10541	-0.05676	0.08794	0.32624	-0.11188	-0.11495	0.26966	76.89102
500	0.40029	-0.01778	0.39375	0.10714	-0.06145	0.08582	0.32105	-0.10973	-0.11780	0.26583	76.39577
1000	0.41988	-0.00778	0.39527	0.11170	-0.07081	0.08628	0.31918	-0.10304	-0.12514	0.26236	75.30420
1500	0.42949	-0.00152	0.39712	0.11113	-0.07438	0.08734	0.31872	-0.10296	-0.12772	0.26182	74.70609
2000	0.43747	0.00142	0.39830	0.11126	-0.07717	0.08806	0.31817	-0.10129	-0.12984	0.26060	74.27497
2500	0.44236	0.00552	0.39940	0.11165	-0.07955	0.08898	0.31814	-0.10064	-0.13097	0.26041	73.94508
3000	0.44673	0.00848	0.40036	0.11119	-0.08091	0.08985	0.31800	-0.10067	-0.13161	0.26031	73.65830
True values	1.24000	0.64000	0.60000	0.04000	0.24000	0.32000	0.12000	0.00800	0.04800	0.06400	

Table 2. The M-SG with $\varepsilon = 0.51$ estimates and errors

t	b_1	b_2	a_1	a_2	a_3	a_4	$a_1 m$	$a_2 m$	$a_3 m$	$a_4 m$	δ (%)
100	0.73053	0.11220	0.49972	0.04855	-0.25820	0.25045	0.21525	0.03473	-0.15876	0.16166	58.88819
200	0.80789	0.26709	0.53606	0.06517	-0.11296	0.23717	0.22648	0.01584	-0.07676	0.16255	44.75201
300	0.86322	0.36566	0.52063	0.02437	0.03215	0.23382	0.20064	0.00874	0.00006	0.13615	34.03530
500	0.97693	0.43648	0.55059	0.05782	0.06939	0.24517	0.18345	0.01717	0.01910	0.12930	25.13617
1000	1.17067	0.58419	0.58524	0.04100	0.17334	0.29245	0.17352	0.01703	0.05203	0.11213	8.64856
1500	1.22174	0.62350	0.58505	0.04267	0.20968	0.29904	0.14635	0.01353	0.06083	0.10119	4.23606
2000	1.23700	0.63758	0.60446	0.04568	0.22321	0.30666	0.14002	0.01186	0.06055	0.09324	2.80236
2500	1.22987	0.63075	0.58947	0.03479	0.23014	0.30778	0.12743	0.01182	0.05396	0.08656	2.18337
3000	1.23510	0.63408	0.59486	0.03539	0.23485	0.31065	0.12866	0.00781	0.05457	0.08083	1.58068
True values	1.24000	0.64000	0.60000	0.04000	0.24000	0.32000	0.12000	0.00800	0.04800	0.06400	

From Tables 1-3 and Figure 2, we can get the following conclusions:

1 The convergence rate of the SG algorithm is slower than the M-SG algorithm and the FF-SG algorithm.

2. The parameter estimation errors become smaller as t increases.

Table 3. The FFSG with $\lambda = 0.85$ estimates and errors

t	b_1	b_2	a_1	a_2	a_3	a_4	a_1m	a_2m	a_3m	a_4m	δ (%)
100	0.49147	-0.02175	0.31086	-0.00130	-0.23086	0.16729	0.29433	0.01455	-0.23961	0.16617	76.37191
200	0.62764	0.14250	0.40633	0.05513	-0.17180	0.18435	0.31902	0.02785	-0.14664	0.17045	61.36174
300	0.70892	0.24394	0.40327	0.02102	-0.05501	0.17822	0.24721	0.01056	-0.00373	0.11341	49.35280
500	0.87127	0.34977	0.49308	0.06011	-0.01799	0.21182	0.20443	0.02290	0.01103	0.10659	35.89739
1000	1.14595	0.56546	0.56673	0.03695	0.12475	0.29153	0.18420	0.01842	0.05335	0.08915	11.77230
1500	1.22275	0.62282	0.58630	0.04434	0.19106	0.30174	0.12513	0.01712	0.05951	0.08137	4.04396
2000	1.23982	0.64415	0.62377	0.06007	0.21857	0.30906	0.13507	0.01011	0.05237	0.07666	2.81662
2500	1.22441	0.62294	0.59533	0.03343	0.23617	0.31288	0.10663	0.01335	0.03743	0.07008	2.02440
3000	1.23132	0.63091	0.59435	0.03473	0.23773	0.31559	0.12360	0.00527	0.04951	0.06913	1.08151
True values	1.24000	0.64000	0.60000	0.04000	0.24000	0.32000	0.12000	0.00800	0.04800	0.06400	

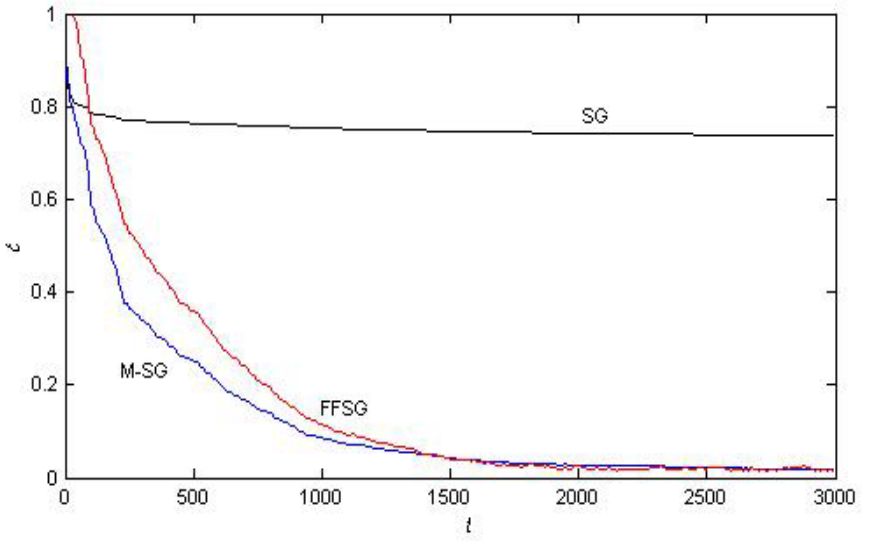


Fig. 2. The parameter estimation errors δ versus t

5 Conclusion

A method to identify a dual-rate system with preload nonlinearity is presented in this paper. By using an appropriate switching function and a polynomial transformation technique, the nonlinear system is described as an analytic form, then all the parameters can be estimated by using the SG algorithm. Furthermore an M-SG algorithm and an FFSG algorithm are given to improve the identification accuracy. The simulation results verify the proposed algorithms. We can also use the iterative algorithm as shown in [18,19] to estimate all the parameters of this dual-rate system with preload nonlinearity.

References

1. Abo-zahhad, M.: Current state and future directions of multirate filter banks and their applications. Digital Signal Processing 13(3), 495–518 (2003)
2. Shi, Y., Ding, F., Chen, T.: Multirate crosstalk identification in xDSL systems. IEEE Transactions on Communications 54(10), 1878–1886 (2006)
3. Kadu, S.C., Bhushan, M., Gudi, R.D.: Optimal sensor network design for multirate systems. Journal of Process Control 18(6), 594–609 (2008)
4. Gudi, R.D., Shah, S.L., Gray, M.R.: Multirate state and parameter estimation in an antibiotic fermentation with delayed measurements. Biotechnology and Bioengineering 44(12), 1271–1278 (1994)
5. Li, D., Shah, S.L., Chen, T., Qi, K.Z.: Application of dual-rate modeling to CCR octane quality inferential control. IEEE Transactions on Control Systems Technology 11(1), 45–51 (2003)
6. Ding, J., Shang, Y., Wang, H.G., Ding, F.: A modified stochastic gradient based parameter estimation algorithm for dual-rate sampled-data systems. Digital Signal Processing 20(4), 1238–1247 (2010)
7. Ding, J., Han, L.L., Chen, X.M.: Time series AR modeling with missing observations based on the polynomial transformation. Mathematical and Computer Modelling 51(5-6), 527–536 (2010)
8. Ding, J., Ding, F.: The residual based extended least squares identification method for dual-rate systems. Computers Mathematics with Applications 56(6), 1479–1487 (2008)
9. Ding, F., Liu, P.X., Yang, H.Z.: Parameter identification and intersample output estimation for dual-rate systems. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans 38(4), 966–975 (2008)
10. Ding, F., Liu, P.X., Shi, Y.: Convergence analysis of estimation algorithms of dual-rate stochastic systems. Applied Mathematics and Computation 176(1), 245–261 (2006)
11. Ding, F., Chen, T.: Combined parameter and output estimation of dual-rate systems using an auxiliary model. Automatica 40(10), 1739–1748 (2004)
12. Ding, F., Qiu, L., Chen, T.: Reconstruction of continuous-time systems from their non-uniformly sampled discrete-time systems. Automatica 45(2), 324–332 (2009)
13. Liu, Y.J., Xie, L., Ding, F.: An auxiliary model based recursive least squares parameter estimation algorithm for non-uniformly sampled multirate systems. Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering 223(4), 445–454 (2009)
14. Han, L.L., Sheng, J., Ding, F., Shi, Y.: Auxiliary model identification method for multirate multi-input systems based on least squares. Mathematical and Computer Modelling 50(7-8), 1100–1106 (2009)
15. Bai, E.W.: Identification of linear systems with hard input nonlinearities of known structure. Automatica 38(5), 853–860 (2002)
16. Voros, J.: Modeling and parameter identification of systems with multisegment piecewise-linear Characteristics. IEEE Transactions on Automatic Control 47(1), 184–188 (2002)
17. Voros, J.: Modeling and identification of systems with backlash. Automatica 46(2), 369–374 (2010)
18. Dehghan, M., Hajarian, M.: Two algorithms for finding the Hermitian reflexive and skew-Hermitian solutions of Sylvester matrix equations. Applied Mathematics Letters 24(4), 444–449 (2011)
19. Chen, J., Wang, X.P., Ding, R.F.: Gradient based estimation algorithm for Hammerstein systems with saturation and dead-zone nonlinearities. Applied Mathematical Modelling 36(1), 238–243 (2012)

Determination of Poisson's Ratio of Kraft Paper Using Digital Image Correlation

Xiaolong Cao, Zhongchen Bi, Xing Wei, and Yong Xie*

School of Packaging and Material Engineering, Hunan University of Technology, Zhuzhou Hunan 412008, P.R. China
hutxy@126.com

Abstract. Kraft paper is the most popular raw material for based-paper packaging containers. Poisson's ratio is an important index to indicate the inherent property of materials. There were many difficulties to measure Poisson's ratio of kraft paper using the traditional contract methods, because of its micro deformation and anisotropy. A simple and efficient non-contract method had been proposed to solve this problem using DIC(Digital Image Correlation) method in this paper. Obtained the relative deformation of samples by calibrated CCD images, Poisson's ratio could be computed. The test results indicated that Poisson's ratios of corrugating medium in the MD(Machine Direction) and CD (Cross-machine Direction) were 0.275 and 0.119, and ones of linerboard were 0.275 and 0.119 in MD and CD, respectively. This study showed that DIC was a new approach to measure Poisson's ratio of kraft paper.

Keywords: Digital image correlation, Poisson's ratio, Kraft paper, Calibration.

1 Introduction

Poisson's ratio is the inherent property of materials, reflecting material changes in stretching deformation. It is used in a wide variety of engineering applications as an indicator of material deformation properties. Poisson's ratio of metal materials is measured with the traditional contact-methods, such as a national standard GB/T 22315-2008 [1] or the electrometric method [2]. However, it is hard to measure Poison's ratios of the thin and softy kraft papers with the traditional contact-methods because deformations in the horizontal direction are too micro to be measured accurately when papers are stretched. The traditional contact-methods make them unapt for the application during tests. Consequently, an easy and efficient method has not been proposed to measure Poisson's ratios of kraft papers.

Digital image correlation method [3] provides a new non-contact method for the displacement and deformation measurements. Eitner et al. used DIC to detect tiny displacements and deformations of solar cells in transparent PV modules for changes of the temperature [4]. Yates et al. characterized crack tip displacement fields based on data obtained from DIC [5]. According to DIC, Sánchez-Arévalo, et al. calculated local

* Corresponding author.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 51–57.
springerlink.com

and global strain measurements to obtain the young modulus for materials, such as commercial aluminum and bovine pericardium [6]. As a result, DIC is characterized by global strain, high efficiency, accuracy, relatively simple application, etc.

In this article, it will be shown that DIC can be applied with success to obtain Poisson's ratios of kraft papers by measuring displacements from experiments.

2 Principles

2.1 Digital Image Correlation

DIC is an image processing methods that the gray values of the object on the surface are analyzed to calculate the deformation and displacement. It depends on principles of relativity in statistics. As shown in Fig. 1, image I and image J are gray images before and after papers are stretched, respectively. The point $P(x, y)$ and $Q(x', y')$ are the center of a subarea f in image I and g in image J, respectively. When f and g are performed with a relativity operation, if f is matched successfully with g, it shows that the position at point P moves to where point Q is after specimens are stretched. And $\boldsymbol{r}$ is the displacement vector of the two subareas. Assume that a matching image can be divided into n subareas in the horizontal direction and m subareas in the vertical direction, a matrix M and N can be given as,

$$M=[p_{ij}],\ N=[q_{ij}],\ i=1,2,\ldots,m,\ j=1,2,\ldots,n. \tag{1}$$

where, p_{ij} is the center point of the i^{th} row and j^{th} column subarea for image I, and q_{ij} is the center point for image J.

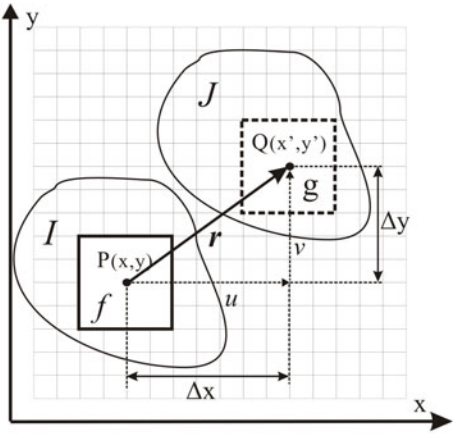


Fig. 1. Principle of DIC

The match can be performed as,

$$x' = x + u(x, y), \quad y' = y + v(x, y) \tag{2}$$

To find the displacement fields, the horizontal u, and vertical v, displacements can be approximated using the Taylor series around a point $P(x, y)$ as,

$$\begin{aligned} x' &\approx x + u + \frac{\partial u}{\partial x}\Delta x + \frac{\partial u}{\partial y}\Delta y + \frac{\partial^2 u}{2\partial^2 x}\Delta x^2 + \frac{\partial^2 u}{2\partial^2 y}\Delta y^2 \\ y' &\approx y + v + \frac{\partial v}{\partial x}\Delta x + \frac{\partial v}{\partial y}\Delta y + \frac{\partial^2 v}{2\partial^2 x}\Delta x^2 + \frac{\partial^2 v}{2\partial^2 y}\Delta y^2 \end{aligned} \tag{3}$$

in which $\Delta x = x' - x$ and $\Delta y = y' - y$.

Relativity operation needs to choose an appropriate function, so called correlation factor, C, defined as,

$$C = \frac{\sum\sum[f(x,y) - \overline{f}]\cdot[g(x',y') - \overline{g}]}{\sqrt{\sum\sum[f(x,y) - \overline{f}]^2 \cdot \sum\sum[g(x',y') - \overline{g}]^2}} \tag{4}$$

where $f(x, y)$ and $g(x', y')$ are gray values of subareas in image I and image J, respectively, $\overline{f}$ and $\overline{g}$ are the average of gray values. Generally, the closer to 1 the maximum value of C is, the better the match is done.

2.2 Calculation of Poisson's Ratio

DIC is used to obtain displacements represented by pixels in the horizontal and vertical direction. In the vertical direction, the average length before and after samples stretched are H and H' , given as,

$$H = \frac{1}{n}\sum_{j=1}^{n}[p_{mj}(y) - p_{1j}(y)], \quad H' = \frac{1}{n}\sum_{j=1}^{n}[q_{mj}(y') - q_{1j}(y')] \tag{5}$$

where $p_{mj}(y)$ is the y-coordinate of a point P at the final row of the matrix M. Consequently, displacements are $\Delta H = |H' - H|$. Likewise, in the horizontal direction, the average widths and displacements are V , V' and ΔV .

Poisson's ratio is an absolute value of the ratio of deformations in horizontal direction divided by ones in vertical direction, defined as,

$$\mu = \left|\left(\Delta d / d\right) / \left(\Delta l / l\right)\right| = \left|\frac{\Delta d}{\Delta l}\right| \cdot \frac{l}{d} \tag{6}$$

in which Δd and Δl are the displacements in the horizontal and vertical direction, d and l are the width and length of samples respectively.

2.3 Calibration

To examine the accuracy of DIC, a simple calibration method can be used. Laplacian of Gaussian (LOG) [7], as an edge detection method, can be used to process images. Laplacian is the second derivative operator, given as,

$$\nabla^2 f(x,y) = \frac{\partial^2 f(x,y)}{\partial x^2} + \frac{\partial^2 f(x,y)}{\partial y^2} \tag{7}$$

$G(x, y)$ is the two dimensional Gaussian function, given by,

$$G(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \tag{8}$$

Detecting the edge of the subject in the image $f(x, y)$ with LOG, the first step is to smooth the image $f(x, y)$ with Gaussian filter, and then the second derivative is done for the smoothed image with Laplacian. That is to say, LOG can be given as,

$$h(x,y) = \nabla^2(G(x,y) * f(x,y)) \tag{9}$$

where $h(x, y)$ is the binary image including the edge information of the subject, $*$ is the convolution operator. By the derivative rule for convolutions, the edge information can be detected by convolving the image $f(x, y)$ with the operator $\nabla^2 G(x, y)$,

$$h(x,y) = \nabla^2 G(x,y) * f(x,y) = \frac{1}{\pi\sigma^4}\left(\frac{x^2+y^2}{2\sigma^2} - 1\right)\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) * f(x,y) \tag{10}$$

Let w be a pixel distance from $h(x, y)$ in images, representing standing for the average size of the subject in the horizontal or vertical direction. Assume that the actual length or width d of the sample is equivalent to the corresponding pixel distance w, then the standard value S can be obtained as,

$$S = w / d \tag{11}$$

Let CL be a spatial displacement, according to Eq. 5 and Eq. 11, given by,

$$CL = S\Delta H \tag{12}$$

The displacements ΔH using in pixel as the unit in images can be transformed into spatial ones CL using in millimeter as the unit, to compare with the test data from EUTM(Electronic universal testing machine, made by SANS Company).

3 Experiments

Kraft papers, including linerboard and corrugating medium, are used in the experiment. Direction CD and MD of linerboard are grouped into A and B, and ones of corrugating medium are done into C and D. The size of all are 100mm× 25mm. Each group has five samples, a total of 20, to be pretreated in 23^0C and 50%RH of Humidity Chamber for 24 hours [8]. Samples are stretched along its length with EUTM, which chooses a 100N sensor and a speed of 5mm/min.

As observed from Fig. 2 (a) and (b), the stretched kraft papers appear to be elastic state and non-elastic state in the whole process of deformations. However, Poisson's ratio is a physical quantity only in elastic state. As is well known, the

force-displacement curve approximates to a straight line during elastic state like Fig. 2 (c) and (d). Thus it is necessary to choose force ranges [0, 1.5], [0, 3], [0, 0.5], [0, 1.2], in which samples A, B, C and D can be kept in elastic state, respectively. Image *I* and image *J* are taken in the beginning and ending of the each force ranges.

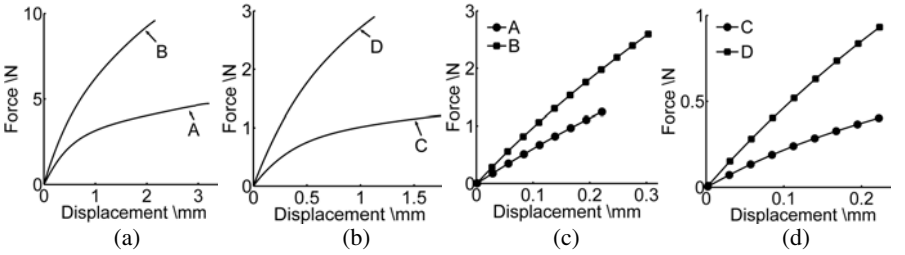


Fig. 2. Force-displacement curve of (a) Linerboard and (b) Corrugating medium; Force-displacement curve in elastic stage of (c) Linerboard and (d) Corrugating medium

4 Results and Discussions

Computing the acquired images of kraft papers with DIC, the displacement field is obtained in Fig. 3.

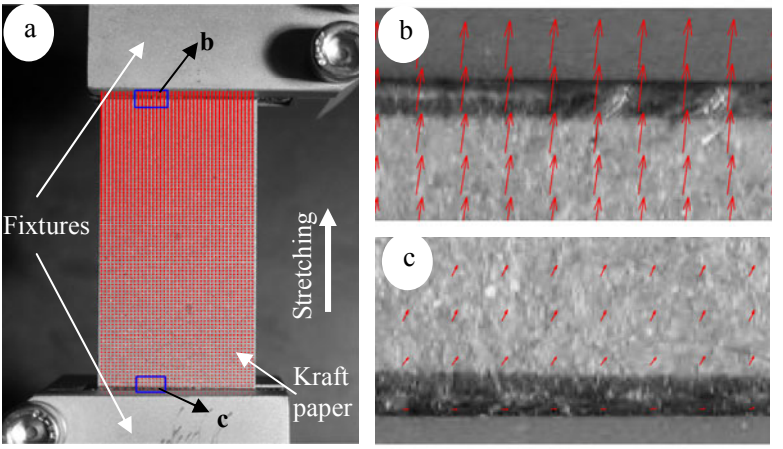


Fig. 3. Displacement field of the stretched sample

It illustrates that part (a) describes the whole situation of the deformation and displacement of stretched samples. As seen, kraft papers have no homogeneous deformations when stretched due to their inherent anisotropy. The level of stretching from the bottom to the top is gradually increased. Part (b) and part (c) show local details of the sample at the top and bottom respectively. The displacements of stretched samples are little at the bottom in the vertical direction, but they are much larger at the

top than at the bottom. Nevertheless, both of those are measured very small in the horizontal direction.

Based on the edge detection method of LOG for the simple calibration, the average of pixel distances w is 854pixel in the horizontal direction. According to Eq. 11, S = 25/854 = 0.0293mm/pixel. Similarly, in the vertical direction w is 1685pixel, S = 0.0297mm/pixel. The standard value S can be obtained by the mean in two directions, 0.0295mm/pixel. The relative error is 0.78%.

Table 1. Experimental data of samples

Group	NO.	Displacement				Relative error[%]
		ΔV [pixel]	ΔH [pixel]	CL[mm]	T[mm]	
A	1	0.390	9.438	0.278	0.278	0.153
	2	0.291	8.591	0.253	0.259	2.152
	3	0.298	9.381	0.277	0.282	1.861
	4	0.371	9.427	0.278	0.286	2.765
	5	0.340	9.349	0.276	0.281	1.854
B	1	0.949	11.405	0.336	0.347	3.044
	2	1.092	12.500	0.369	0.378	2.447
	3	1.207	13.009	0.384	0.392	2.099
	4	1.112	12.393	0.366	0.383	4.547
	5	1.007	11.986	0.354	0.358	1.232
C	1	0.708	11.236	0.331	0.339	2.226
	2	0.702	11.593	0.342	0.340	0.587
	3	0.681	11.438	0.337	0.334	1.025
	4	0.679	11.450	0.338	0.344	1.809
	5	0.658	12.026	0.355	0.364	2.535
D	1	1.563	11.738	0.346	0.354	2.183
	2	1.507	11.229	0.331	0.336	1.410
	3	1.588	11.237	0.331	0.334	0.749
	4	1.521	10.914	0.322	0.328	1.838
	5	1.475	10.481	0.309	0.320	3.374

According to the standard value S, the pixel distances are transformed into spatial displacements in Table 1. ΔH can be obtained by Eq. 5. CL are calibration displacements by Eq. 12, $CL = S\Delta H = 0.0295\Delta H$, and T are test displacements from EUTM.

Comparing CL with T, both of displacements are very micro but their differences are so small, and the relative error is less than 5%. It indicates that the experiment based on DIC is reliable and trustworthy. Table 2 is formed by Eq. 6. As shown in Table 2, Poisson's ratio approximates to each other in the same group. So compared with observational data, the results have a good agreement. The averages of Poisson's ratios of Linerboard are 0.073 and 0.175 in CD and MD respectively, and those of Corrugating medium are 0.119 and 0.275 in CD and MD respectively.

Table 2. Poisson's ratios

Group	Poisson's ratio					Average
	1	2	3	4	5	
A	0.083	0.068	0.064	0.079	0.073	0.073
B	0.166	0.175	0.186	0.179	0.168	0.175
C	0.126	0.121	0.119	0.119	0.109	0.119
D	0.266	0.268	0.283	0.279	0.281	0.275

5 Conclusions

This study shows that Poisson's ratios can be determined by direct observations of surface displacements using DIC in stretched kraft papers. Most usefully, it has proved to be a very useful technique for helping to characterize Poisson's ratio of the thin and softy materials.

It is available that increasing the amount of effective pixels means the precision between pixels improved by operating sub-pixel interpolation algorithm. That can enhance the reliability of the experiment.

Acknowledgments. We gratefully acknowledge the financial supports of the National Natural Science Foundation of China (NSFC) (NO.61170101, NO. 60774069).

References

1. GB/T22315-2008, Metallic materials - Determination of Modulus of Elasticity and Poisson's Ratio. China Standards Press, China (2008) (in Chinese)
2. Shao, X.Z., Shao, M.H., Bi, Y.F., Sun, L.J.: Testing Method of Asphalt Mixture Poisson's Ratio. J. Tongji Univ. (Nat. Sci.) 34, 1470–1474 (2006) (in Chinese)
3. Sutton, M.A., Wolters, W.J., Peters, W.H., Ranson, W.F., McNeill, S.R.: Determination of Displacements Using an Improved Digital Correlation Method. Image Vis. Comput. 1, 133–139 (1983)
4. Eitner, U., Kontges, M., Brendel, R.: Use of Digital Image Correlation Technique to Determine Thermome - chanical Deformations in Photovoltaic Laminates: Measure- ments and Accuracy. Sol. Energy Mater. Sol. Cells. 94, 1346–1351 (2010)
5. Yates, J.R., Zanganeh, M., Tai, Y.H.: Quantifying Crack Tip Displacement Felds with DIC. Sol. Eng. Fract. Mech. 77, 2063–2076 (2010)
6. Sánchez-Arévalo, F.M., Pulos, G.: Use of Digital Image Correlation to Determine the Mechanical Behavior of Materials. Mater. Charact. 59, 1572–1579 (2008)
7. Marr, D., Hildreth, E.: Theory of edge detection. Proc. R. Soc. Lond. B. 207, 187–217 (1980)
8. ISO 187:1990, Paper, board and pulps - Standard atmosphere for conditioning and testing and procedure for monitoring the atmosphere and conditioning of samples (1990)

Geomagnetism-Aided Navigation Based on Matching Algorithm for Underwater Vehicles

Yue Zhang*, Chong Kang, Hui Li, and Yuefang Yang

College of Science, Harbin Engineering University, Nantong Str. 145,
150001 Harbin, China
sinozhangyue@hotmail.com

Abstract. Geomagnetic assisted inertial navigation is an original navigation approach of autonomous underwater vehicles (AUV), and it has a high performance of concealment, stability and precision, which applies ideally to autonomous navigation of AUV. The ICCP algorithm's principle, selection of navigation parameters and realization method are analyzed in this paper and a method for extraction of contour is presented in this paper by comparative testing. It provides reference for improving the long range navigation accuracy and autonomous ability.

Keywords: Autonomous underwater vehicles, geomagnetic matching, inertial navigation, ICCP algorithm.

1 Introduction

Geomagnetic assisted navigation for underwater robot is a method of integrated navigation system (INS), which corrects errors by geomagnetic data. With time of navigation increased, INS will produce the errors accumulated over time, and military underwater robot run more risks of exposure through using satellite navigation, radio navigation and other forms of navigation system to correct it, also a waste of time and energy.

The combination of geomagnetic navigation and conventional INS has many advantages, such as its passivity, no electromagnetic radiation, full-time work, all terrain, miniaturization and low energy consumption, etc. It is geomagnetic navigation algorithm that forms the core of improving the navigation accuracy. In the present paper, one geomagnetic assisted navigation algorithm is introduced for underwater robot [1].

* About the author: Yue Zhang(1985-), male, graduate student.
The author's teacher: Chong Kang(1970-), male, professor.
Fund project: National Natural Science Foundation of China (61174192).

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 59–65.
springerlink.com

2 Geomagnetic Assisted Navigation for Underwater Robot

2.1 Geomagnetic Assisted Navigation System

The geomagnetic assisted navigation system consists of geomagnetic detection module, inertial navigation module and geomagnetic data graph module. The basic principle is as follows: geomagnetic detection module preprocess data like diurnal correction, noise removal, and Carrier magnetic field interference compensation, by which geomagnetic measured sequence are processed; the navigation computer receives the sequence; and by extrapolation the input function parameters is generated. After measured latitude and longitude information by the INS are transmitted to the navigation computer, matching algorithm realizes data fusion, then the correction of latitude and longitude information are returned to the INS[2].

2.2 Selection of the Geomagnetic Navigation Parameter

The geomagnetic field varies with space r and time t, its Intensity is expressed as:

$$B_{(r,t)} = B_m(r,t) + B_a(r) + B_d(r,t) \tag{1}$$

Commonly we use geomagnetic anomaly $B_a(r)$ and geomagnetic anomaly maps as measured value and reference frame of geomagnetic assisted navigation for underwater robot. After removing the disturbed field intensity $B_d(r,t)$ and the main field intensity $B_m(r,t)$ from the total amount of the geomagnetic field intensity $B_{(r,t)}$ measured by geomagnetic sensors, we get the geomagnetic anomalous intension $B_a(r)$.

3 Technical Route and Realization of the Algorithm

3.1 Principle and Process of ICCP algorithm

INS provides a sequence of track points' position value (φ_i', λ_i') $i = 1,2,\ldots n$. ((φ_i', λ_i') represents latitude and longitude). At the same time, geomagnetic sensors record the geomagnetic field measurements of corresponding points. Because there are accumulated errors, drifting and random influence of environment, there are also errors between measured track points' position value (φ_i', λ_i') by INS and the real track points' position value (φ_i, λ_i) $i = 1,2,\ldots n$. The purpose of the algorism is to eliminate the errors by finding the closest contours points on the stored geomagnetic anomaly map compared with real-time measurement and performing iteration.

Figure 1 shows an iterative process: hollow points represent the position of INS measurements, solid points represent the real position, and the side ones represent the nearest contour points. Family of curves c_i is a contour set extracted by the sensor's

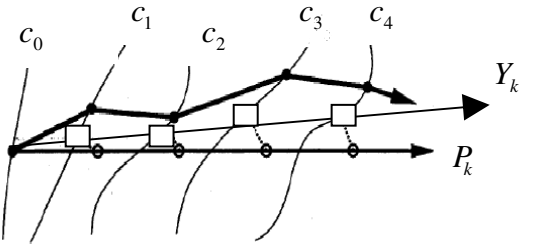


Fig. 1. Schematic diagram of ICCP algorithm

measurements of the corresponding INS positions. However, the corresponding geomagnetic values of INS measurement positions (φ_i', λ_i') $(i = 1,2,\ldots n)$ compose the set P_k, each $p_{i,k}$ of which is used in the reference geomagnetic map to find the closest contour point $y_{i,k}$, coordinates notes for $(\varphi_i^c, \lambda_i^c)$, the set composed of closest contour points Y_k. And vector norm just can describe the similar degree of two sequences (vector). So find a minimum rigid transform Γ_k to make sure objective function d_k is also the most minimum.

$$d_k = \frac{1}{n}\sum_{i=1}^{n}\left\| y_{i,k} - \Gamma_k p_{i,k} \right\|^2 \tag{2}$$

The former Γ_k rigid acts on p_k .We get point set p_{k+1}, next, find a rigid transform Γ_{k+1} to make objective function d_{k+1} minimum, by parity of reasoning, until the iteration termination condition meet.

According to the analysis, ICCP algorism can be described as "Initial alignment——contours extracting——finding the closest contour points——rigid transforming——applying the transformation——determining whether to stop the iteration", of which, rigid transforming and contours extracting are the key factors in the algorism.

3.2 Rigid Transformation

The prerequisite of rigid transformation is to make sure that equation (2) is minimum and "Quaternion Law" is the most common and easier to obtain the orthogonal rotation matrix method. Find a rigid transformation Γ to make the measured value of point sets P_k and the closest contour point sets Y_k minimum distance: First rotate the P_k so it is on Y_k direction. Then translate, so that the sets' centers of mass coincide. Set: the rotation matrix of P_k is R, translation vector is t, namely: $\Gamma P_k = R \cdot P_k + t$ .Two sets' centers of mass are:

$$\tilde{y} = \frac{1}{n}\sum_{n=1}^{N} y_n = \frac{1}{n}\sum_{i=1}^{n}(\varphi_i^c, \lambda_i^c), \tilde{p} = \frac{1}{n}\sum_{n=1}^{N} p_n = \frac{1}{n}\sum_{i=1}^{n}(\varphi_i, \lambda_i) \tag{3}$$

Quaternion rotation matrix method can be used to indicate the unit quaternion $q=(q_0,q_1,q_2,q_3)^T$, and $\sum_i {q_i}^2=1$. Quaternion method initially used in three-dimensional case, quaternion element of which has the following relationship with rotation angle θ and rotation axis $\tilde{v}$:

$$q_0=\cos\left(\frac{\theta}{2}\right),(q_1,q_2,q_3)=\sin\left(\frac{\theta}{2}\right)\tilde{v} \tag{4}$$

Geomagnetic matching is two-dimensional problem, so rotation axis $\tilde{v}=(0,0,1)^T$, unit quaternion is expressed as $q=(\cos(\theta/2),0,0,\sin(\theta/2))^T$. Based on mathematical principles of ICCP algorithm, the rotation matrix can be constructed as follows [3]:

$$R=\begin{pmatrix}\cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{pmatrix} \tag{5}$$

To find the rotation angle, first construct a Hamiltonian 4×4 real symmetric matrix W:

$$W=\begin{pmatrix}S_{11}+S_{22} & 0 & 0 & S_{21}-S_{12}\\ 0 & S_{11}-S_{22} & S_{12}+S_{21} & 0\\ 0 & S_{12}+S_{21} & S_{22}-S_{11} & 0\\ S_{21}-S_{12} & 0 & 0 & -S_{11}-S_{22}\end{pmatrix} \tag{6}$$

Solution of $S_{11},S_{12},S_{21},S_{22}$ is as follows:

$$S_{11}=\sum_{i=1}^{n}(y_i-\tilde{y})(y_i-\tilde{y})^T \qquad S_{12}=\sum_{i=1}^{n}(y_i-\tilde{y})(p_i-\tilde{p})^T$$

$$S_{21}=\sum_{i=1}^{n}(p_i-\tilde{p})(y_i-\tilde{y})^T \qquad S_{22}=\sum_{i=1}^{n}(p_i-\tilde{p})(p_i-\tilde{p})^T$$

The four eigenvalues of the matrix W are real, and given by:

$$\lambda=\pm\left[(S_{11}+S_{22})^2+(S_{21}-S_{12})^2\right]^{1/2},\pm\left[(S_{11}-S_{22})^2+(S_{12}+S_{21})^2\right]^{1/2} \tag{7}$$

Note that λ_m is the largest eigenvalue, the corresponding eigenvector q calculated by the following formula:

$$(S_{11}+S_{22}-\lambda_m)q_0+(S_{21}-S_{12})q_3=0 \tag{8}$$

Rotation angle:

$$\tan\left(\frac{\theta}{2}\right)=\frac{S_{11}+S_{22}-\lambda_m}{S_{12}-S_{21}} \tag{9}$$

After determining the rotation matrix R, translation vector can be expressed as: $t=\tilde{y}-R\tilde{x}$, the rigid transformation Γ is solved.

3.3 Extraction of Contours

For the purpose of contours' extraction is to create contour equation and then to find the closest contour points. Two common methods are Kriging method and radial basis function method. For linear unbiased and minimum variance estimation, each sample is assigned a certain weight. Unknown quantity of given area is estimated using weighted average method. The Specific approach is to construct the interpolation function as a function of spatial point [4]:

$$F(P)=\sum_{j=1}^{n} a_j \phi(\|P-P_j\|) \tag{10}$$

In the formula, P_j is a set of measured value points by geomagnetic sensor corresponding to the P_k set. "Given factor" a_j is a function of one variable (radial basis function).When $F(p)$ meets interpolation condition $F(P)=c_i$, available is linear equations:

$$F(P)=\sum_{j=1}^{n} a_j \phi(\|P-P_j\|)=f_i \tag{11}$$

$\Phi=(\phi_{ik}),\phi_{ij}=\varphi(\|P_i-P_j\|)$, $a=(a_1,a_2,\cdots,a_n)^T, c=(c_1,c_2,\cdots,c_n)^T$ Then equation (11) can be expressed by matrix as: $\Phi a=c$, from which we can compute c and then $F(P)$. The choice of radial basis function has the following common functions: GAUSS functions, thin splines, Multiquadri function and so on. Here we choose Multiquadri function [5]:

$$\varphi(t)=(t^2+a^2)^{\beta}, \phi(t)=(t^2+a^2)^{-\beta} \tag{12}$$

Generally, a、β can be set. For comparing the two methods, we choose a $93m\times 63m$ area to do the experiment. Every 3 meters to set a data collection point, it constitutes one 31×21 grids of area. At the 704 intersection points, the geomagnetic sensors measure the total amount of the geomagnetic field intensity as the sample, as follows:

Table 1. Geomagnetic value measured by sensors in 31×21 grids of area

	l_1	l_2	l_3	l_4	l_5	l_6	l_7	l_8		l_{32}
r_1	54677	54687	54697	54696	54724	54735	54748	54734	……	54827
r_2	54680	54689	54707	54722	54742	54741	54752	54759	……	54835
r_3	54638	54667	54707	54734	54751	54754	54760	54788	……	54843
r_4	54269	54579	54699	54744	54768	54774	54772	54791	……	54840
r_5	*	54523	54693	54737	54764	54771	54776	54778	……	54806
…	……	……	……	……	……	……	……	……	……	……
r_{22}	54700	54746	54752	54753	54754	54753	54751	54749	……	54629

For a visual observation, we use Surfer9.0 simulation software to generate contour maps as follows:

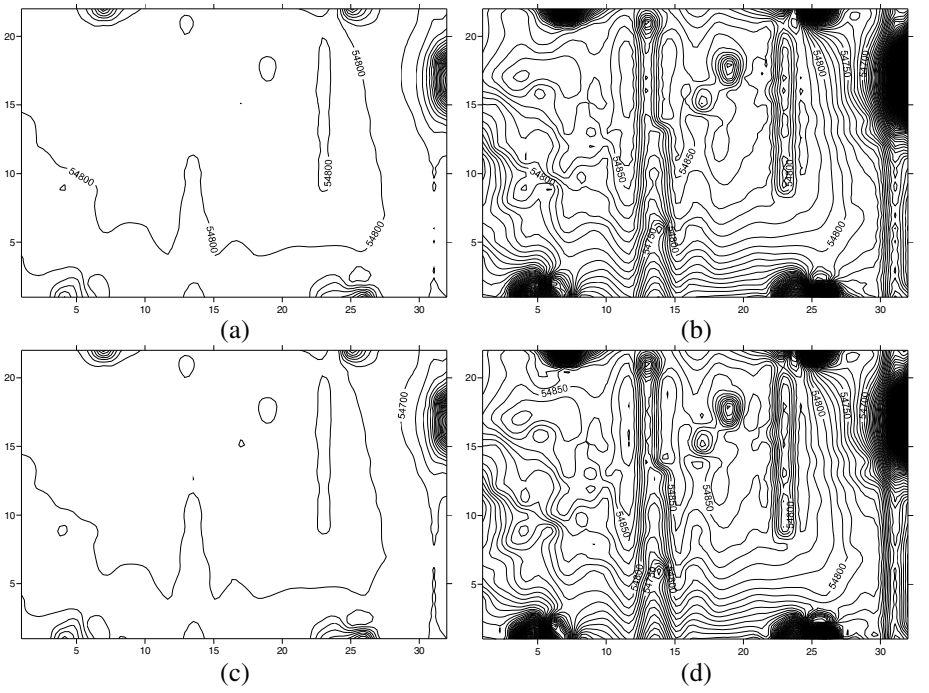


Fig. 2. Comparison of the contour maps between two different gradients with two interpolations ((a)Gradient of 100 with Kriging method. (b)Gradient of 10 with Kriging method. (c)Gradient of 100 with radial basis function interpolations. (d) Gradient of 10 with radial basis function interpolations.)

It can be seen from the figure above, the contour maps drawn by radial basis function interpolations are more suitable for ICCP algorithm in two different gradients, from the angles of the features of details and the gently degrees.

3.4 Initial Value and Terminal Conditions of Iteration

The selections of the initial value set method are always initial transform by random rotation and translation, or detailing three dimensions to definite complete set of local minimization. These two methods both have strengths and weaknesses, the former one with low amount of calculation but sometimes impossible convergence of the optimum global solution in limited number of time, the latter one with good convergence but high amount of calculation [6].

Four terminal conditions of iteration:

① $|R_{k+1} - R_k| / |R_k| < \varepsilon_{Rr}$ $|t_{k+1} - t_k| / |t_k| < \varepsilon_{Rt}$。 The ratio of incremental of rotation and translation to the original data is lower than the threshold ratio.

② $\left|R_{k+1}-R_k\right|<\varepsilon_r, \left|t_{k+1}-t_k\right|<\varepsilon_t$。 The absolute value of rotation and translation are both lower than the threshold ratio.

③ $\frac{1}{N}\sum_{i=1}^{n}\left\|y_{i,k}-R_k p_{i,k}-t_{i,k}\right\|^2<\varepsilon_{LSE}$。 The distance change in equation (2) is lower than the threshold value.

④Number of iterations exceeds a threshold value: $K > K_{\max}$.

4 Summary and Outlook

The research of geomagnetic matching algorithm is raised in the integrated navigation. The article carries out the basic principle of ICCP algorithm and its realization method. According to the technical route, some of the key technologies are introduced, of which in process of rigid transformation, there is also some weakness when random rotation and translation method is adopted.

In some condition, optimum global solution can not be available, so how to realize it is still a consideration. On the other hand, the reliability of ICCP algorithm is in urgent need to be improved, also for the condition of the corresponding geomagnetic maps' suitability.

References

1. Cai, Z., Wei, H., Ren, Z.: Research on underwater magnetic navigation technology. National Defense Science & Technology (3), 28–29 (2007)
2. Meng, J., Sun, F.-P., Zhu, X.-H.: Geomagnetic Model and Geomagnetic Matching Navigation. Science of Surveying and Mapping 35(4), 20–24 (2010)
3. Qiao, Y., Wang, S., Zhang, Q.: Selection of the Characteristic Variable of Geomagnetism for Matching. Seimological and Geomagnetic Observation and Research 28(1), 42–47 (2007)
4. Liu, R., Wang, C.: A new matching algorithm for fast magnetism. In: The National 13th Space and Control Technology Academic
5. Liu, Y., Wu, M., Hu, X., Xie, H.: Contour Constraint Based Geomagnetic Matching Method. Chinese Journal of Space Science 27(6), 505–511 (2007)
6. Ding, Y., Wang, J., Xu, F.: Application of Magnetic Navigation to Underwater Vehicle. Torpedo Technology 17(3), 47–51 (2009)

A Novel Approach to Provide Mobile RFID-Based Services Combined with Data Mining Technique

Juyoung Kang[1], Jinlong Piao[2], Hyeon-Seok Kim[2],
Seong Baeg Kim[2], and Chan Jung Park[2]

[1] Dept. of Computer Science and Engineering, Ewha Womans University,
Seoul 120-750, Korea
jykang@ewhain.net
[2] Dept. of Computer Education, Jeju National University, Jeju 690-756, Korea
herodragon107@hotmail.com, khsbasac@nate.com,
{sbkim,cjpark}@jejunu.ac.kr

Abstract. In advance of RFID technology, various types of services based on mobile RFID technology have been developed. However, there still exists limitation on RFID-based services such as transferring the simple information of exhibits at exhibition halls or managing the exhibits stored in warehouses. In order to improve the service qualities at exhibition centers, showrooms, or museums, we have developed a system, which can provide the disruptive services using mobile RFID technology. Our system accumulates the information such as how many people visit a specific exhibit, what the demographic information the visitors have, and what kinds of items the visitors see, and so on whenever visitors read a tag on an exhibit with an RFID reader connected to a mobile device. The accumulated information is stored in a database to provide an efficient search service and to bring a better service to the visitors. We have developed an application program for analyzing the visitors' data retrieval pattern from the database. Using this program, our system can get the information about the movement paths of the visitors. Also, by adopting data mining technology, we also propose a novel approach to enhance RFID-based services so that visitors can have a customized real-time question-answering service from assistants in a certain exhibition hall or museum whenever they are intrigued to know about exhibits.

Keywords: Mobile RFID-based Service, Customized Real-Time Helper Service, Data mining technique, Exhibition Management.

1 Introduction

Recently, as ubiquitous computing has been widespread throughout the whole world, various kinds of research on ubiquitous technology are underway. In particular, the research on the service system in a ubiquitous environment that can provide the information about the status of facilities and the degree of usage has actively been pursuing. Among them, the research on ubiquitous sensor network (USN) is one of

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 67–74.
springerlink.com

the essential part of ubiquitous computing. In addition, RFID technology is the most key component for USN [1].

RFID (Radio Frequency Identification) is a technology that identifies a thing without contact with it [2][3]. RFID technology consists of tag, reader, and antenna. A unique code is stored in tag and a reader is used for the purpose of decoding the code. An antenna is the device, which activates the reader's function. Typically, an RFID system takes the way that the code is used to get the information in the database, after the reader identifies the code in tag and transfer it to a PC or a mobile device [4]. RFID technology can be used to manage the stock in a large scale storehouse. For example, Incheon International Air Cargo in South Korea has recently built a new RFID system to improve the efficiency of the container management [5]. Subsequently, the quantity of physical distribution coming into the port grows up due to the increase of the processing efficiency of the port.

We can also find out another example at museums [6][7][8] or exhibition halls [9]. Visitors can use a hand-held mobile device provided by a museum or an exhibition hall agency. From that device, they can get the extensive information about the exhibits very easily. However, these kinds of RFID-based services cannot provide any interaction with the visitors. Even though the visitors can get information quickly and easily, there is no way for museum or exhibition agencies to get the information and the feedbacks from their visitors. That is because the existing RFID-services are one-sidedly. On the other hand, when the visitors have intriguing questions, they want to have immediate and authoritative answers from the agencies for museums or exhibit halls. However, it is difficult for individuals or small group of visitors to get a customized helper service promptly. Since this kind of service requires a lot of budget, museums or exhibit organizations can only hire the limited number of helpers. Thus, we should devise a new way how to provide a customized real-time helper service to visitors within a given budget.

On the other hand, as the complexity of the applications in IT realm has increased, new techniques such as AI technique or database technique are combined to solve the problems. Among them, data mining technique has been widely used in various kinds of areas recently. In particular, data mining technique is helpful when meaningful information should be extracted from a huge volume of datasets. Since massive amount of information is generated in RFID-based services, we can offer an efficient mobile RFID-based service if we apply data mining technique to this.

Currently, there has been no approach to provide real-time customized interaction between visitors and helpers in museums or exhibit halls. Also, there has been little research on how to use both RFID technology and data mining technique. In this paper, we describe a system, which is developed for the improvement of mobile RFID-based service for museums or exhibit halls, to overcome the limitation mentioned above. Our system was upgraded by adding data mining technique from the previous research based on RFID technology [10]. Therefore, we developed a system using mobile RFID and data mining techniques for improving a museum service considering the characteristics of the museum. Also, we propose how to provide a real-time customized service through interaction with visitors, who have an intriguing question about an exhibit under the human resource-restrained environment. The basic structure of the system is almost identical with the existing RFID system. The key component of the system developed is an application program

running on PDA, which can provide the information needed for managing the museum using database operation and data.

The rest of the paper is organized as follows. Section 2 briefly examines the case studies of the application developed using RFID or data mining technology. Section 3 describes our proposed system based on RFID and mobile devices such as PDAs for museums. Section 4 proposes how to use data mining techniques in the museum. Section 5 presents an approach to provide a real-time customized service to the visitors in the museum. In Section 6, we conclude this paper.

2 Backgrounds

In this Section, we briefly describe the meaning of data mining and RFID. And then, we examine a few case studies of mobile RFID-based service.

Data mining is a sequence of processes to discover new patterns from large datasets [11]. It is an interdisciplinary area because Statistics, Artificial Intelligence, and Database techniques handle data mining issues. There are two types of methods in data mining: one is predictive and the other is descriptive [11]. Among predictive methods, there are classification, regression, and outlier detection. Among descriptive methods, there are association rules and sequential patter mining, clustering, and so on. In this paper, we focus on clustering algorithms and association algorithms.

Already mentioned, RFID is a technology that identifies a thing without contact with it. Among the components of RFID, the tag differs in the operation depending on the type : passive or active type. An active type tag has its own power supply to reduce reader's power consumption and increase the identification distance. However, there is a problem that the tag's time span is short because of power supply. A passive tag responds to the wave sent from a reader. This type tag has low power consumption. However, it has relatively short identification distance and the reader's power consumption is high.

As described in [10], we can find a few examples in Jeju Stone Park [6], National Palace Museum of Korea [7], and National Museum of Korea [8] in Korea. Also, in [9], RFID and data mining technique are considered together. But this research focused on CRM(Customer Relationship Management). In addition, this system did not provide customized real-time helper system. However, in our proposed system, we include the customized real-time helper system.

3 The Proposed RFID-Based System

3.1 Mobile Device Application Program

The system we developed consists of a mobile device application program, a database, and a database application program. The database application program displays the sorted data retrieved from the database. They were developed using Visual Basic .Net. The application program's interface on a mobile device is composed of menus. The main menu consists of Information Display Window, Memo Window, and a few buttons including Port Search. The information of the exhibits at

the museum is represented on Information Display Window. A user can write a simple memo on Memo Window. Each button's description is shown in Table 1.

Table 1. Each button's function of the application

Button name	Description
Port search	Search the port between PDA and the reader.
Delete	Remove all information on Information Display Window.
Save	Save the information on Information Display Window to a text file.
Help	Display how to use this program on Information Display Window

The mobile device's user couldn't use the Port Search button. Once the Port Search button is pushed, a pop-up window asking for a password appears. If the password is not correct, it doesn't work. The mobile device is rented to the visitors after the port is connected by the museum's staff. If the visitors push the Save button, the content of Memo Window is stored into a file named 'yyyy-mm-dd name.txt' and the visitors can get the file later.

3.2 The Design of Database Tables

For storing the information of the exhibits and the accumulated information of tags, we developed a database by using MS SQL. Our databases are shown in Table 2. It describes the contents of each table designed in the database. The *Device_info* table and the *Visitor_info* table are updated when the visitor's information is registered. When each of the data contained in the *No_reading* table, the *Age_info* table, the *Route_info* table, and the *Gender_info* table is used in database application programs, the tables mentioned above are joined with the *Ex_info* table to get the exhibit's name. Once the *Visitor_info* table and the *Device_info* table are joined, the mapping information between visitors and mobile devices is provided to the application programs. And then, the gender information and the age information of the mobile device user, which is calculated, are used to update the *Age_info* and the *Gender_info* table. The *Route_info* table is updated with the *Device_info* table and the *Ex_info* table together.

Table 2. Each table's contents

Table name	Description
Ex_info	Exhibit's information.
Visitor_info	Visitor's information.
Age_info	Information accumulated by age.
Gender_info	Information accumulated by gender.
No_reading	Number of tag reading
Route_info	Viewing sequence information
Seq_info	Sequence information between the exhibits.
Device_info	Mobile device's status for rent

4 Applying Data Mining Technique

In order to improve the quality of services for visitors, personalized exhibition programs can be recommended based on the individual preferences. In addition, if we obtain the characteristics of visitor groups, we can develop new programs or differentiate the programs based on the visitor's watching history. Data mining techniques, which extract hidden patterns or trends from large amount of data, can be used to discover useful information for high quality museum services.

4.1 Visitor Segmentation Using Clustering

Visitor segmentation involves the identification of groups of visitors with similar characteristics in touring patterns. There has been much research on visitor segmentation in market basket analysis [12] and on visitor segmentation in the area of tourism and hospitality, based on data mining techniques. Clustering analysis technique, which divides data into groups of similar objects, can be applied to the data collected from museum services to do visitor segmentation. Although rather scalable clustering algorithms such as ROCK or OPTICS show comparable performance and capable of obtaining high quality clusters [13], most commercial mining products implement more simple algorithms like K-means and CLARANS[14]. K-means is a partition clustering technique that attempts to find a user-specified number of clusters (K), which are represented by their centroids. Since K-means algorithm has some limitations related with handling noises and outliers or with choice of initial centroids, mediods based algorithms like CLARA and CLARANS have been proposed. Since collected datasets from museum services is rather smaller than usual market basket or financial datasets, which have hundreds of attributes for analysis, we can obtain high quality results using partitioning based clustering algorithms. Therefore, we adopt K-medoids approach to the task relevant data for visitor segmentation. First, we pre-process target attributes for each Visitor_info from related tables. Birth, job, sex and address attributes from Visitor_info table and list of Tag Ids from joining Route_info and Visitor_info tables should be extracted as relevant data. These attributes have different data types such as categorical, numeric and binary, we compute dissimilarity between objects of mixed variable types by combining the different variables into a single dissimilarity matrix. After we normalize the values of variables into the interval [0, 1] as a form of a matrix, mining algorithm can be applied to the matrix.

From the clustering analysis, we discover the groups of visitors with similar preferences, and we can advertise special programs or events to the proper visitor groups. Moreover, if we collect more information on using PDA services by users while they are watching the galleries, preferred services by a specific group can be identified, and our system can automatically recommend a proper type of service or a matched program to visitors.

4.2 Developing Personalized Recommendation Program Using Association Rules

Association rules mining generates rules which represents relationships between attributes in data. For example, we can have a rule such as {Exhibit 1 in History,

Artifact 3 in History} → {Artifact 10 in Arts, Exhibit 11 in Asian History} by analyzing the list of items which a visitor watches. Those kinds of rules give us an intuitive understanding about visitors as a form of frequent pattern as well as make us possible to predict visitor's further choice of exhibits or services [15].

Association rule mining is to discover rules which satisfy user specified threshold, minimum support [13]. Support and confidence is the measurement of interestingness of association rules. Support determines how often a rule is applicable to a given dataset, while confidence determines how frequently items in right-hand side of the rule appear in the transactions that contain the left-hand side of the rule. When we have an association rule a form of $X \rightarrow Y$, where X and Y are disjoint item sets, the formal representation of support and confidence are as follows.

$$\text{Support, } s(X \rightarrow Y) = p(X \cup Y)$$

$$\text{Confidence } c(X \rightarrow Y) = p(Y \mid X)$$

Among many scalable algorithms which have been developed in association rule mining area, Apriori-based approaches [16] are most simple and intuitive for extracting frequent patterns from data. In order to extract relationships between a set of exhibits which a visitor watched, we can apply apriori-based algorithm to the Seq_info table with a pre-defined minimum support value (pattern extraction step). When we obtain frequent patterns with enough support count from the algorithm, we can generate association rules from the patterns. The object of rule generation is to extract all the high-confidence rules from the frequent item sets found in the pattern extraction step, when the minimum confidence is given (rule generation step). Generated rules give us an explorative description about the collected data from visitors, therefore we can understand that how often the exhibit are visited and how many visitors watch specific set of exhibits together.

In addition, by integrating association rules and classification approaches, we can construct rule-based classifier using the generated rules. CBA [17] and CPAR [18] make us possible to classify a new record, which is a new visitor's path, and to predict the next movement to the exhibit of the visitor using associative classification. As a result, we can provide recommendations of exhibitions to a visitor based on the prediction results extracted from the history of visitor's movements.

5 Interaction Service between FAQ Processing Center and Visitors

Most visitors couldn't fully understand an exhibit by only watching the description of it. Sometimes, it's necessary to get more detailed explanation from a helper. However, when individual or small group visitors look around exhibits, it's not easy to get a service from a helper because the number of helpers available would be limited. In particular, we can see straightforward that curious students will have many intriguing questions about exhibits. To tackle this situation effectively, we propose a service that a visitor can communicate with a helper at FAQ processing center to get some specific information about an exhibit. We are developing the interaction service as shown in Fig. 1.

We can see in the figure that the system consists of mobile device, FAQ processing center, and FAQ database. Using the client program on mobile device, a visitor can

search FAQ lists to find out an answer to an intriguing question about an exhibit. If he couldn't find the answer from the FAQ lists, he can send the question to a helper residing in the FAQ processing service center with RFID tag id using wireless LAN. Then, a helper responds quickly to him using the FAQ processing service program, which is similar to a messenger program like MSN. Also, the helper needs to update FAQ database to list up FAQs from most frequent to least frequent question depending on how many times each question is counted.

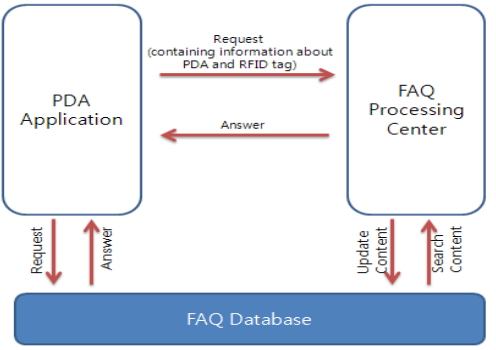


Fig. 1. Customized interaction service system between mobile device and the center

6 Conclusion

We developed a system that analyzes the pattern of the exhibition tours and provides museums or exhibition organizations with the useful information using RFID and data mining technique. With the proposed system, museums and exhibition organizations can provide more improved and high quality services to visitors and can provide a systematic and cost-effective management of exhibits to staffs and helpers of the organizations. Also, the accumulated information can help the organizations' management. In addition, with data mining technique, we presented a new approach to provide a customized real-time interaction service to the visitors for the purpose of answering visitor's questions about the exhibit.

Acknowledgements. This paper is results of a study on the "Human Resource Development Center for Economic Region Leading Industry" Project, supported by the MEST and NRF.

References

1. Kang, P.G., Kim, J.H., Choi, J.S., Choi, W.I.: Design and Realization of Facility Management System using Removable RFID. Journal of Korea Multimedia Society 10(4), 492–505 (2007)
2. Weinstein, R.: RFID: A Technical Overview and Its Application to the Enterprise. IT Professional 7(3), 27–33 (2005)

3. Floerkemeier, C., Sarma, S.: An Overview of RFID System Interfaces and Reader Protocols. In: Proceedings of the 2008 IEEE International Conference on RFID (2008)
4. Lee, J., Choi, S., Kim, J., Ko, Y.: The Contents Management System using RFID in Wireless Environment. In: Proceedings of the Korean Institute of Information Scientists and Engineers Conference, vol. 33(2), pp. 818–823 (2006)
5. Digital Times, http://www.dt.co.kr/contents.html?article_no=2010071202010560739010
6. Ahn, U.S.: Jeju Stone Park's Audio Guidance System and the Exhibit Management System using RFID tag. Korean Local Information Research & Development Institute (2007)
7. National Museum of Korea, http://www.museum.go.kr
8. National Palace Museum of Korea, http://www.gogung.go.kr
9. Kim, D.H., Kang, M.S., Park, C.J.: Development of RFID-based Automatic Analytical Information Generation System for Ubiquitous CRM in an Exhibit Hall. Journal of Korea Multimedia Society 12(1), 85–96 (2009)
10. Kim, H.S., Park, C.J., Kim, S.B.: Improvement of Jeju National Museum Service using Mobile RFID. In: Proceedings of the 2nd ACM International Conference on Interaction Sciences: Information Technology, Culture and Human, pp. 1066–1071 (2009)
11. Witten, I.H., Frank, E., Hall, A.M.: Data Mining: Practical Machine Learning Tools and Techniques, 3rd edn. Morgan Kaufmann (2011)
12. Choi, P.B., Grossman, E., Gunopulos, D., Kamesam, P.: Identifying Prospective Customers. In: Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 447–456 (2000)
13. Tan, P.N., Steinbach, M., Kumar, V.: Introduction to data mining. Pearson Addison-Wesley (2006)
14. Ng, R.T., Han, J.: Efficient and Effective Clustering Method for Spatial Data Mining. In: Proceedings of the 20th VLDB Conference, pp. 144–155 (1994)
15. Huang, Y.P., Chuang, W.P.: Improving the Museum's Service by Data Mining and Location-Aware Approach. In: Proceedings of the IEEE International Conference on Systems Man and Cybernetics, pp. 2646–2651 (2004)
16. Agrawal, R., Imielinski, T., Swami, A.: Mining Association Rules between Sets of Items in Large Databases. In: Proceedings of the ACM SIGMOD Conference on Management of Data, pp. 207–216 (1993)
17. Liu, B., Hsu, W., Ma, Y.: Integrating Classification and Association Rule Mining. In: Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, pp. 80–86 (1998)
18. Yin, X., Han, J.: CPAR: Classification based on Predictive Association Rules. In: Proceedings of the 3rd SIAM International Conference on Data Mining (2003)

LS-SVM Based Human Recognition on Automotive Active Safety System

Wang Duolin, Qin Guihe, Dong Jinnan, and Yuan Shuai

College of Computer Science and Technology, Jilin University,
Changchun, China
wateryew@hotmail.com

Abstract. In this article, a human recognition method, using wavelet transform and least square support vector machine (LS-SVM), is introduced. This had been performed in automotive active safety system. This method takes two-dimension discrete wavelet transform (2D-DWT) to do pretreatment, and extracts coefficients of correlation as eigen-coefficients, and then uses LS-SVM to search humanoid in certain area. The searching result will be presented to controller of a vehicle for making decision. A real vehicle experiment proves the validity of proposed method.

Keywords: Human Recognition, LS-SVM, Wavelet Transform, Active Safety System.

1 Introduction

The traditional passive safety system has been far behind the need of modern traffic system. Thereby, active safety technology will be the focus of future automotive research and development. The active safety system enables a vehicle actively taking prevention. Beside ESP, EBA, LDWS and EBD, which are handling aspects of active safety equipments, there are some other techniques, such as image processing and radar sensor. The image processing is a hot field of collision warning system in vehicle active safety technology, whose principle is to detecting human in front of a moving vehicle by using image processing technology.

Human recognition algorithms commonly use contiguous frames differencing, which mainly rely on detecting moving objects by finding difference between a frame and its following [1]. This kind of recognition asks for moving information, and is suitable for fixed camera and simple background. Some kind of human recognition algorithms use feature matching [2], which need an appropriate human feature matching method, for instance head and shoulder model or a ratio of length to width of human shape. The results of methods above are highly depending on body position, and the decisions are not accurate enough, because selected feature templates are not well fitted.

In view of previous research methods cannot meet the requirements of moving vehicles; this paper provides LS-SVM based human recognition. In this method, 2D-DWT with Haar wavelet is used for preprocessing images and finding corresponding

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 75–80.
springerlink.com

wavelet coefficients as the eigen-coefficients, and then LS-SVM is selected to detect all area that may contain humanoid by constructing learning machine. This human recognition does not need moving information, and has features of real-time and accuracy, therefore, it is good for moving devices, e.g. cars.

2 Wavelet Transform in Feature Finding

Wavelet transform is developed in middle of 80s in last century, which is wildly used for signal processing, image compression, pattern recognition, quantum physics and many nonlinear fields. Since an image under different illuminations has same wavelet transforms, wavelet transform is good to be used for obtaining mutation factor between human and background, which is called wavelet coefficient[3].

According to the continuity of signal *f(t)*, wavelet transform can be divided into continuity wavelet transform (CWT) and discrete wavelet transform (DWT). In computer applications, continuous wavelet must be discretized. In this paper, 2D-DWT is employed.

Selection of an appropriate wavelet basis is very essential. Haar wavelet is fast and easy to do image analysis, hereby, Haar wavelet is chosen as wavelet basis for 2D-DWT to decompose humanoid image. It is a differential function, and is able to present as following:

$$\phi(\chi)=\begin{cases}1 & 0\le t<1/2\\ -1 & 1/2\le t<1\\ 0 & other\end{cases}. \tag{1}$$

After original image is decomposed by wavelet, four images are produced, which contains different frequency factors. One sub-image keeps low frequency part of the origin, which is the main characteristics of the original image (first picture in Fig. 1); sub high frequency parts are saved in two pictures, one owns vertical detail component LH(second picture in Fig. 1) and another contains horizontal detail component HL(third picture in Fig. 1); the origin's high frequency part has the information of diagonal detail component HH(fourth picture in Fig. 1).

Fig. 1. Human image decomposed by wavelet

The HL and LH include edge information, brightness contrast information and mutation information. After processed with Haar wavelet, both HL component and LH component are given to LS-SVM learning machine as input. With treated

components, training speed and recognizing speed are increased, the number of input dimensions of classifier is reduced, and recognition accuracy is improved.

3 LS-SVM for Human Recognition

3.1 LS-SVM Algorithm

In 1999, J.A.K.Suykens and J.Vandewalle introduced LS-SVM [4], and changed the error of optimization from first order into secondary order. Thereby, a series of different equality constraints were launched, which covers a problem of solving quadratic programming into a problem of solving a set of linear equations. Finally, quadratic programming problem is changed to calculating a matrix's inverse, and using least squares in SVM is achieved.

Assumption: $\{(xi, yi) \mid i = 1,2,\ldots,N\}, xi \in R^n$ are input vectors, $y_i \in \{1,-1\}$ are output variables. LS-SVM classifier can be built as a optimization problem:

$$\min_{w,b,e} J_3(w,b,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{N} e_i^2 \quad . \tag{2}$$

subject to the equality constraints $y_i[w^T \varphi(x_i) + b)] = 1 - e_i, (i = 1,\ldots,N)$.

γ is punishment weight, $e_i \in R$ are error variables. Building mapping functions from input space to K-dimension space $\varphi : R^n -> R^k$, solving this optimization problem, defining corresponding Lagrange function.

$$L(w,b,e,a) = J_3(w,b,e) - \sum_{i=1}^{N} a_i \{ y_i [w^T \varphi(x_i) + b] - 1 + e_i \} \quad . \tag{3}$$

where $a_i \in R^n$ are Lagrange multipliers. The conditions for optimality can be written immediately as the solution to the following set of linear equations

$$\begin{bmatrix} 0 & y^T \\ y & ZZ^T + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ \vec{1} \end{bmatrix} \quad . \tag{4}$$

where $Z = [\phi(x_1)^T y_1, \phi(x_2)^T y_2, \ldots, \phi(x_N)^T y_N]^T$,

$$y = [y_1, y_2, \ldots, y_N]^T, e = [e_1, e_2, \ldots, e_N]^T, a = [a_1, a_2, \ldots, a_N]^T, \vec{1} = [1,1,\ldots,1]^T$$

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$$

Select kernel function to get decision function of LS-SVM.

$$y(x) = \mathrm{sgn}[\sum a_i y_i K(x, xi) + b] \quad . \tag{5}$$

a, b are solution of linear equations. Kernel function can be polynomial, RBF, MLP, etc.

3.2 Human Recognition Method

Human recognition method is split into two parts: LS-SVM classifier training and human recognition. In this paper, training method uses samples library to train classifier for building recognition system, and gives specified positive samples and negative samples to test the system. When recognition results are incorrect, the un-recognized samples will be given to the system with expectations. Repeated training and testing will finally produce an efficient recognition system. Images captured by car camera are able to be input for the trained recognition system to be analyzed. Because the size of humanoid in real images is not fixed, an humanoid image was done angular transformation, brightness changing in 0.5 to 1.2 times magnification increments by a step of 0.1 to make copied samples for the library. Those copies will be used to train classifier separately, which ensures the real-time [5].
Using LS-SVM for training has four steps:

Step1: Take samples from the library, and make transform with Haar wavelet. Use HL and LH components as input vectors x_i *(LH, HL),i=1,2,…,N* for LS-SVM, and give the corresponding expectations $y_i \in \{1,-1\}$.
Step2: Use LS-SVM algorithm to compute a and b with input from step1.
Step3: For specific vectors x_i to be tested, select a kernel function $K(x, x_i)$ to calculate decision function (5). From the value of $y(x)$, which is 1 or -1, x_i can be decided to be humanoid or not. In this paper we choose polynomial kernel [6].
Step4: Review the output results, if classification error occurred in step3, all incorrect testing vectors will be given to LS-SVM as input to repeat step1.
Step5: Repeat the above steps until all samples completed.

Accordingly, the training of LS-SVM classifier is completed. Then human recognition is able to start.

The human recognition method has three steps:

Step1: Setup zoom ratio to 0.5.
Step2: Move detection window to cover all possible areas. If there is human detected, the location will be marked and searching ends, otherwise go to step3.
Step3: Increase the zoom ratio by a step of 0.1. If the zoom ratio equals to 1.2, the searching ends and a message of non-humanoid is returned, otherwise repeat step2.

4 Real Vehicle Experiment

4.1 Build Sample Libraries

To training SVM classifier, first of all we should build sample libraries. Pictures in sample libraries were taken by the camera which was fixed at front windshield. Negative samples are some of the road pictures which were randomly selected(Fig. 2 shows an example). Positive samples are fixed size with human centered, therefore, raw picture has to be resized (Fig. 3 shows an example).

For ensuring validity of the training and real-time of recognition, a positive image was done angular transformation, brightness changing in 0.5 to 1.2 times magnification increments by a step of 0.1 to make copied samples.

Fig. 2. Non-humanoid images

Fig. 3. Humanoid images

4.2 Experiment Results

In real vehicle test, a total 400 photographs were taken, there are 200 humanoid and 200 for non-humanoid. The human recognition rate is 92.0%. And an average recognition time is 0.812s.The result is shown in Fig. 4.

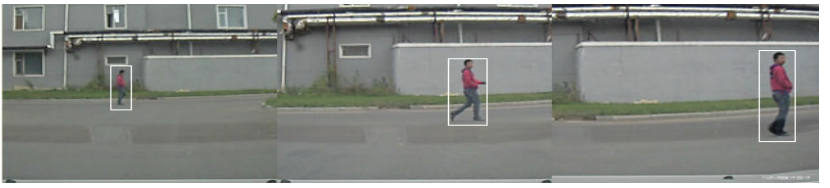


Fig. 4. Real vehicle human recognition

5 Conclusions

We did the research on 2D-DWT with Haar wavelet and LS-SVM in this paper, and performed LS-SVM based human recognition. We built remote experimental platform, and put the training and recognition in powerful PC. After the experiment in the real vehicle and remote server, the result from experiment shows: the algorithm is simple and has low time complexity and space complexity, which ensures the need of real-time in human recognition on automotive active safety system.

References

1. Zhong, Z.: Intelligent Surveillance Based On Technology For Abnormal Human Behavior Identification. Ph.D.Dissertation, Shanghai Jiao Tong University (2008)
2. Wang, J., Liu, W., Wang, J.: A Moving Object Detection and Recognition Method in Video Sequences. Computing Technology and Automation 26(3) (2007)
3. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 3rd edn. Prentice Hall (2008)
4. Suykens, J.A.K., Vandewalle, J.: Least Squares Support Vector Machine Classifiers. Neural Processing Letters 9, 293–300 (1999)
5. Grahn, J., Kjellstrom, H.: Using SVM for Efficient Detection of Human Motion. In: IEEE International Workshop on visual Surveillance and Performance Evaluation of Tracking and Surveillance, Beijing, China, pp. 231–238 (2005)
6. Shuji, Z., Frederic, P., Matthieu, C.: Spatio-Temporal Tube data representation and Kernel design for SVM-based video object retrieval. Multimedia Tools and Applications 55, 105–125 (2011)

Study on Forces Simulation of Gecko Robot Moving on the Ceiling*

Zhiwei Yu[1,**], Jianmin Chen[1,2], and Zhendong Dai[1]

[1] Institute of Bio-inspired Structure and Surface Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
yuzhiwei@nuaa.edu.cn
[2] College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Abstract. Gecko can freely move on the ceiling surfaces because of the directional adhesive structures. Dynamic model of gecko robot is created with Matlab/SimMechanics, A elastic damping model of the adhesive material on robot's foot is built. The crawling behavior of the robot on the ceiling is simulated with planned gait. The robot can move on the ceiling in simulation. The experiment results show that locomotion of the robot is similar with the dynamics of geckos moving on the ceiling. It is proved that the elastic damping model built is reasonable. The most important perturbation in motion is induced by the change of alternation between the foot's stance period and swing period. It provides references for the future experiment of gecko robot moving on the ceiling.

Keywords: Gecko robot, ceiling, adhesion, motion simulation.

1 Introduction

In recently robots are in application widely, the controllable mobile system (i.e. three dimensional space surface barrier-free robot) that can move freely on the smooth or rough surfaces is an important branch of robot, it has a wide range of applications and important value [1]. Gecko has four feet and can move freely and flexibly on the ground, wall, ceiling and so on. The gecko's adsorption principle and mobile way provide a new concept for breaking through the traditional crawling robot's limits, so gecko robot becomes a new research direction [2-3]. In 2000, Autumn and other researchers reveal adhesion properties and adhesion mechanism of gecko's feet, the researchers have studied for dry adhesion mechanism of adhesion materials inspired by geckos and applied them in the robot [4].

In the world many kinds of gecko climbing robot have been developed. Stanford have developed StickyBot. It is similar with gecko in shape and adsorption principle

* Project supported by the National Natural Science Foundation of China(60910007, 51105201), Project supported by the Postdoctoral Scientific Foundation of China (20100471339).
** Corresponding author.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 81–88.
springerlink.com

which can moving on the vertical plane [5]; Carnegie Mellon university developed Waalbot which has the adhesion wheeled structure [6]. In China Nanjing University of Aeronautics and Astronautics(NUAA), Beihang University and so on are also doing the research of gecko robot [7-10].

NUAA has been doing research of geckos for years. At present the new gecko robot of IBSS can crawl on vertical plane, but it is quite difficult and challenging to make gecko robot based on dry adhesion move on the ceiling. In this paper we establish the model of the gecko robot based on SimMechanics, use elastic damping model to simulate the adhesion force and ansys the results of experimental on ceiling. It provides references for the future experiment of gecko robot moving on the ceiling.

2 Structure of Gecko Robot

Inspired by geckos we design a new gecko robot[11](Fig.1). The length of crus is 60.5mm and the length of thigh is 49.5mm.The length and width of body is 205mm and 69.7mm. The legs of robot are the same, every leg has 3 servos. Feet and legs connected according to globe joints. The gesture of foot in the foot's stance period keeps the same and it is in the natural constraints state in the swing period.

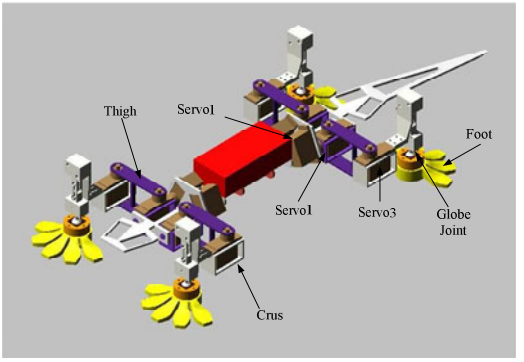
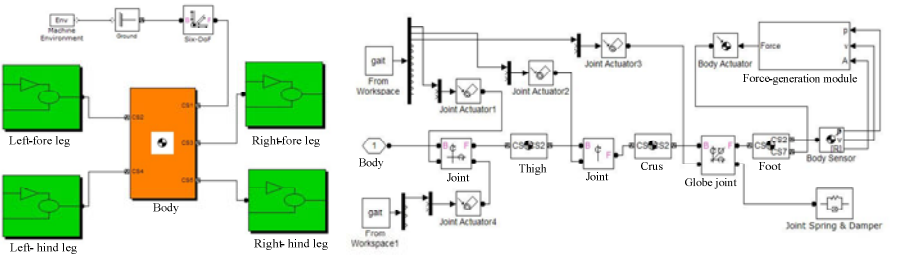


Fig. 1. Structure of gecko robot

3 Dynamic Model of Gecko Robot

SimMechanics can provide a suitable way to define the rigid body, quality characteristics, movement and Kinematic constraints so on. It can be used for the simulation of rigid body systems[12]. Dynamic model of gecko robot is created with Matlab/SimMechanics (Fig.2(a)) and every leg connect with body according to joints. The built models of every leg is the same, we show the left-fore leg for example(Fig.2(b)). The workspace of Matlab outputs the gait data. Force-generation module generates force between foot and ceiling and then drives the foot.



(a) The block diagram of gecko robot **(b)** The block diagram of the left-fore leg

Fig. 2. Dynamic model of gecko robot

A typical elastic damping model consists of an ideal elastic element and an ideal damping element (elastic element stress-strain relationship: $\tau = E\gamma$, damping element stress-strain relationship: $\tau = F\dot{\gamma}$). Total stress can be expressed as the following formula (elasticity: E, viscosity: F, strain: γ and strain rate: $\dot{\gamma}$).

$$\tau = E\gamma + F\dot{\gamma} \tag{1}$$

Similarly we use the following formula to express the force generated by feet (elasticity: Kp, viscosity: Kv, foot speed: V and displacement: ΔP).

$$F = Kp \cdot \Delta P + Kv \cdot V \tag{2}$$

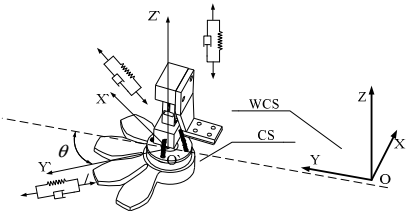


Fig. 3. The elastic damping model of the adhesive foot

As Figure.3 shown, $OXYZ$ is world coordinate system (WCS) and $O'X'Y'Z'$ is foot's coordinate system(CS). In CS, O' is the center of foot, Y' is of the axis of symmetry for foot and Z' is the vertical axis of foot. From study on the adhesion force of gecko feet, we find that feet generate directional adhesion force for attachment and detachment [4]. From the experiments of wall-climbing we know that it requires feet to provide greater adhesion force in the foot's stance period but smaller adhesion force for detachment. We calculate directional adhesion force(F_i) of foot by setting the feet'coefficients (Kpx/Kvx, Kpy/Kvy and Kpz/Kvz) in CS. We also set the maximum adhesion force in different direction. If $F_i <= [F_i]$, the foot can adhere to ceiling safely , it would detach if $F_i > [F_i]$.

Simmechanics's sensors get the displacement and velocity of feet in WCS, but we need the displacement and velocity of feet in CS to calculate the directional adhesion

force. The gesture of feet in stance period keep the same and it's in the natural constraint state in swing period. We assume that CS doesn't rotate in stance period and the $O'X'Y'$ plane of CS is parallel to the OXY plane of WCS at the same time. So if know the initial direction of foot, the gesture of foot can be determined and the displacement and velocity of feet can be transfered from WCS to CS in stance period. We define the Y' axis as the direction of foot and θ is the angle between Y' axis and Y axis(Fig.3). In simmechanics body sensors get $\mathrm{Rot}(\theta,\phi,\psi)$ which is a matrix that is used to perform the rotation of CS in WCS[13]. θ can be known from $\mathrm{Rot}(\theta,\phi,\psi)$.

$$\mathrm{Rot}(\theta,\phi,\psi)=\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}=\begin{pmatrix} C\theta C\phi & C\theta S\phi S\psi - S\theta S\psi & C\theta S\phi C\psi + S\theta S\psi \\ S\theta C\phi & S\theta S\phi S\psi + C\theta C\psi & S\theta S\phi C\psi - C\theta S\psi \\ S\phi & C\phi S\psi & C\phi C\psi \end{pmatrix} \tag{3}$$

$$\theta = \arctan\left(\frac{a_{21}}{a_{11}}\right) \tag{4}$$

$\mathrm{Rot}(\theta,\phi,\psi)$ represents a rotation whose yaw, pitch and roll are ψ,ϕ and θ. θ represents the rotation between the Y' axis and Y here. As knowing the θ the displacement and velocity of feet in CS can be calculated.

$$\begin{pmatrix} \Delta x' & V_{x'} \\ \Delta y' & V_{y'} \\ \Delta z' & V_{z'} \end{pmatrix}=\begin{pmatrix} \cos\theta & \sin\theta & 0 \\ \sin\theta & -\cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} \Delta x & V_x \\ \Delta y & V_x \\ \Delta z & V_x \end{pmatrix} \tag{5}$$

$\Delta x', \Delta y'$ and $\Delta z'$ represent displacements of three directions in CS and $V_{x'}, V_{y'}$ and $V_{z'}$ represent velocitys of three directions in CS. At last it needs to transfer the force from CS to WCS.

$$\begin{bmatrix} F_x \\ F_y \\ F_z \end{bmatrix}=\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{bmatrix} F_{x'} \\ F_{y'} \\ F_{z'} \end{bmatrix} \tag{6}$$

Force-generation module outputs the adhesion force(Fig.4). Sensors get displacement, velocity and $\mathrm{Rot}(\theta,\phi,\psi)$ and then input into Force-generation module. Function of different directions (X_function, Y_function and Z_fucction) calculate directional forces in CS. Transfer_function transfers forces from CS to WCS.

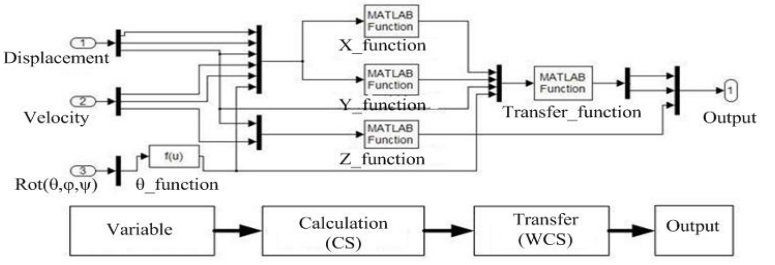


Fig. 4. Force-generation module

4 Simulation and Results

Before simulation we should plan gait: period (T) is 8s, stride length is 53.33mm, the displacement of foot relative to the body is 40mm and legs raise 20mm highest in the swing period. In a cycle robot moves straight and legs swing are in order as following: left-fore leg, right-hind leg, right-fore leg and left-hind leg. Based on the actual robot we set the appropriate parameters for each component including size, quality, center of mass and inertia properties so on. In WCS X axis is the lateral direction of robot, Y is the forward direction and Z is the normal direction. The direction of gravity acceleration is parallel to Y axis and the value is 9.8m/s^2, robot weight is 4.23N. The adhesion properties of every foot is the same, feet coefficients ($Kpx'/Kvx'/[F_{x'}]$, $Kpy'/Kvy'/[F_{y'}]$ and $Kpz'/Kvz'/[F_{z'}]$) are as following: $Kpx'_{+}=Kpx'_{-}=-1200$, $Kvx'_{+}=Kvx'_{-}=-15$, $[Fx'_{+}]=[Fx'_{-}]=-3\text{N}$; $Kpy'_{+}=-600$, $Kvy'_{+}=-10$, $[Fy'_{+}]=1.5\text{N}$, $Kpy'_{-}=-2000$, $Kvy'_{-}=-20$, $[Fy'_{-}]=4.5\text{N}$; $Kpz'_{+}=-2000$, $Kvz'_{+}=-25$, $[Fz'_{+}]=5\text{N}$, $Kpz'_{-}=-3000$, $Kvy'_{-}=-25$. Then the crawling behavior on the ceiling is simulated with planned gait and the process in the first cycle is as following(Fig.5) :

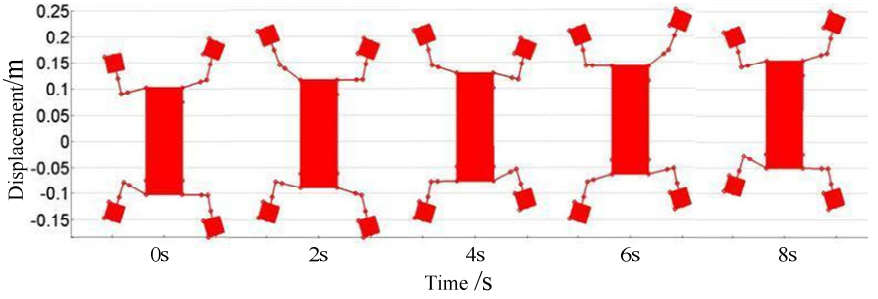
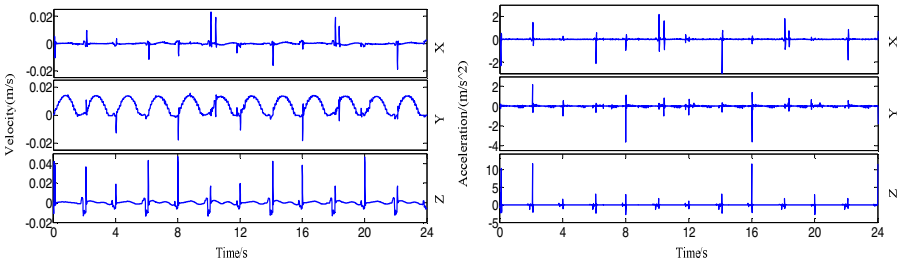


Fig. 5. Crawling process in the first cycle

4.1 Velocity and Acceleration of Body

Fig.6(a) shows the velocity of body in WCS in three cycles. V_x ranges from -0.02m/s to 0.02m/s, the highest V_y is 0.015m/s . V_z changes in the range from -0.015m/s to 0.05m/s. Figure.6(b) shows the acceleration of body. a_x is smaller than others, ranging from -2m/s^2 to 2m/s^2. a_y ranges from -4m/s^2 to 2m/s^2 and the change of a_z is the largest in three directions. Regardless of speed or acceleration there is perturbation in motion, they can converge to a stable state eventually. Obviously the most important perturbation in motion is induced by the change of alternation between the foot's stance period and swing period.

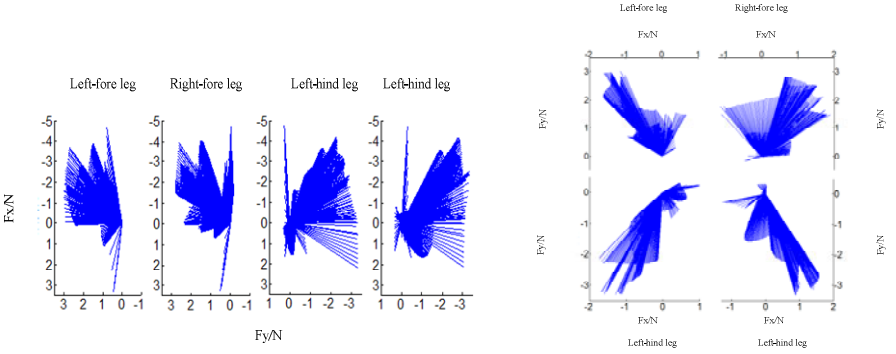


(a) Velocity **(b)** Acceleration

Fig. 6. Velocity and acceleration of body

4.2 Reaction Force between Feet and Ceiling

We get the force between feet and ceiling in three periods. *F*x, *F*y and *F*z represent force of three directions in WCS. From Fig.7(a) we know that normal force *F*z is negative at the most of time and the feet supply adhesion force .The weight of robot is balanced by the normal force. When the normal force is positive it is thought as the pre-stress of foot and the fore-feet is higher than the hind-feet. The result obviously show that when moving on the ceiling all *F*x and *F*y push away the center of the body mass(Fig.7(b)).It interesting that geckos'feet pull toward the midline of body and all lateral and fore-aft force push away the center of the body[14].



(a) Fz and Fy **(b)** Fx and Fy

Fig. 7. Reaction Force between Feet and Ceiling

The impulse I_i of the directional forces and gravity is calculated by integrating the foces(F_i) during the crawling time. We calculate the impulses of F_x, F_y, F_z and gravity in the third period(Table 1). There are no significant differences between left-legs' impulses of F_x and right-legs'. Fore-legs impulses are equal to hind-legs'

approximately. In normal direction the impulse of the normal force and gravity's is equal approximately.

$$I_i = \int_0^T F_i \cdot dt \qquad i=\text{X,Y,Z,G} \tag{7}$$

Table 1. Four feet impulse for the third period

Foot	F_x/ N·s	F_y/ N·s	F_z/ N·s	G/ N·s
Left-fore	-2.4020	7.2880	-8.3389	/
Left-hind	-1.4747	-5.8990	-8.4813	/
Right-fore	1.6122	6.4202	-8.3663	/
Right-hind	2.2591	-7.8147	-8.6459	/
Total	-0.0054	-0.0055	-33.8324	33.8376

Fig.8 shows us the force of right-hind leg in WCS and CS. In CS the force of X' direction has positive and negative fluctuation and is smaller than the force of Y' direction. In WCS, *F*x is negative, *F*y is positive and the peak value of *F*z approaches in the weight of robot. The force generated by leg has the periodicity. When right-hind leg is in the swing period, the reaction force between feet and ceiling is zero. There is preload force when foot incipiently contact to the ceiling. Right-hind leg's reaction force curve clearly reflects the preload and there is perturbation at the time. During separating from the ceiling, the force of right-hind leg would change into zero and then results in detachment. In the stance period the foot provides the necessary adhesion to ensure the stability and safety, but there are obviously perturbations at the time: 0s, 6s, 14s and 22s for example. We find that the perturbations usually occur during the change of alternation between the foot's stance period and swing period happens.

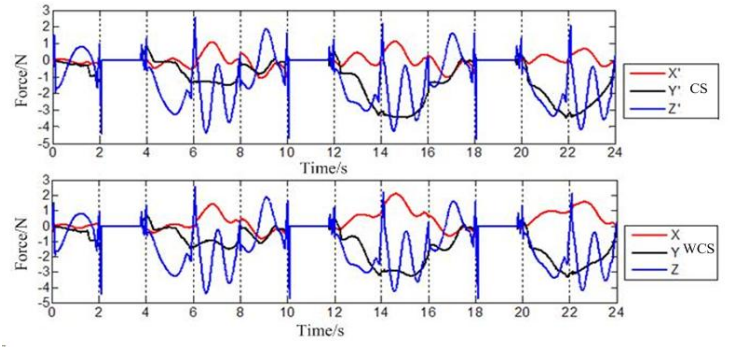


Fig. 8. Reaction force of right-hind leg in WCS and CS

5 Conclusions

We build the dynamic model of gecko robot based on Matlab/SimMechanic and the crawling behavior on the ceiling is simulated. Some conclusions can be obtained from simulation results.

(1) A elastic damping model of the directional adhesive material on robot's foot is built. In some way the simulation resuts are similar with the mechanical rules of geckos during moving on the ceiling. This proves that the model is reasonable.

(2) The main perturbation in locomotiom is induced by the change of alternation between the foot's stance period and swing period. We should reduce thhe influence in the future experiments of gecko robot moving on the ceiling.

References

1. Dai, Z.D., Sun, J.R.: Research Progress in Gecko Locomotion and Biomimetic Gecko-robots. Progress in Natural Science 16(5), 519–523 (2006)
2. Wang, T.M., Meng, C., Pei, B.Q., et al.: Summary on Gecko Robot Research. Robot 29(3), 290–297 (2007)
3. Ji, A.H., Dai, Z.D., Zhou, L.S.: Research Development of Bio-inspired Robotics. Robot 27(3), 284–288 (2005)
4. Autumn, K., Chang, W.P., Hsieh, T., et al.: Adhesion Force of A Single Gecko Foot-Hair. Nature 405(8), 681–684 (2000)
5. Kim, S., Spenko, M., Trujillo, S., et al.: Smoothvertical Surfaceclimbing Withdirectional Adhesion. IEEE Transactions on Robotics 24, 65–74 (2008)
6. Menon, C., Murphy, M., Sitti, M.: Gecko Inspired Surface Climbing Robots. In: Proceedings of the 2004 IEEE International Conference on Robotics and Biomimetics, pp. 431–436 (2004)
7. Wang, T.M., Meng, C., Guan, S.G., et al.: Structure Design of Gecko Robot with Compliant Shank. Journal of Mechanical Engineering 45(10), 1–5 (2009)
8. Dai, Z.D., Sun, J.R.: A Biomimetic Study of Discontinuous-constraint Metamorphic Mechanism for Gecko-like Robot. Journal of Bionic Engineering 4, 091–095 (2007)
9. Ruan, P., Yu, Z.W., et al.: Gait Planning and Simulation of Gecko Inspired Robot Based on ADAMS. Robot 32(4), 499–509 (2010)
10. Wang, H.J., Mei, T., et al.: Adhesion Array Design of a Novel Biomimetic Gecko Crawling Robot. Robot 28(2), 191–194 (2006)
11. Liu, X.Y., Dai, Z.D., Zeng, X.L.: A Guantitative Research on Gekko Gecko's Appendicular Muscle. Anatomay Research 27(4), 292–294, 301 (2005)
12. Huang, Y.A., Ma, L.Y., Liu, H.M.: Modeling Simulation and Senior Engineering Application Based on MATLAB7.0/Simulink6.0. Tsinghua University Press, Beijing (2005)
13. Xiong, Y.L.: Fundamentals of Robot Techniques. Huazhong University of Science and Technology Press, Wuhan (2008)
14. Wang, Z.Y., Wang, J.T., Ji, A.H., et al.: Locomotion Behavior and Dynamics of Geckos Freely Moving on the Ceiling. China Science Bulletin 55(9), 841–848 (2010)

Research and Development of Intelligent Slide Type Bathing Technical Aids*

Chunjing Tao[1], Hao Zhang[2], Jian Huang[2], and Qiang Xue[3]

[1] National Research Center for Rehabilitation Technical Aids,
Ronghua Road, Beijing Economic and Technological Development Area NO.1,
Beijing, China, 100176
taochj@gmail.com
[2] Huazhong University of Science & Technology, China
[3] Tianjin University of Science & Technology, China

Abstract. To overcome the bathing difficulties and solve the safety problem of severe impairment people, an intelligent slide type bathing technical aids is developed based on the ergonomic idea and modular mechanism design. The bathing technical aids can realize assisted docking and transferring. It has multiple functions including intelligent control, safety alarm, network centralized monitoring, hydrotherapy and so on. The development of slide type bathing technical aids provides a scientific basis for severe impairment people bathing and can improve the professional and technical level in intelligent bathing technical aids of China.

Keywords: Severe impairment people, Intelligent, Slide type, Bathing technical aids, Auxiliary bath.

1 Introduction

In China, due to traffic accidents, injuries and diseases, the absolute value and the proportion of the number of patients with severe disabilities arise everyday. In addition, there are about 150 million elderly people in our country, which also has the most disabled people in the world. The huge population base makes the aging become a serious social problem plaguing us within the next decade. When people get old, they are more likely to become disabled, which makes China has a much larger population of severe impairment people than other countries. The severe impairment people in disabled crowd cannot live by themselves because of prolonged bed rest. They need family and medical care, are under great psychological pressure, and often suffer from the complications. There are over 600 million key entitled groups in the whole country. Among them, 11,033 are staying in hospital. Most of them are very poor, old, disease-prone and completely bedridden. They do need not only the care and help from family, but also the social rehabilitation products to solve practical complications [1].

* Project Support: National science and technology support project Research and development of severe impairment people's bathing technical aids (2009BAI71B01).

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 89–95.
springerlink.com

Bathing is a challenge for severe impairment people, while they do need bathing because of the following three reasons. First, to maintain the people's skin hygiene, promote their skin blood circulation and reduce the complications. Secondly, to check whether there is damage in the paralytics' body before bathing. Thirdly, for severe impairment people, bathing is a systemic motion mode. Currently, our country can only produce relatively simple small bathing technical aids, such as shower chairs, bathtub handrail, walk-in bathtub, massage bathtub and so on. In China, the research of bathing technical aids for disabled people is just in its bud state [2]. Current technical equipments are relatively old. The research and development of bathing technical aids to promote the living standards of disabled people are far behind the developed countries. Thus, it cannot meet the needs of dysfunction such as elderly and people with disabilities. Especially in the docking, falling or drowning alarm and other bathing automation aspects, before bath and after bath are far behind the developed countries. Therefore, we can say that there is no "true auxiliary bath" in China [3]. From a structural design perspective, the types and functions of current Chinese bathing aids are very limited. They do not fully take into account the differences in the different users and cannot realize real-time monitoring or control.

Therefore, in order to meet the current health needs of severe impairment crowd, aiming at their bathing difficulties and securities, we research and develop safe, practical and reliable intelligent slide type bathing technical aids. They are used in concentrated support institution and part of families. They will solve the problem of bathing, improve the quality of life, and reduce the risk of complications（such as bedsores, muscle atrophy and joint contractures）for prolonged bed rest people.

2 Main Design

2.1 Functional Analysis of the Intelligent Bathing Technical Aids

The concept of the intelligent bathing technical aids proposed in this paper is shown in Fig. 1. Before the bathing, a nurse transfers the severe impairment patient who is about to bathing to the bathtub by some shifting device. Then, she transfers the patient to the load-bearing platform in the bathtub by the slide of the docking mechanism. Load-bearing platform can lower to a proper position by the elevator mechanism in the bathtub, such that the patient can soak in water. During the bathing, the intelligent bathing controller can adjust the temperature automatically. Advanced bating functions including the bubble bath and the eddy current bath are also available. The voice interface input lets the severe impairment people can make some real-time adjustment by their wish. Once emergency such as drowning happens during the bathing, intelligent bathing controller will signal out an alarm rapidly.

2.2 Overall Design Ideas

In terms of the actual needs of auxiliary bath for severe impairment people, by combining multiple disciplines including the medical engineering, the rehabilitation engineering, the mechanical design, the control science, we carry out multi-aspect

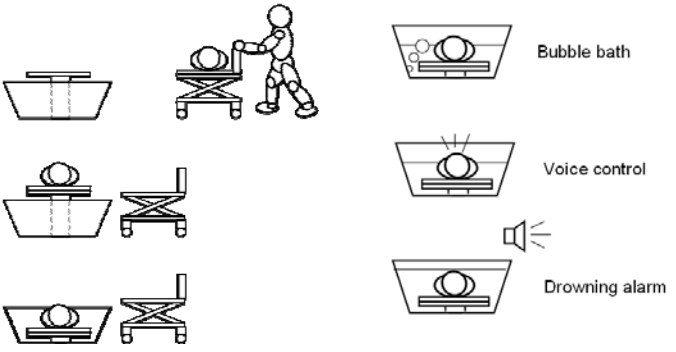


Fig. 1. Concept of intelligent bathing technical aids. (a) shifting process of the intelligent bathing technical aids; (b) part of the intelligence bath functions.

researches of the bathing technical aids. These researches comprise of structure design, intelligent capabilities, securities and hydrotherapy and so on. The function and design requirements of the intelligent bathing technical aids are: the design of bath equipment based on ergonomic, voice control, safety alarm, network centralized monitor, and operators of human-machine interface are simple. The overall design block diagram is shown in Fig. 2.

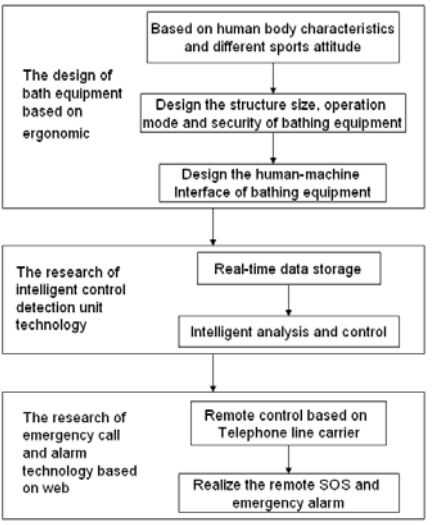


Fig. 2. The overall design block diagram of bathing technical aids

2.3 Main Mechanical Structure

The main mechanical structure of slide type bathing appliance consists of three parts: shifting device, folding bed bath and bathtub (include waterproof lifting gear) (see Fig. 3 for details).

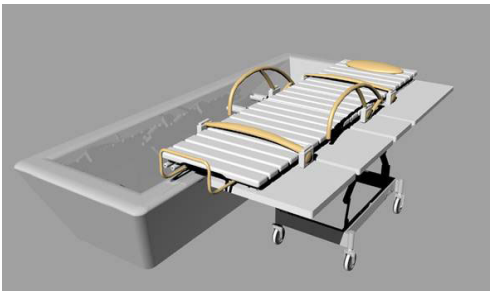


Fig. 3. Main mechanical structure of the slide type bathing appliance

The shifting device can realize the movement of severe impairment people. During the movement, people can lie flat as well as sit. At the same time, it can also help severe impairment people to go up and down, and transfer them from shifting device to elevator mechanism in the bathtub, thus facilitating the balneation of severe impairment people. Safety issue must be considered during the institutional and structural design process, so that the shifting device can realize the safety smooth transition during the bathing process of severe impairment people. The device must cooperate smoothly with severe impairment people, coincide with ergonomics, and prevent users from uncomfortable.

In the slide type bathing aids, the bath bed is put on the shifting device. It is a detachable bath bed, can realize the going up and down of severe impairment people during the bathing process. The bath bed can also assist users to raise back and bend leg during the bathing process.

Considering the special needs of disabled people and referring to the design of ordinary bathtub[4], the principles of the overall design of the bathtub are: a smooth docking with lifting device; minimizing the unused space and save water to the greatest extent under the premise that meet the bathing requirements; reasonable structure and reliable strength; using antibacterial material to prevent cross-infection; cost savings. The lifting structure in the bathtub is realized by linkage, it can be combined with a bathing bed, and assist severe impairment people to go up and down, raise back and bend leg during the bathing process. The schematic diagram is shown in the following Fig. 4.

3 Control System Design

3.1 Water Temperature Control

The Micro-Computer (MCU) is the core of water temperature control. By cyclically measuring the temperature using multi-channel temperature sensors, we adopt

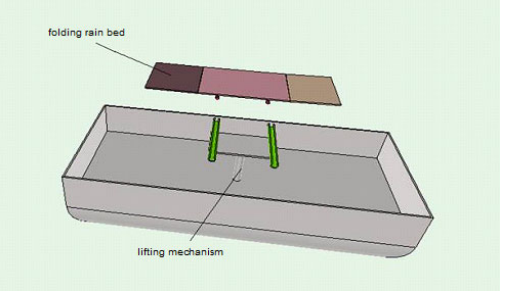


Fig. 4. The overall structure of the bath

fuzzy-control or intelligent PID-control to realize temperature regulation. The temperature can be preset from keyboard, and is displayed on a LCD display. Isolating the control signals is realized by using relays. By isolating the strong electric heating circuit and the MCU, it will guarantee the work of MCU more stable.

3.2 Voice Control

In order to improve the intelligence of bathing technical aids, except the external keyboard we designed the voice-input interface for patients by applying the voice recognition technology. In this module, the features of input voice are extracted and recognized by artificial neural network (ANN). The recognized results are used as an remote control input to regulate the water temperature or switch the bathing mode. The system can automatically identify a specific person's voice command after learning.

3.3 Safety Module Design

The main goal of safety module is to protect the safety of bathing people. It consists of intelligent bathtub environment monitoring subsystem, bath condition monitoring subsystem, leakage protection device, alarming subsystem, fall and drowning detection device and so on. The relationship between the subsystems is shown in Fig. 5.

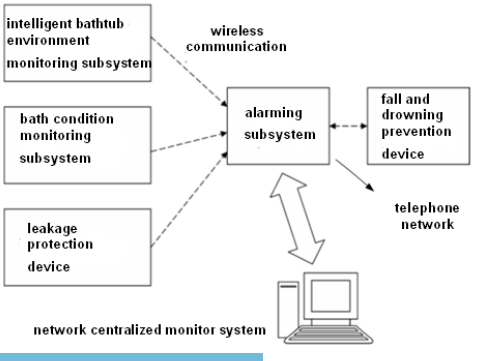


Fig. 5. Diagram of the safety protection module

The most important part of the safety modules is fall and drowning detection device. We combine the two different functions together in a small equipment. The fall-drowning detector is placed in the bathing cap. The main idea is showed in Fig. 6.

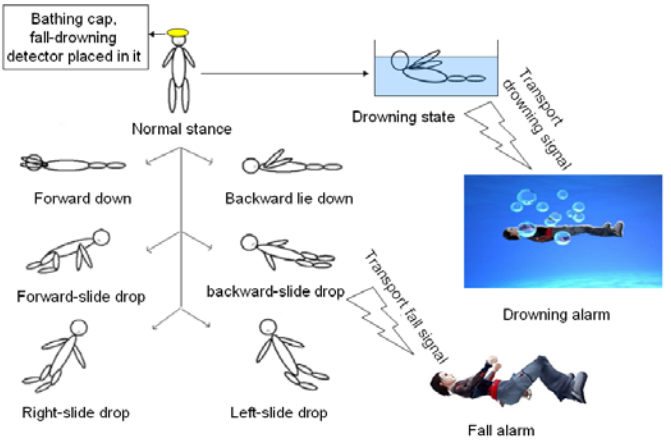


Fig. 6. Concept of fall-drowning detect

The realization of the system is accomplished by the combination of a three-axis acceleration sensor, a point-in flood sensor, an MCU, and a wireless communication module. The experimental device has been designed and is shown in Fig. 7.

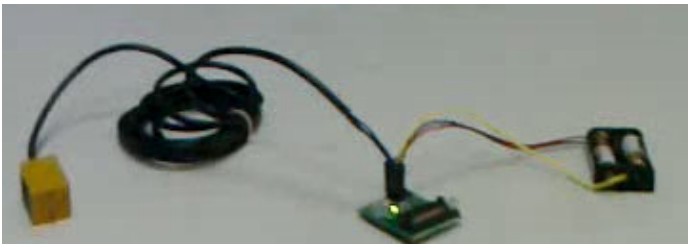


Fig. 7. Implementation of the fall-drowning detector

3.4 Hydrotherapy Module Design

The hydrotherapy module takes advantage of water vortex jet and bubble bath hydrotherapy treatment. Bubbles come from the gas generator in the bottom of the bathtub. Eddy current bath is realized by six water spray nozzle in the bathtub. The effective time, strength and temperature of eddy current are all automatically adjustable.

4 Conclusion

By combining many modules into a whole body, we construct an intelligent slide type bathing technical aids. In addition to solving the bathing problem of severe impairment people, the bathing aids also showing the whole society's care to disabled veterans, the elderly and disabled people, which plays a important role in building a harmonious society and solving the major issues brought by the coming aging era.

References

1. Weaver, C.M., Martinez, R., Maier, C.A., Cerqueira, C., Foulds, R.A.: Design of a wearable eye-gaze communication system for people with severe neuromuscular impairment. In: Proceedings of the IEEE 31st Annual Northeast Bioengineering Conference, April 2-3, pp. 26–27 (2005)
2. Yukawa, T., Nakata, N., Obinata, G., Makino, T.: Assistance system for bedridden patients to reduce the burden of nursing care (first report — Development of a multifunctional electric wheelchair, portable bath, lift, and mobile robot with portable toilet). In: 2010 IEEE/SICE International Symposium on System Integration (SII), December 21-22, pp. 132–139 (2010)
3. Huang, W., Luo, Y., Qiao, X.: The integer bath cabinet function design study on modern life style. In: 2010 IEEE 11th International Conference on Computer-Aided Industrial Design & Conceptual Design (CAIDCD), November 17-19 (2010)
4. Chih, H.K., Chen, T.L., Jain, A., Kemp, C.C.: Towards an assistive robot that autonomously performs bed baths for patient hygiene. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 18-22, pp. 319–324 (2010)

Intelligent Assistive and Robotics Development in China

Xiao-yu Zhang*

National Research Center for Rehabilitation Technical Aids, 100176 Beijing, China
zxy1949519@163.com

Abstract. Basic concepts of intelligent assistive device, and product groups. intelligent prosthetics, intelligent wheelchair, intelligent mobility assistive device, intelligent home automation and environment control of assistive rehabilitation training robot for assistive device, intelligent life, intelligent assistive device development tasks and future trends.

Keywords: intelligent appliances, status, trends.

1 Introduction

Rehabilitation assistive device is disabled people (including the elderly, persons with disabilities, trauma) compensation or improve function, improve the quality of life, one of the most direct and effective means to enhance social life involved in capacity. Compensation or alternative to the physical barriers for persons with disabilities, enhance or improve their activities of daily living for the elderly, sick person for non-surgical, non-pharmaceutical engineering help restore it to health, rehabilitation assistive device is to help people with disabilities return to society's most basic and the most effective means, for some physical dysfunction, configure assistive device even is the only means of rehabilitation.

2 The Concept of Artificial Intelligence

Intelligent assistive device designed to allow a variety of machines think like a person, with "wisdom" that mimic human brain function command to manipulate the body's other organs, to the "knowledge". Intelligent auxiliary with is used such of science principle, in auxiliary with implementation system in the combination has imitation human brain command body organ action of necessary module, makes the auxiliary with has "perception outside environment changes of ability", "analysis judgment reality situation of ability", "manipulation other organ of ability", "feedback manipulation results of ability", full imitation has human feeling organ collection information, brain analysis inductive finishing information, limb obey brain directive for action of ability, makes the auxiliary with can quickly perception, And make immediate adjustments to accommodate the requirement to complete a task.

* Author: Xiao-yu Zhang (born May 1949), male, professor. National Research Center for Rehabilitation Technical Aids, No. 1, Ronghua Road, Beijing Economic Development Zone.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 97–106.
springerlink.com

Application of intelligent assistive device object for the disability of the elderly and for persons with severe handicaps: my elderly, disability has reached 9.4 million people, there will be some disability 18.94 million people; bedridden, life cannot be take care of some 27 million people; 374 million Chinese families, families with long-term bedridden patients accounted for 8% family households throughout the country. Due to the lack of social care, a disability affecting at least two families of the elderly, at least tens of millions of households is plagued by disability care for the elderly, nursing and auxiliary personnel shortages in the future. One of the most lack is advanced home care equipment and intelligent assistive device. Intelligent assistive device appears on the concept of relatively new, February 2011 Community Development Division of the Department of science and technology in the collection "Eleven-Five" ten major scientific and technological achievements in the field of social development, "research and development of intelligent assistive device of life of disabled people" included.

3 Intelligent Assistive Products Group

Intelligent assistive device includes not only a smart prosthetics, currently, intelligent mobility assistive device, smart home smart wheelchair assistive and environment control, intelligent assistive living, assistive intelligent information exchange, also includes intelligent care robots, intelligent rehabilitation training instrument, in which intelligent assistive device width of life, is the biggest and most pressing needs. "Eleven-Five" during, in China technology competent sector and industry competent sector actively promoting intelligent auxiliary with research and development, Technology Department by technology support plans, and 863 plans, national technology plans support has multiple specifically-oriented heavy of residual impaired who of major research project, part project has made has group of high level of achievements, include several a intelligent life auxiliary with, signs sexual new products, will gradually push market, these results for intelligent auxiliary with products of further research lay has good of Foundation, Effectively promoting the development of China's intelligent assistive device.

4 Domestic Intelligent Lower Limb Prosthesis in China in 1994 and Have a Smart Knee Joint Research

2009 by Hebei University of technology, and national rehabilitation auxiliary with Research Center, and Tsinghua University joint development success domestic intelligent prosthetic pressure knee (Figure 1), used has four connecting rod and pressure cylinder of integration design, using sensor implementation gait and step speed recognition, intelligent controller can under wearing who of walking speed automatically adjustment cylinder valve open of, real-time adjustment prosthetic of swing speed, filled has domestic intelligent lower limb prosthetic area of blank.

Fig. 1. Domestic smart prosthetics pressure of knee joint

5 To Some Extent for Intelligent Wheelchair

It is based on intelligent wheelchair wheelchair rehabilitation robot. Is the application of Intelligent robot technology in intelligent wheelchair power wheelchair, blending a variety of fields of study include machine vision, robot navigation and positioning, pattern recognition, sensor fusion and the user interface, also known as intelligent wheelchair-type mobile robot. Smart wheelchairs and compared to traditional wheelchairs, "the brain", "eyes" and "ear", they are computer controlled systems, cameras and laser probes, also has a microphone. In China "Eleven-Five" planning focus support of service robot project--help old assistive robot system, national 863 plans intelligent wheelchair topics has development out in the end wheelchair (Figure 2) implementation has difference speed driven control, and smooth up stopped, and stability driving, and real-time avoidance barrier parking function; high-end wheelchair (Figure 3) seat can free lifting, back free tilt, bed and wheelchair implementation has butt, convenient ride personnel self-care Shang bed; automatically avoidance barrier bypass, drop, and pour doubled automatically alarm, also may on using who health state real-time detection.

Fig. 2. Domestic mid-tier wheelchair

Fig. 3. High-end domestic intelligent wheelchair

In China "Eleven-Five" planning focus support of service robot project--help old assistive robot system, national 863 plans intelligent wheelchair topics has development out in the end wheelchair (Figure 2) implementation has difference speed driven control, and smooth up stopped, and stability driving, and real-time avoidance barrier parking function; high-end wheelchair (Figure 3) seat can free lifting, back free tilt, bed and wheelchair implementation has butt, convenient ride personnel self-care Shang bed; automatically avoidance barrier bypass, drop, and pour doubled automatically alarm, also may on using who health state real-time detection.

6 Intelligent Mobile Auxiliary

Recession started with studies for many of the lower limb function of the elderly and the disabled, using intelligent assistive device correctly standing and walking training on lower limb function of recovery, and ensure the health of the body as a whole is of great significance. Homemade paraplegics walk training system. By Tsinghua University, and national rehabilitation auxiliary with Research Center, joint development of domestic paraplegia patients walking training system (Figure 4) the system breakthrough has joint reverse driven and waist with dynamic institutions, and movement trajectory planning and the hip knee coordination control, key technology, design has under patients height automatically generated gait trajectory of intelligent more joint coordination movement control system, help paraplegia and the other lower limb movement function obstacles patients for walking training, recovery its walking function.

Fig. 4. Domestic paraplegics walk training system

7 Intelligent Home Controlling Assistive Intelligent Environments Environment Control of Auxiliary

With help of severe controlling for household electrical appliances in the home for the disabled device, can be used to open the door, opened the curtains, dial and switch lights, TVs, and so on. System is the key to providing a use of disabled people are human-machine interface between function and electrical equipment. At the Shanghai World Expo in Shanghai eco home, there's one called "Ling Ling" home surveillance robot (Figure 5), this personal robot voice interaction is the intelligent home of "big", able to accept user instructions over the phone, telephone and other means, without compromise to accomplish tasks. Through professional Web video communication technology, home surveillance robot for the mobility, monitoring, cleaning and control home appliances in one. The owner in the company, to advance opened the air conditioners, water heaters in homes, on business, home lighting, television didn't pass, at any time you can turn off.

Fig. 5. Domestic home surveillance robot "Ling Ling"

8 Assistive Intelligent Life at Home and Abroad the Assistive Device

Includes a smart diet assistive intelligent life, may become incontinent nursing bed nursing assistive intelligent, smart, intelligent environment control system and nursing care robots, major types of daily living, a considerable number of intelligent living-assistive products has been commercialized and used directly in institutions or in the family.

8.1 Obey the Intelligent Nursing Machine Heavy of Limb Obstacles

Obey the intelligent nursing machine (Figure 6) application c language technology of RTX embedded operating system work, was nursing who wearing Shang nursing set Hou, intelligent nursing machine of induction device can automatically recognition

obey the of excretion status, select corresponding program set mode (with manually operation), automatically for rushed water extraction excrement, and washed anal, and urethral mouth and the near parts, and heater drying body, series of of Humanized nursing work, and records excretion number.

Fig. 6. Severe limb handicapped toilet smart care machines

8.2 Intelligent Based Urine Device

2006 national rehabilitation auxiliary with Research Center, and Beijing Aerospace Aviation University, units composition production, and learn, and research consortium, development has a intelligent set urine device (Figure 7), with humidity sensor of set urine interface and negative pressure air pump are seal to and has liquid bit sensor of barrels cover connected, its work principle is: loaded in set urine interface top of humidity sensor test to patients has urine liquid Shi, immediately to intelligent control circuit issued signal, started negative pressure air pump work, automatically help patients urination, entered storage liquid barrels in the; When the patient if there is no urine, intelligent control circuit parts an electrical signal is issued closed negative pressure air pump, when the liquid height of urine in the barrel when the predetermined value is reached, people are prompted on the display time.

9 Smart Care Beds at Present of Intelligent Multifunctional Nursing Bed

It is more similar to the robot. Domestic special life for disabled people (Figure 8) breaking through intelligent control, human-computer interaction and physiological information detection, anti-decubital ulcer key technologies developed 3 types with different capabilities, different grades of accommodation beds: product is modular in design, Plug and play. According to the Division of function modules, nursing bed should have five functions, each function can be increased or decreased as needed.

Fig. 7. Smart draw the urine

Fig. 8. Domestic special life for disabled people

10 Daily Life Care Nursing Robot System for Daily Life

Already exists on the current international system. This system in place for disabled people is to stay in bed, the body can't roam free in case, by setting the bed nursing unit (robot parts) to dietary care for disabled people, including storage cabinets required to take on shelf goods and other services. Bedridden and sound awareness for disabled people, this robot can be used in the wards, family, in the appropriate bodies, with the implementation of the care project: transformations, urine excretion, body movement, body position diet cuisine, messaging (communications), emergency escape (Figure 9).

10.1 Nursing Robot Unveiled in Nanjing Nanjing Domestic

First "nursing robot" serving the mobility such as paralysis, its most basic function is to replace nursing staff, patients themselves "care" themselves. "As long as the activity in patients with one hand, you can complete the operation, press the button to finish the basic living needs. "This" care robot "on the bed there is not much difference between usual and ordinary, similar to that of hospital beds in the ward,

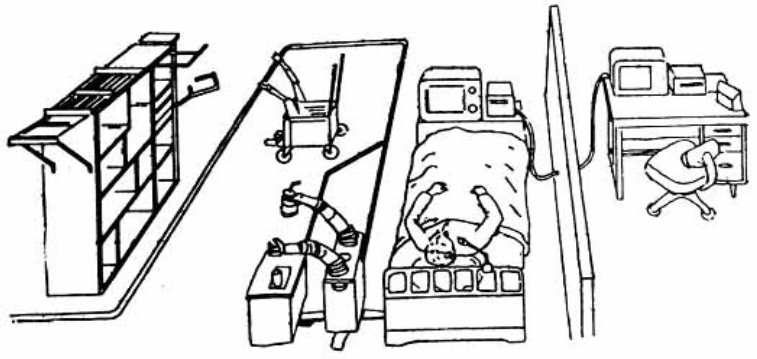


Fig. 9. Daily nursing robot system

next to the fence can adjust the height. Lying in bed in patients with sleep tired, push a button, the bed immediately became Chair; think of eating, reading the newspaper, and then press the, became backrest Chair; want to go to the toilet, press, butt below the active oval bezel will automatically disappear, followed by potty to make up.

10.2 Smart Care Robot Unveiled Beijing International High-Tech Expo

May 18, 2011, a medical care robot (Figure 10) held in Beijing China International Exhibition Center "14th China Beijing international technology industry exposition" with the audience, the robot and patient old man language exchanges, consultation, video playback and simple care.

Fig. 10. Smart care robot

11 Intelligent Robot Rehabilitation Robot

It was developed in recent years a new motor neurons in rehabilitation technology, as an important branch of the medical robot, has become a research hotspot of the international robotics.

Exoskeleton-type robot. So-called exoskeleton is a wearable bionic devices, artificial intelligence, In the health field, the application of the exoskeleton is not only

patients with mechanical legs to paralysis, it can help the natural gait of paraplegic patients out of wheelchairs, walk in; it can teach them how to learn to walk again.
Zhejiang University, state key Lab of fluid power transmission and control in the late 90 's commitment to the theoretical research on man-machine integration, flexible exoskeleton and gradually derived human-computer intelligent technology, flexible exoskeleton of upper limbs developed control system as shown in Figure 11, is actually an intelligent upper limb orthosis.

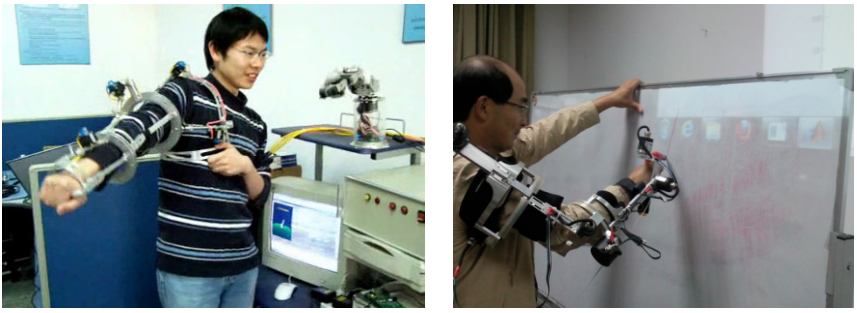


Fig. 11. Flexible exoskeleton controlled feedback systems, Zhejiang University

12 Intelligent Assistive Device Development Tasks and Future Trends in Recent Years

International intelligent assistive technology as innovation active in the field, information, networks, microelectronics, biomaterials, precision machining and other areas of advanced technology and innovation and intelligent assistive device penetration and integration is accelerating, creative ability of the developed countries and industry development in a strong position. Intelligent assistive device be the focus of global industrial competition area. Intelligent assistive device in China facing a major challenge: intelligent assistive device companies in China in the early stages, industrial Foundation is weak, the relatively low level of product development, high-end intelligent assistive device core technology and key components still cannot grasp, it is difficult to contend with foreign enterprises, industry at a disadvantage in the competition.

National "Twelve-Five science and technology development plan" in accelerating the development of population health technology, improving national health support capabilities, specifically for the elderly, the disabled population, enhance life assistive device development. Around the "rehabilitation for all" demand, according to Pratt and Whitney, intelligent, personalized development trend, alternative research structure, functional compensation, skills training, environmental technology products. Focused on research and development of intelligent speech functional compensation of audio-visual and assistive device (artificial vision, artificial larynx, artificial cochlea), a new type of artificial limb (smart and the energy-storing prosthetic knee joint, muscle and nerve control artificial limbs), intelligent assistive

living, cognitive function training system of the elderly, rehabilitation training system based on neural signals intelligence and other products to enhance the overall level of China's intelligent rehabilitation products.

References

1. The Chinese National Standards. Disability aids classification and terminology. GB/T16432-2004/ISO9999, p. 35 (2002)
2. Lin, L.-M.: Modern Rehabilitation Medicine Project. Shanghai Jiaotong University Press (1992)
3. Wang, J.: The basis of rehabilitation engineering. Xi'an Jiaotong University Press (2008)
4. Zhang, X.Y.: Disability aids assembly guide. China Personnel Press (2006)
5. Zhang, X.Y.: Wheelchair use conservation techniques in Chinese society Press (2010)
6. Yang, C.J., Chen, Y., Lu, Y.X.: Man-machine integration of theory and application of intelligent systems research and exploration. Journal of Mechanical Engineering (2000)
7. World Health Organization. Less resource settings in the supply of manual wheelchair guide (2008)

Design and Analysis of Built-In Test for the Absolute Locating Sensor of Maglev Train

Jun-ge Zhang[1], Ning He[2], Te-fang Chen[1], and Song Xue[2]

[1] Department of Information Science and Engineering,
Central South University, 410075 Changsha, China
[2] Engineering Research Center of Maglev Technology,
National University of Defense Technology, 410073 Changsha, China
junge1986@126.com

Abstract. Built-in test (BIT) of absolute locating sensor of Maglev Train system is an important technique for improving testability, maintainability and fault diagnosis capability of absolute locating sensor system. The basic principle of absolute locating sensor of Maglev Train is presented in the paper, and it analyses the issue of the absolute locating sensor in detail. An optimization design of BIT is made in order to deal with the problem. The experiment result shows the system can satisfy the practicable application of the absolute locating sensor of Maglev Train.

Keywords: Maglev Train, Absolute Locating Sensor, BIT, Fault detection.

1 Introduction

Built-in Test technology is an automatic testing capability of fault test and isolation provided by the equipment or the system itself, which is the common technologies to overall design and subsystems design, condition monitoring, fault diagnosis and maintenance decision-making of complex systems and equipment [1,2]. Foreign countries have done a lot of accumulated on research of BIT since 1970s, not only in theory, but also with the task in the application of the model and made a series of research results. Almost all the U.S. military or civil aircraft avionics and weapons systems equipped with BIT, running for their own identification and rapid detection of faults, which greatly improve the reliability of the systems, maintainability and testability and reduce life-cycle cost. BIT research domestic is still in early stages, mainly in the model task applications. BIT basic theory and method is relatively weak [1].

Maglev train is a new kind of railway transport using electromagnetic force to achieve the body of non-contact support and guidance, emerging through the linear motor traction [3-6], its traction control, zoning power, speed and position, and designated parking are closely related to the absolute position information of the train, as the core components of the maglev train, the absolute locating sensor train plays an important part in the safe operation of trains [6]. So it is significant to design a reliable and stable absolute locating sensor to improve the maglev train system security and reliability. To improve the testability and diagnostic capability of the system, a BIT

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 107–115.
springerlink.com

system taking computer technology as the core has been designed, by the sensor software self-test program of the absolute locating system, the key modules test itself to confirm whether it is working properly. When the system does not work properly, BIT can process failure module giving an alarm; when the system is working properly the system will give "system self-test normal" instruction.

2 System Structure and Testability Requirements of BIT

The absolute locating sensor of Maglev Train gets the information on the location mark along the fixed track by an absolute locating reader scanning the mark in order to obtain the absolute position information of the train. Absolute locating sensor model is shown in Figure 1.

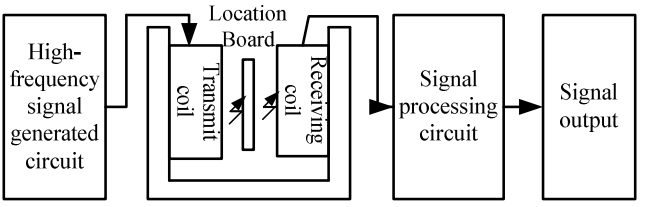


Fig. 1. Absolute locating sensor model

Absolute locating sensor is a complex circuit system, the hardware block diagram of the circuit is shown in Figure 2, the entire system is made up of five parts, which contain the power module, high frequency power transmitter module, receiver coil signal processing module, the digital processing module, and serial port transmission module [7].

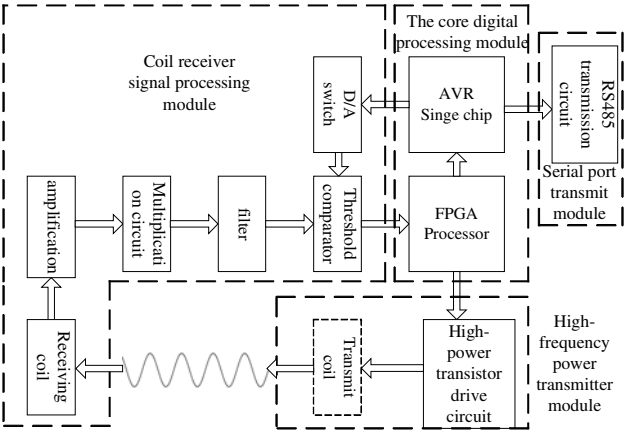


Fig. 2. Absolute locating sensor hardware circuit diagrams

The high frequency power transmitter module generates a high frequency square wave with FPGA to drive the transmit coil by high-power transistor; the signal processing module of receive coil includes LC resonance amplification, band-pass filter, synchronous detection circuit, comparator circuit; digital processing module using FPGA processor for signal processing and protocol generation, AVR microcontroller is responsible for the threshold voltage generator; Serial port transmit module using the 485 send circuit [8-10].

In the absolute locating sensor system, system test and state real-time monitoring subsystem unit is not only an important part of the system but also part of other modules. In other words, there is the subsystem level BIT and system-level BIT. This is because the components do not have the state detection circuit module fault location and isolation function, so system-level BIT is necessary. In addition, the absolute locating sensor system needs to collect the status of each subsystem and key parameters to help commanders make decisions when the system fails. Therefore, system-level BIT is more essential. In short, it is important to design a BIT system to improve the reliability, testability and maintainability of absolute locating sensor system.
The functions of absolute locating sensor BIT are as follow:

(1) Real-time monitoring: within a specified time period of the sensor works, real-time automatic testing work state and fault detection system and display test results;
(2) Functional tests: check the performance and function of the system;
(3) Fault display: to display work status and failure mode of the system, alarm information is given, including the failure of the specific location, number, type, degree of hazard, which are provided to the operator as command and decision-making basis;
(4) Fault isolation: BIT isolates the system fault to the circuit board; part of the fault is isolated to critical components.

3 Design and Analysis of BIT for the Absolute Locating Sensor

3.1 Absolute Locating Sensor Fault Analysis

3.1.1 Interrupt Fault of Sensor Coil Signal

The transmitting and receiving coils of Absolute locating sensor are the carrier of detected and received signal, they directly affect the accuracy of reading the sign board location information. When transmitting or receiving coil circuit failure occurs, the sensor receives signal will remain zero. Taking the transmit coil is a serial connection into account, if the coil circuit broke off, then all the receiver coil signal will remain zero; and all the receiver coils are independent, if a receiver coil signal is always zero value, then the coil circuit failure occurs.

3.1.2 Sensor Signal Processing Amplifier Circuit Fault

The sensor's analog circuit fault is divided into the following areas: operational amplifier circuit failure, filter circuit failure, the multiplier circuit failure, operational amplifier circuit failure is the main fault of the sensor’s signal processing circuit. Amplifier is a weak signal processing system with high sensitivity, very susceptible to outside interference signals, when the operational amplifier circuit input signal floating, the output voltage signal will be uncertain. High-frequency low-noise

operational amplifiers are used here, while the resistor and capacitor components chose are high-temperature coefficient, which can reduce the impact of noise and drift, but since absolute locating sensor is long-running working in a complex electromagnetic environment, including train suspension, guidance and other sources of interference effects of the magnetic field, amplifying circuit has higher probability of failure, which is mainly reflected in the distortion form of amplified waveform and the instability amplification and so on.

3.1.3 Processor Fault

FPGA and AVR is the core digital processing unit of the sensor, the process of generating waveform, testing signal and code combined and send of entire system are finished in the processor. Processor itself is a high-security high-reliability electronic component, but not exclusively the possibility of fault owing to the production process and devices with aging issues. Processor failure is mainly reflected in the instruction processing fault and unit circuit fault two aspects.

3.1.4 Communication Fault

Absolute locating sensor communicates to the communication network of speed and locating system through the RS485 circuit, communication fault will lead directly to the output of the sensor failure. Unlike common serial port data transmission circuit, RS485 circuits communicate with differential transmission, which will help eliminate the common mode interference of transmission. In the complex electromagnetic interference condition of Maglev train, the common fault of RS485 differential data transmission circuit is bit errors in data transfer.

3.2 Design and Analysis of Absolute Locating Sensor BIT

Absolute locating sensor plays an important part in a maglev train speed and position system , the state of it works is directly related to the security and stability of maglev train, how to test the sensor efficiently and accurately in the limited time is a question must be considered when absolute locating sensor designed. Absolute locating sensor's BIT is the primary methods and means to monitor the key function of the system and to detect and isolation the system's fault. Testing object, measuring parameters and testing functions (monitoring, inspection or quarantine) of the system BIT should be determined according to requirements and diagnostic program.

The type of Absolute locating sensor BIT is start-mode BIT by the execute style, it will start up before the task by the triggered command. Absolute locating sensor will start BIT to test each channel function when receiving a command sent by the control system. When performance exceeds the threshold or fault occur the indication that self-test abnormality is given.

The absolute locating sensor BIT can be classified to centralized BIT according to the structure, each subsystem have no their own micro-processors and micro-diagnosis, the control, diagnosis and decision making function of the whole system relies on a microprocessor ,the sub-system will collect the information and uploaded to the main microprocessor by data bus. The main microprocessor will give a comprehensive diagnosis to the functional status of the entire system by analyzing the received data

and comparing to the historical information stored, then output the status information to the display device and store this state information [11].

4 Absolute Locating Sensor BIT Design

The paper design a fault detection system for the sensor to deal with the common faults, it is shown in Figure 3, the fault detection system consists of hardware platforms, software platforms and alarm platform.

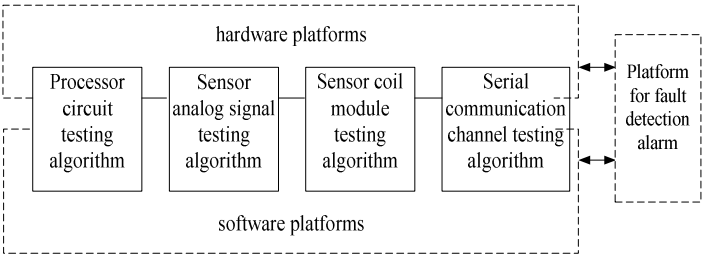


Fig. 3. Absolute positioning sensor fault detection system block diagram

4.1 Processor Test Method

As the main function of CPU is to execute instructions, you can determine it works well or not by detecting whether the CPU can correctly perform all its commands. Considering the design requirements of testing methods to the BIT software design, we choose the functional testing method based on instruction execution, the condition of their test results validity is that the test program to run properly. An auxiliary circuit called "watch-dog" is used here to avoid the test program "runaway" .The main part of the circuit is counter and timer, if the counter / timer does not receive any clear signal within a certain time, it will overflow and alarm.

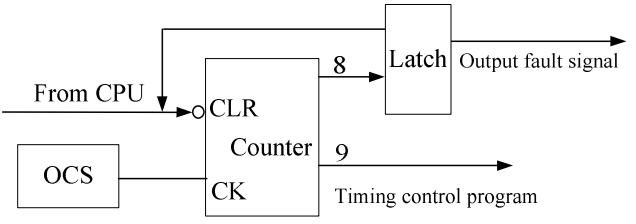


Fig. 4. Watchdog circuit

Thus, we determine whether the microprocessor can work correctly following the procedures destined for by adding clear instruction to the sub-process. If the microprocessor fails, then it can not send clear signals to counter / timer, once the

counter / timer does not receive any clear signal within a certain time, it will overflow and alarm, which is the sign of processor fault, then automatically clear to zero and resumes counting.

4.2 Sensor Signal Processing Circuit Testing Methods

4.2.1 D / A Converter Circuit Testing Algorithm

The structure of the whole detection system is shown in Figure 5. It should be noted that the test method is avail only when there is no trouble in the A / D circuit, when the A / D and D / A go wrong, the fault test on the D / A circuit is meaningless.

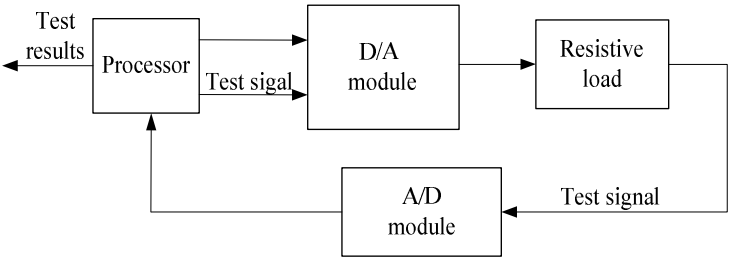


Fig. 5. D / A converter circuit fault detection schematic

D / A converter circuit fault detection principle is: firstly the processor send a test signal, the test signal pass through the D / A converter circuit followed by a resistive load, the load resistances should be as large as possible to avoid the voltage changes caused by load effects .Then sample the signal by A / D closed-loop circuit, and compared with the test signal, if the difference exceeds the threshold, then the D / A converter circuit fault will be reported.

4.2.2 The Threshold Comparator Circuit Testing Algorithm

The fault of Threshold comparator circuit is the comparing value appears out "fixed value" constant "1" or constant "0." For this fault model, fault detection circuit design is shown in Figure 6. When the test begins, processor sends two-way signals high and low, which will input to the non-inverting and inverting terminal Respectively After the isolation circuit ,the processor test and record the output of the comparator. Then exchange the two signals sent by the processor, test again and record the output of the signal. If the two comparator output signal does not change, we can say that the comparator goes wrong.

4.2.3 Receiving Coil Signal Processing Circuit Test Algorithm

Receiver coil signal processing circuit includes LC resonant circuit, band-pass filter circuit, the synchronous detection circuit and low pass filter circuit. This part of the circuit is relatively complex, closely associated with each other, so the fault detection and accurate locating is difficult. Here we treat the part as a whole to design fault detection circuit. The fault of the circuit includes the amplification factor offset, the

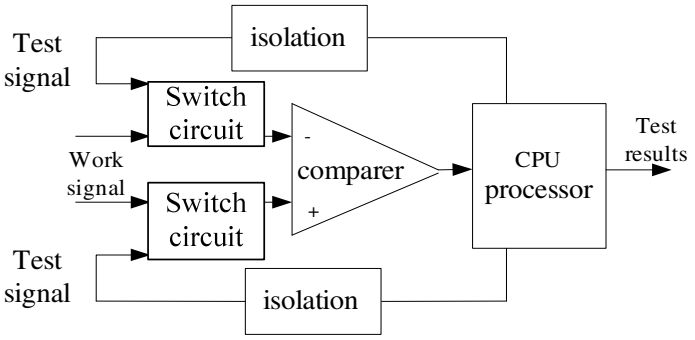


Fig. 6. Compares the schematic circuit fault detection

multiplier output constant value, filter circuit disorders. The input of Signal processing circuit is a sine wave signal, the output DC signal. The fault detection circuit use External incentives signal source, we can get a sine wave signal which is the same amplitude and frequency as the receiving coil signal by method of processor software occurs. The signal generated by the processor will send to the input signal processing circuit by switching way, then we can determine whether the processing circuit fault occurs by detecting the difference of output DC voltage comparing to the voltage under normal circumstances. Fault detection principle in shown in Figure 7.

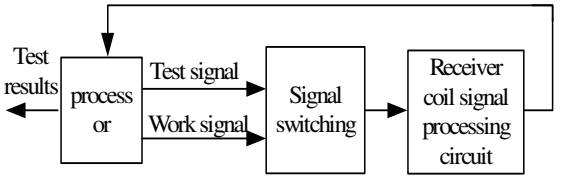


Fig. 7. Receives the signal processing circuit fault detection coil schematic

In practice, the signal processing circuit is the core of the system circuit, fault detection circuit switching will have some impact on the system stability, so the signal processing circuit fault detection is the last steps in the whole system self-test process.

4.3 Serial Communication Module Test Methods

For the communication channel detection algorithm, we use hardware to send signal, then receives the signal by wrap circuit and use software to proofread. The processor connect then input and output of the serial channel directly by controlling the wrapping switch at the beginning of the test, send a test signal through the output, using the signal input receives compared with the output signal the to test the serial channel. Detection principle is shown in Figure 8.

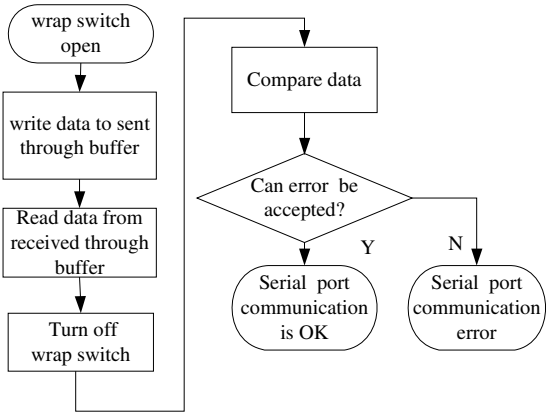


Fig. 8. Serial communication channel failure detection flow chart

Software verify testing method is adding a parity bit in the process of serial communication, which will achieve on-line testing, such as the use of parity method, Hamming parity law and so on. This approach can effectively detect the line fault and do not affect the normal operation of the system when sending and receiving data in data-packets mode. In order to detect the reading code faults, the communication protocol has done some improvements in the sensor. When the sensor works well, the data transmission format is 12-bit data, which is the integration of three indication plate coded information. The data store in two registers with two high bit supplement 1, if the sensor communication unit receives the data which is less than 0xC0, the upper sensor communication unit communication error can be reported.

Serial communication hardware circuit includes optical isolation chip, driver chip and voltage conversion chip, and the communications circuits is in a relatively long signal transmission line, it is vulnerable to outside noise interference, if the circuit interfere by high amplitude pulse, then it is easy to destroyed voltage conversion chip and isolation devices, resulting in communication circuit. As there is more failure point in the communication circuit, the node-scan detection method can be used to achieve its fault location. We need to know all aspects of the circuit firstly, and then analyzed to find the key points of the various parts, according to the current state of the signal in these key points, you can find the failure part then achieve the fault location.

4.4 Sensor Fault Self-detection Algorithm for Code Reading

In addition to the hardware fault detection method, the sensor cols reading fault can also be tested by software. Sign board is positioned on both sides of the track in a certain distance. Take Shanghai high-speed maglev train test line for example, the top 5 Signs of code as 1.

When the sensor of the train continuously passing by the locating sign board in normal circumstances, the coded information read out should meet the relationship table, if the results read out do not match the relationship with the table twice, you can determine the sensor fault, and the sensor fault will be reported. We can amend the fault code according to the reading history combined with the sign board code distribution table.

Table 1. Location board code distribution

Number	Location board distance (meter)	Location board code		
		High-bit	Middle-bit	Low-bit
1	20.124	1001	0001	0000
2	88.236	1001	0001	0001
3	97.524	1001	0001	0010
4	204.852	1001	0001	0011
5	288.444	1001	0001	0100

5 Conclusions

The basic principle of absolute locating sensor of Maglev Train is presented in the paper, and it analyses the issue of the absolute locating sensor in detail. An optimization design of BIT is made in order to deal with the problem. The experiment result shows the system can satisfy the practicable application of the absolute locating sensor of Maglev Train.

References

1. Wen, X.S., Xu, Y.C., Yi, X.S.: Intelligent built-in test theory and application. National Defense Industry Press, Beijing (2002)
2. Tian, Z., Shi, J.Y.: Systems analysis and design verification testing. Beijing Aerospace University Press, Beijing (2003)
3. Wu, X.M.: Maglev train. Shanghai Science and Technology Press, Shanghai (2003)
4. Wu, W.Q.: Rail transport operation control and management. Tongji University Press (2004)
5. Henning, U.: Long Stator Propulsion System of TRAN-SRAPID Berlin-Hanbur. In: Yananashi: The 15th International Conference on Magnetically Levitated Systems and Linear Drives, pp. 274–279 (1998)
6. Liu, H.Q.: German magnetic levitation train Transrapid. University of Electronic Science and Technology Press, Chengdu (1995)
7. Qian, Y.C., Han, Z.Z., Shao, D.R., Xie, W.D.: High-speed maglev train speed and position system design. Computer Engineering 31(3), 12–14 (2005)
8. Long, Z.Q., Li, X.L., Zhou, W.W., Liu, S.S.: Maglev train location and speed research. National University of Defense Technology 25(4), 25–26 (2003)
9. Guo, X.Z., Wang, Y., Wang, S.X.: High speed maglev train locating system. Southwest Southwest Jiaotong University 39(4), 455–459 (2004)
10. Chen, Z.Y.: Maglev Locating System Design. Tongji University, A Master's Degree Thesis 3, 50–59 (2006)
11. Chen, X.Z.: Radar Built-in test (BIT) system design. Electronic Measurement Technology 31(3), 134–137 (2008)
12. Zhou, G.F., Chen, T.F., Cui, X.Q.: Locomotive on-line fault diagnosis expert system. Changsha Railway University 20(1), 105 (2002)

The Tag Estimation and Frame Length Determination with Capture Effect in Dynamic Frame Slotted ALOHA about RFID*

Liu Jing, Wu Haifeng**, Li Yuexun, Tan Yuan, and Deng Zhongting

School of Electrical and Information Technology, Yunnan University of Nationalities, 650500 Kunming, People's Republic of China
{ningning87109,liyuexun810,pllpty,seamild}@163.com
whf5469@gmail.com

Abstract. In order to enhance radio frequency identification (RFID) tag identification efficiency, this paper proposes a Vogt estimation method (Cap Vogt) and an optimal frame length scheme under the capture environment for a dynamic frame slotted ALOHA anti-collision RFID system. There is a growing awareness of the reading leaking problem in tag estimation, the estimation method we proposed solves the problem in a sense. Meanwhile the simulation results of the tag estimation shows the superiority of the method comparing to the conventional method. In addition, this paper derives an optimal frame length for obtaining the maximum radio channel efficiency when the capture effect exists. The optimal frame length always less than the tag number.

Keywords: Radio frequency identification (RFID) anti-collision, dynamic frame slotted ALOHA(DFSA), tag estimate, frame length, capture effect.

1 Introduction

Radio frequency identification (RFID) tag collision is that multiple access problem in natural. So now many scholars consider the anti-collision theory about multiple access into RFID tag anti-collision, and proposed two algorithms: tree algorithm [1] [2][4] and ALOHA algorithm[3][5][7][8][9]. ALOHA algorithm mainly applied to the multiple access because its limited function of mobile station and powerful function of the base station.

Now ALOHA algorithm are widely applied in the RFID tag anti-collision system, because the reader for the active, tags for the passive in the usually RFID system, and there are much less tags. The ALOHA algorithm can be divided into pure ALOHA, slotted ALOHA and dynamic framed slotted ALOHA. The system efficiency of pure ALOHA and slotted ALOHA is much low, so to improve the system efficiency, we proposed the dynamic framed slotted ALOHA. The frame size of which dynamically

* This work is supported by Innovative Research Team in Yunan University of Nationalities, and the Scientific Research Foundation of Yunnan Provincial Department of Education (2010Y423).

** Corresponding author.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 117–123.
springerlink.com

changes according to specific case to ensure the maximum system efficiency. However, the frame size often determined by the tag number, and we usually don't know the tag number. Therefore there are two problems we must consider to improve the performance of dynamic frame slotted ALOHA: (1) Tag estimation; (2)Determine the frame size according to the tag is estimated.

The tag reading leakage problem is a common phenomenon in the tag estimation. This paper proposes a VOGT tag estimation method which is an estimation method under the capture effect. In this method, we also estimate the capture effect existence probability γ. This method is much more accurate and it solves the tag reading leaking problem. Another problem is that determine the frame size according to the tag number to improve the system performance of dynamic frame slotted ALOHA in RFID.

This paper proposes the relationship of tag number and frame size according to the specific situation of the collision slot, idle slot and success slot. This relationship can guarantee the system channel utilization rate reaches a maximum value.

2 System Description and Tag Estimation with the Capture Effect

In the RFID dynamic frame slotted ALOHA anti-collision system, the slot number in an information frame is L, if the number of tags is n, and there no capture effect. The probability of the idle, success and collision in a slot are p_0, p_1, p_k

$$p_0 = \left(1-\frac{1}{L}\right)^n \tag{1-1}$$

$$p_1 = \left(\frac{n}{L}\right)\left(1-\frac{1}{L}\right)^{n-1} \tag{1-2}$$

$$p_k = 1-\left(1-\frac{1}{L}\right)^n - \left(\frac{n}{L}\right)\left(1-\frac{1}{L}\right)^{n-1} \tag{1-3}$$

So if no capture effect, the average slot number of idle, success and collision in an information frame are:

$$a_0 = L\left(1-\frac{1}{L}\right)^n \tag{2-1}$$

$$a_1 = n\left(1-\frac{1}{L}\right)^{n-1} \tag{2-2}$$

$$a_k = L\left[1-\left(1-\frac{1}{L}\right)^n - \left(\frac{n}{L}\right)\left(1-\frac{1}{L}\right)^{n-1}\right] \tag{2-3}$$

But in the realistic systems, the capture effect usually exists. Now if the capture probability is γ ,and the current frame size is L, the value of a_0, a_1, a_k will be:

$$a_0(L,n) = a_0 \tag{3-1}$$

$$a_1(L,n) = a_1 + \gamma a_k \tag{3-2}$$

$$a_k(L,n) = a_k - \gamma a_k \tag{3-3}$$

The traditional Vogt estimation[6] with no capture effect is a method which through comparing the actual and theory slot number of idle, success and collision that results in the minimum error to estimate the tag number, but a serious problem in this method is the leaking reading problem. Which lead to low throughput of the system. When in capture environment, the key idea is to estimate n and γ using the value of (3-1), (3-2), (3-3). The Vogt estimate under the capture effect can be expressed as:

$$\hat{n} = \arg\min_{(\tilde{n},\gamma)\in\Omega} \| A(\tilde{n}) - A \| \tag{4}$$

Where the $\|\cdot\|$ is Euclidean norm, $A = [a_0, a_1, a_k]^T$ is the actual slot number of idle, success and collision under the capture environment.

$A(\tilde{n}) = [a_0(L,n), a_1(L,n), a_k(L,n)]^T$ is the slot number of idle, success and collision under the capture effect and its Probability is unknown. Assemble Ω can be expressed as:

$$\Omega = \{(\tilde{n},\gamma) \mid a_1 + 2a_k \le \tilde{n} \le N, 0 \le \gamma \le 1\} \tag{5}$$

where N is the maximum tag number.

3 Determine the Frame Size

In the RFID dynamic frame slotted ALOHA, there will be more idle slots if the number of slot more than the tags, to the contrast, there will be more collision slots. All these would lead to low identification efficiency. Literature[5][7][8] all think that under no capture environment, only the slot number equal to the tag number can obtain the maximum system throughput.

So in capture environment, the frame size and the tag number should satisfy certain relation. Suppose the capture probability is γ , L is the current frame size and n is the tag number. In realistic RFID systems, the values of a_0, a_1, a_k are provided by the reader by estimating c_0 (idle slots), c_1 (success slots), c_k (collision slots). If

no capture effect, that will be: $a_0 = c_0$, $a_1 = c_1$, $a_k = c_k$; To the contrast, the values will be different, the value same to the (3-1),(3-2),(3-3).

Based on (2-1),(2-2),(2-3),(3-1),(3-2),(3-3) and (4), the system throughput with the capture effect can be written as:

$$P_s = \frac{n\left(1-\frac{1}{L}\right)^{n-1} + \gamma L\left[1-\left(1-\frac{1}{L}\right)^{n} - \left(\frac{n}{L}\right)\left(1-\frac{1}{L}\right)^{n-1}\right]}{L} \tag{6}$$

We can obtain the optimum frame length when the system throughput maximum. Letting the differentiation of (3) to be zero:

$$\frac{dP_s}{dL} = \frac{n\left(1-\frac{1}{L}\right)^{n-2}\left[\gamma + n(1-\gamma) - L\right]}{L^3} = 0 \tag{7}$$

We can derive that the optimum frame length is:

$$L' = \gamma + n(1-\gamma) \tag{8}$$

We can learn that the Optimum frame length decreases when γ increases from 0 to 1, and it always less than the tag number n .

The anti-collision algorithm with capture is summarized in the following table:

For each frame{

initial parameters:

1. $k = 1$ // frame counter
2. $L_k = L_{knk}$ // initial frame length

Repeat:

3. $k = k + 1$ // increase the frame counter
4. Observe a_0 , a_1 , a_k // slots number of previous frame
5. Estimate n_k and γ_k using VOGT estimate under capture effect
6. $L_k = \gamma + n_k(1-\gamma)$ // determine optimum frame length
7. reader-tag interrogation within L_k slots

until $a_1 = 0$ & $a_k = 0$ }

4 Analysis of the Simulation Results

The experimental results of this paper are all obtained by Monte Carlo 50.The performance of Vogt estimation under capture environment is simulated. Figure 1 shows the tag estimation when the capture probability is unknown, it's much more accurate. Fig.2 shows the comparison between the estimation under the capture environment and there no capture effect. We can learn that the error is smaller under

the capture environment than that of no capture effect. So the reading leaking problem is less serious under the capture environment. Figure 3 shows the throughput of the original DFSA without capture effect and DFSA under the capture environment.

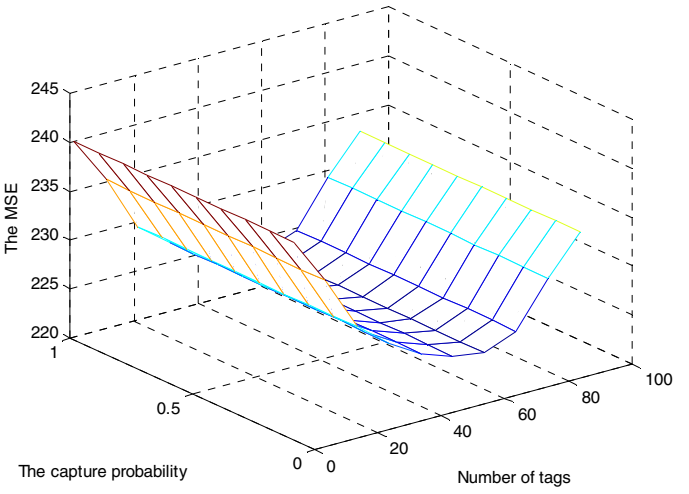


Fig. 1. The Vogt estimate error under the capture environment

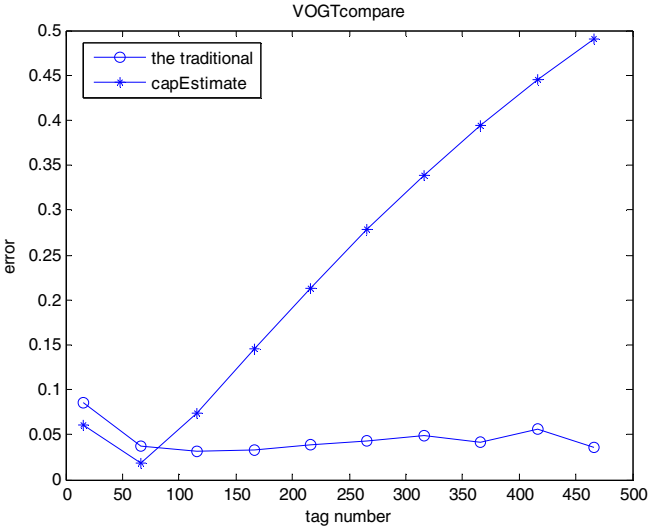


Fig. 2. The estimation error under the capture environment comparing with when no capture effect

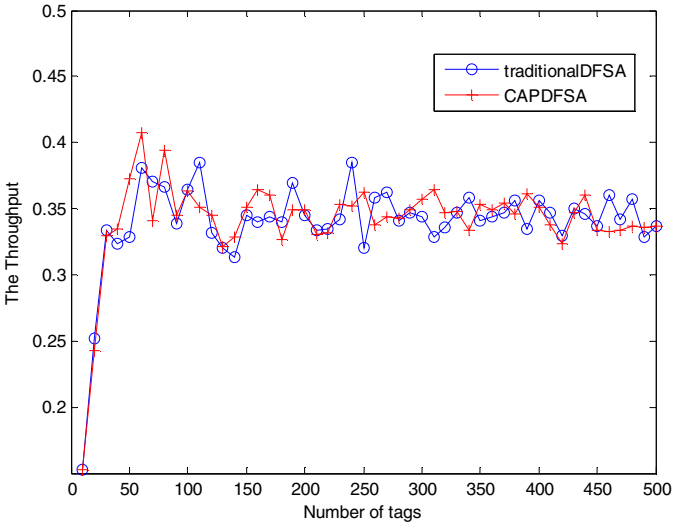


Fig. 3. The optimum throughput of DFSA under the capture environment is higher than that of no capture effect

5 Conclusions

This paper proposes a Vogt estimation method under the capture environment, and we also put forward the optimum frame length in dynamic frame slotted ALOHA with the capture probability. The simulation results show that the method of estimate capture probability is much more accurate than that of no capture effect.

References

1. Capetanakis, J.I.: Tree algorithms for packet broadcast channels. IEEE Transactions onInformation Theory 25(5), 505–515 (1979)
2. Hush, D.R., Wood, C.: Analysis of tree algorithm for RFID arbitration. In: Proceedings of IEEE International Symposium on Information Theory, p. 107. IEEE, Cambridge (1998)
3. Myung, J., Lee, W., Shih, T.K.: An adaptive memoryless protocol for RFID tag collision arbitration. IEEE Transactions on Multimedia 8(5), 1096–1101 (2006)
4. Lai, Y.C., Lin, C.C.: A pair-resolution blocking algorithm on adaptive binary splitting for RFID tag identification. IEEE Communication Letters 12(6), 432–434 (2008)
5. Schoute, F.C.: Dynamic frame length aloha. IEEE Transactions on Communications 31(4), 565–568 (1983)
6. Vogt, H.: Efficient Object Identification with Passive RFID Tags. In: Mattern, F., Naghshineh, M. (eds.) PERVASIVE 2002. LNCS, vol. 2414, pp. 98–113. Springer, Heidelberg (2002)

7. Floerkemeier, C.: Transmission control scheme for fast RFID object identification. In: Proceedings of the 4th Annual International Conference on Pervasive Computing and Communications Workshops, pp. 457–462. IEEE, Washington D.C., USA (2006)
8. Cha, J.R., Kim, J.H.: Novel anti-collision algorithms for fast object identification in RFID system. In: Proceedings of the11th International Conference on Parallel and Distributed systems, pp. 63–67. IEEE, Washington D.C., USA (2005)
9. Chen, W.T.: An accurate tag estimate method for improving the performance of an RFID anti-collision algorithm based on dynamic frame length ALOHA. IEEE Transactions on Automation Science and Engineering 6(1), 9–15 (2009)

Research the Power Enterprise Data Warehouse Modeling Technology Based on Business Intelligence

Zhao Zhaolin

Department of Computer and Information Science,
Fujian University of Technology, Zipcode 350108,
Fuzhou, Fujian, China
zhaozl@fjut.edu.cn

Abstract. The paper researched the electric power enterprise data warehouse modeling technology based on business intelligence. Application of business intelligence and data warehouse modeling techniques can create a data warehouse model which can support user's decision and analysis .The model can help users find the law from the data, predict trends , assist the user to make the right decisions and guide the organization's development through reports, random queries, OLAP and other functions. The use of the model for the power enterprise can improve management level, promote the standardization and scientific, provide reliable historical data for business decision-making, ensure the feasibility of decision making, strong competition, and achieve concept of business intelligence applications.

1 Introduction

Management information systems for the development of electric power enterprise has brought tremendous boost. But a variety of data and reports in the existing management information systems can not meet the current demand, an urgent need for business activities from which to explore the laws and market trends, and get the decision-making information of compete the market. In this demand driven, the power enterprise require the use of business intelligence and data mining technology to build the data warehouse model, and meet a variety of unexpected the leaders demand inquiry in time. So the leadership get the desired data in time , decision-making related to accurate and reliable information, reduce decision-making costs , improve the efficiency of decision making , improve the overall competitiveness. At present, it is feasible that the applications of data warehouse technologies in the electric power enterprise because the construction of electric power enterprise information system resources has been massive data.

The data warehouse modeling based business intelligence is the the basic link for the electric power enterprise data warehouse construction. A good data model can help the power enterprise to make full use of data in the data warehouse, make comprehensive application and analysis Flexibility and timely. Therefore, the power enterprise can ensure the feasibility of the decision-making, reduce the decision-making costs, improve decision-making efficiency, and enhance enterprise competitiveness.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 125–132.
springerlink.com

2 Research of Business Intelligence and Data Warehouse Modeling Techniques

2.1 Applications of Business Intelligence Technology in Electric Power Enterprise

Business Intelligence (Referred to as BI) is designed for enterprise data access, management, analysis, display and other uses of data support systems. After BI is aplied, the performance of enterprises is improved significantly. The key of BI is extract useful data from the operations support system, network management system, customer relationship management system, financial management system and many different operating systems. The data is cleaned to ensure data accuracy, then through the extraction, transformation and loading into a enterprise-class data warehouse, so a global enterprise data view is obtained. On this basis, the appropriate query tool, data mining tools and OLAP tools are used for analysis and processing, and finally the knowledge presented to managers for decision-making [1].

The electric power enterprise data warehouse based on BI apply large amounts of data which the power companies in all areas of everyday business applications generate. The significant added value for corporate data decision-making is brought because of the analysis based on these data, such as an independent market research, evaluation results, evaluation of external consultants.

2.2 Research of Data Warehouse Data Modeling Techniques

There are based on 3NF data model, star schema and snowflake models [2]for data warehouse data modeling.

The Normal Form is the basic theory for relational data model. The process of a relational model from first to fifth for lossless decomposition is called normalization [3]. Currently in the data warehouse modeling commonly use 3NF, which has a very strict mathematical definition. The 3NF for relational data model must meet the following three conditions: (1) a unique value for each property, it does not have the ambiguity; (2) each non-primary property must be fully dependent on the primary key; (3) each non-primary attributes can not rely on the properties of other relations. The 3NF is the definition of the relationship between non-primary key and the main attributes. If only the first condition is satisfied is referred to as 1NF; if the first and second conditions are referred to as 2NF, and so on.

Although the data warehouse model is based on the standardized data model, standardized data model for data warehouse practice is not ideal. In the environment of a fully normalized, the data in the tables is relatively small, and the processing of these tables need the application program to be dynamically interconnected operation, which requires many I/O operations between the different tables. The time cost is too high for these multi-table joinning operation, and it is very negative for the decision-making efficiency. Therefore, the data warehouse requires a non-standardized data

processing, in order to reduce the demand for tables joinning, improve performance for the data warehouse, and improve query efficiency.

Star model is a multidimensional data model, which consists of a Fact Table and a set of Dimension Table component. Each Dimension Table has a dimension as a primary key, all of these dimensions are combined into the Fact Table's primary key.In other words, the each primary key of Fact Table is the dimension table foreign key. The non-primary attributes in the Fact Table are called the Fact, which are calculated values generally, and mostly dimensions are text, time and other types of data. The characteristics of star model are as follows: (1)existing a large number of preprocessing, query speed is very fast;(2)the modeling process is relatively slow, existing large amount of data redundancy, and the hardware requirements is high. [4]

Snowflake model is an extension to star schema. Design star model to determine the target entity and dimension entities in the concept model. In order to analysis a more in-depth the relevant dimensions, the physical dimensions of the star model increase in-depth detailed instance of the class, so snowflake model is designed. Essentially snowflake model is the star model in the dimension table for the standardized processing. Snowflake model will bring good ability for the subject's queries response, but the snowflake model will bring the difficult of different tables joinning operation.

3 The Data Warehouse Data Modeling Process in the Electric Power Enterprise Based on Business Intelligence

According to these studies, combined with the feature that hardware and infrastructure are good in the electric power enterprise, the paper uses star model as the electric power enterprise data warehouse model. It not only provide readily accessible, but also provide to screening, combined and cut for a large number of physical indicators. In this paper, the analysis of generate electricity and power supply in the field of power grid operation of the electric power companies is as an example of the data warehouse data modeling process.

3.1 Concept Model Design

Accord to the actual situation, The electricity enterprise data warehouse is divided into six basic areas: human resource field, finance field, power grid operation field, material supply field, plan making field, marketing of electric power field. The various subject field model can be builded, subject analysis of electric power grid operation field as shown in Figure 1.

After the subject field to determine, the main problem of concentrated in the data warehouse is how to analyze data, and find out the potentially valuable data.

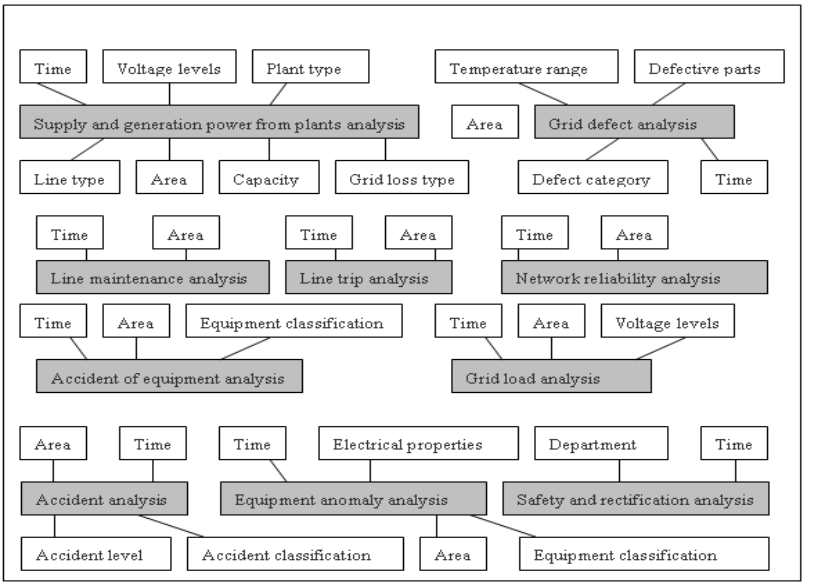


Fig. 1. Subject analysis of electric power grid operation field

Table 1. Information package diagram of generate electricity and power supply analysis in electric power enterprise

Dimension definition	Date	Time	Voltage levels	Grid loss type	Supply area	Plant type
Class definition	Year	1 hour	500KV	Main grid loss	Area 1	Thermal power plants
	Quarter	A half hour	220KV	500KV grid loss	Area 2	Hydropower plants
	Month	5 Minute	110KV	220KV grid loss	Area 3	Local Power Plants
	Day	1Minute	…	…	…	…
Indicator Definition	Regional rate of force= Active power/ Reactive power Network loss rate=(Grid electricity- Used power)/100/ Grid electricity…					

Information package diagram can further refine the concept model. It is a information packaging technology for the analysis subject domain, which reflects the multi-dimensional data inside the computer in aggregation, and reflects the conceptual meaning of all the different information aggregation platform, that include definition of indicators, defined dimensions, and define the category etc, Information package diagram of generate electricity and power supply analysis in electric power enterprise as shown in Table 1.

3.2 Logical Model Design

The data need to be converted from concept data model to logical data model . Basic relation model of logical data of generate electricity and power supply analysis in electric power enterprise as shown in Figure 2.

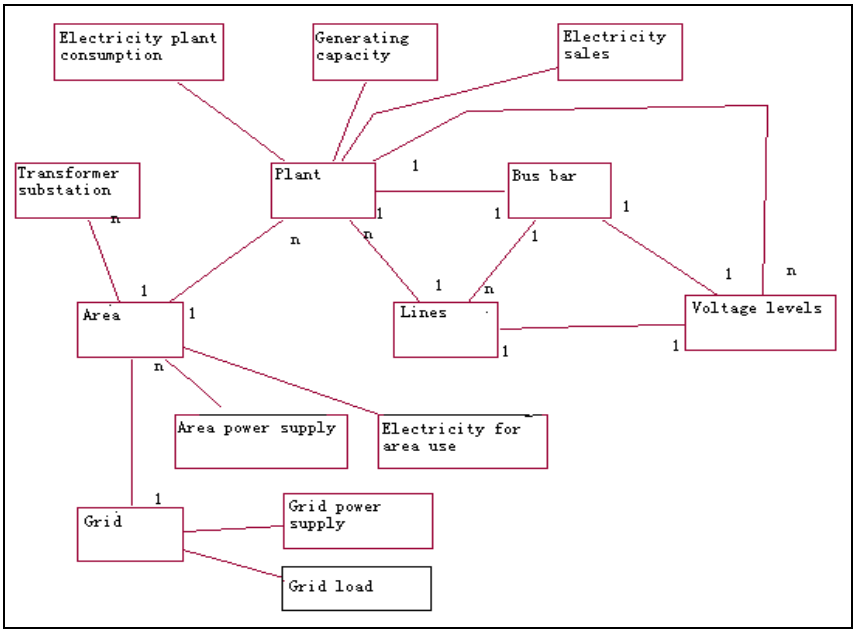


Fig. 2. Basic relation model of logical data of generate electricity and power supply analysis in electric power enterprise

It is the purpose of the logical data model that data is queried based on a subject. Star model will involve each data of a subject domain analysis together, and the related factors of the analysis of any subject in electric power enterprise are likely to come from different DBS. Therefore, it is necessary that the basic relationship between the logical data in the analysis of a subject is summary[5]. After the overall subject of a basic relationship logical data analysis is complete, the star model diagram of a subject analysis can be designed. The star-scheme diagram of generate electricity and power supply analysis in electric power enterprise as shown in Figure 3 [6].

3.3 The Division of Granularity Level

The detail data and the data in pasted 5 years are stored in the electric power enterprise data warehouse. If the historical data more than 5 years, it will store the backup storage device. So the amount of data in the data warehouse can be estimate based on 1 year's data. After data warehouse tables storage space is calculated, the way of granularity level division can be determine. Generally if the number of rows of data table

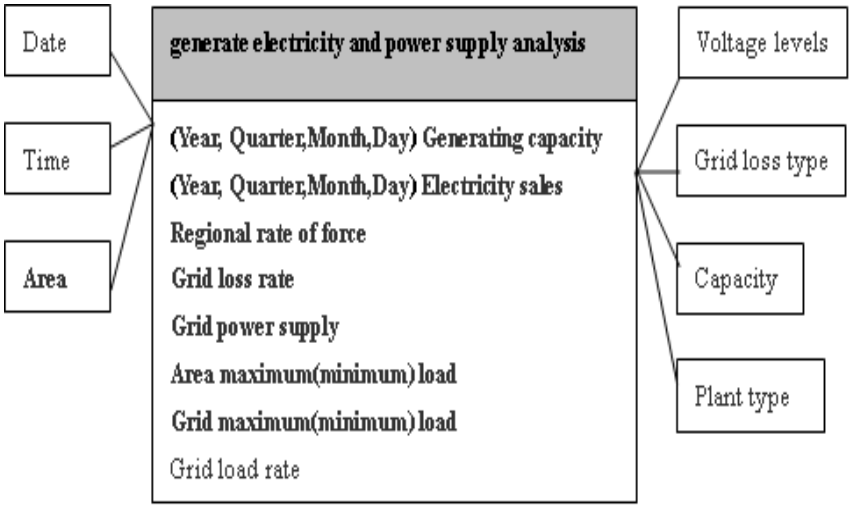


Fig. 3. The star-scheme diagram of generate electricity and power supply analysis in electric power enterprise

in the first year reached 100,000 lines or so, it should consider adding a comprehensive level. If the number of lines in the first year reached more than 1,000,000 lines, then multiple data granularity will be used.

For example, the amount of generation and distribution in the electric power enterprise is about 32 million lines, and the analysis of power generation and distribution of electricity take for account some factors such as season, weather, peak electricity .So granularity level of power generation and distribution data must be rich in some. Different way of data stored determine the amount of storage, while a direct impact on query response time.

3.4 Global Logic Data Model Design

Based on the above analysis, a situation analysis of the fact table and dimension tables in parallel can be created based on the an global OLTP environment.Then the whole logical model of generate electricity and power supply analysis in electric power enterprise as shown in Figure 4.

In summary, the data in data warehouse come from the most details data. Details data is the sources for a subset of data and summary data. The difference of details in the data warehouse environment and in the operational environment is as follow:the details in the data warehouse include only the necessary data to support decision making, and include time properties. The data in the data warehouse data is derived data and summary data based on non-standardized processing which are added because of requirements of business processes and query. For example, the "near future generation" and "regional total power generation" add into the data warehouse, query performance can be improved. In the data warehouse, the details of data storage time is shorter, and general data storage a long time.

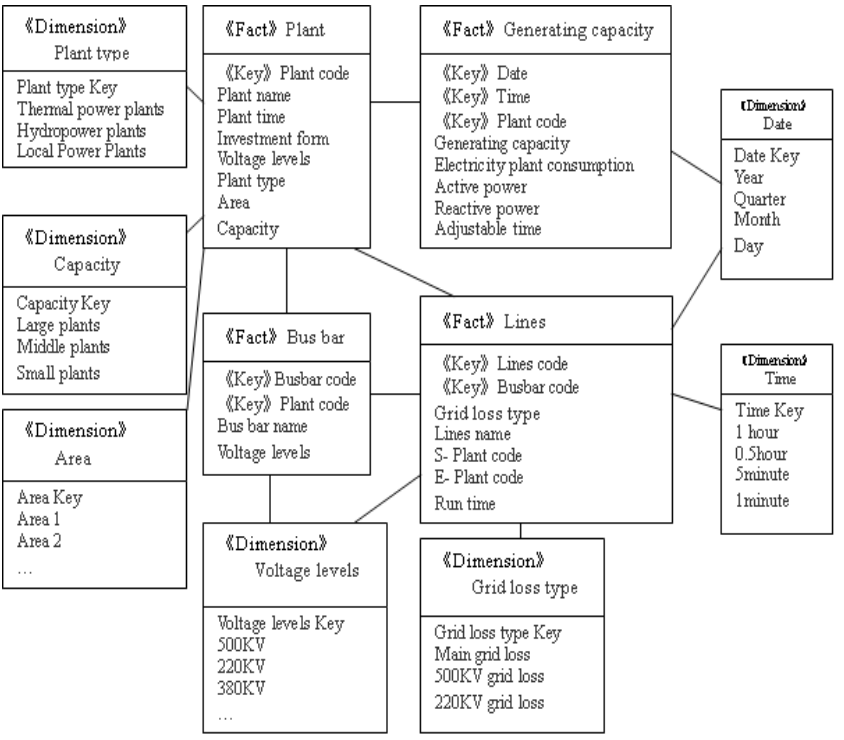


Fig. 4. The whole logical model of generate electricity and power supply analysis in electric power enterprise

4 Application Value

The data warehouse modeling based business intelligence is the the basic link for the electric power enterprise data warehouse construction. According to these studies, combined with the feature that hardware and infrastructure are good in the electric power enterprise, the paper uses star model as the electric power enterprise data warehouse model. The data warehouse based on the model in applications have the following benefits: (1) ensure internal and external data consistency. Marketing data, large customer data, statistical data, measurement data, call center data and other data are centralized management for the electricity companies to provide a unified data view.(2) raise the level of data standardization. The confusion of data management is changed in the past by the construction of the system. (3) improve the overall management level. With the intensive use of data warehouse based on the model by all staff in the electric power enterprise, all business processes are more standardized and standardized .So that labor productivity and utilization are improved. At the same time ,the competitiveness ,management quality and operational efficiency of power companies are improved. (4)the demand of the electric power enterprise users can be better treatmented, and detect problems and treatment more timely.(5) improve

reliability and reduce line losses. So more power to the community for good service, and promote of social productive forces.

The electric power enterprise data warehouse data model based on business intelligence provides flexibility to users. The reports, random queries, OLAP and other functions help users find the law from the data, predict trends , assist the user make the right decisions and guide the organization's development. The use of the model can improve management level, promote the standardization and scientific, strong competition for the electric power enterprise. A power company's concept of business intelligence applications such as "fast to meet customer demand, once success "will be achieved.

References

1. Zheng, H.: The construction of BI system of China's telecom enterprise comprehensive. Communication World, Beijing (2006)
2. Chen, J.: Data warehousing and data mining technology, pp. 145–173, 231–243, 385–398. Electronic Enterprise Press, Beijing (2002)
3. Liu, Y.-S.: The modern database technology, pp. 15–52. National Defence Enterprise Press, Beijing (2001)
4. Chen, Y.: Data Warehouse technology and its applications, pp. 17–19, 43–57. Dalian Maritime University Press, Dalian (2002)
5. Giovinazzo, W.A.: Object-oriented data warehouse design. Xiaoxiang studio translation, pp. 71–133. Posts and Telecommunications Press, Beijing (2000)
6. Schmuller, J.: UML-based, case and application, pp. 24–33. Posts and Telecommunications Press, Beijing (2002)

Adaptive Real-Time Communication Scheme for Mobile Robot Control

Fang Guo and Jiayong Duan

Department of Electronics and Information Engineering,
North China Institute of Science and Technology,
East Yanjiao, 101601, Beijing, China
{Guofang1980,dl020702}@ncist.edu.cn

Abstract. An adaptive real-time communication scheme is proposed in this paper. The dynamic configuration of system is achieved through analyzing the success rate of real-time transmission under various conditions. The proposed scheme is applied to control the mobile robot with visual system and manipulator. Through the application of this scheme, the memory-greedy visual system inter-connects with robotic/manipulation control system harmoniously.

Keywords: Real-time communication, robotic control, adaptive configuration.

1 Introduction

In robotics and manipulator control systems, real-time communication plays an important role and has received widely attention in the last decades. Except for many specialized technologies which require specific hardware for either network nodes or infrastructure components, Ethernet with real-time extension has been developed to control the robotic manipulators since its publication in the 802.3 standard. The utilization of this low-cost, reliable and efficient technique has received widespread attention, and much research effort has been shown in real-time robotic manipulator control over Ethernet recently [3][5].

When more and more manipulators are mounted on the mobile robotic platform recently, the demand for real-time control of manipulator has become a critical issue since the mobile functionality of the robot challenges the reliability and efficiency of the system over various view points: 1) The limitation of computation power, which is a scarce resource on a mobile robot, restricts the utilization of high power-dissipation hardware. 2) The mobile robots usually fuse information resources from multiple sensors (e.g., visual information from camera, spatial information from laser and linguistic information from microphone), the exhaustive information brings huge memory transfers thereby raising violation of real-time communication constraints. 3) Once the robot starts moving, all the tasks turn to time critical since the temporal spatial relationships of the perception demand the reflection of tasks in a fixed time.

The rest of the paper is organized as follows. In section 2 we introduce the background and review the state-of-the-art real-time robotic manipulation control. Section 3 describes the configuration of our robot platform and details how to

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 133–136.
springerlink.com

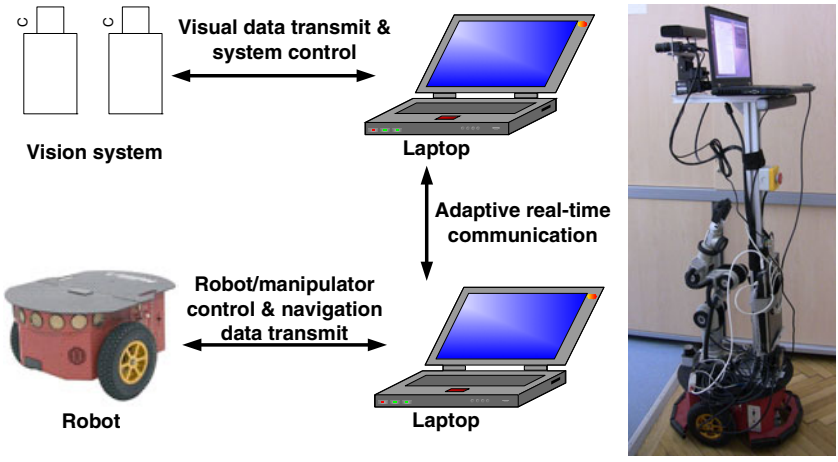


Fig. 1. Left: experimental scenario setup, right: test robot controlled by two laptops, the laptop on the top bracket connects to the vision system using usb and firewire, the laptop on the below bracket controls the robot and Katana 450 robotic arm through usb connections. Two laptops are connected via network interface using a cross connected cable.

autonomously reconfigure the parameters to satisfy the real-time communication under various conditions. A conclusion is given at the end of the paper.

2 Related Work

Increasingly numbers of devices (such as laser scanner, stereo cameras, RGB-D camera, sonar transmitter and olfactory device) have been mounted on the state-of-the-art mobile robots as information resources to obtain versatile functionalities for performing diverse tasks. The complex data process streams have raised new challenges in terms of the real-time performance and robustness of the system. The vision data, as the most important and popular parts in all the information resources, has been utilized for triggering various events or controlling the robotic manipulators [1][2][4]. Since the visual information process usually involves large memory usage thereby influencing the system performance, a specific real-time communication scheme which considers influence of memory transfer behavior is required.

Kiszka [3] developed a Real-Time Driver Model (RTDM) to unify the interfaces for developing device drivers, thereby processing real time control/communication with various devices (e.g. serial, image devices and CAN protocol stack). However, the pressure measuring which tests the performance of system under high-load conditions is neglected.

Palli [5] et al. introduced a real-time control system design for robotic wheelchairs based on RTAI, and evaluated the performances of the system with particular attention to the execution time of the various tasks, communication delays due to network communication etc. However, their system is not adaptive to the dynamic system burden.

Zhou [6] et al. have provided clearly evidences that RTAI and RTnet are capable of meeting deadlines under high CPU loads. However, they didn't give a possible solution to the invalidation of real-time constraints caused by memory transfers exceeding CPU cache.

3 Real Time Communication Based on RTAI

An additional protocol layer named RTmac has been utilized in RTnet/RTAI to avoid collisions on the Ethernet, and it adopts the Time Division Multiplexing Access (TDMA) process to handle the transmission media access control. The synchronization of network nodes is required for the correct slot timing. It is achieved by defining one node in the network as master, and the remaining nodes as slaves will adjust their clocks with respect to the *Start of Frame* (SoF) packets which are periodically broadcasted by the master node.

The duration between two consecutive SoFs is called synchronization round. In every synchronization round, arbitrary amounts of slot could be assigned to each node. The medium accessibility of each node can be ensured only in the specific assigned slots of each node.

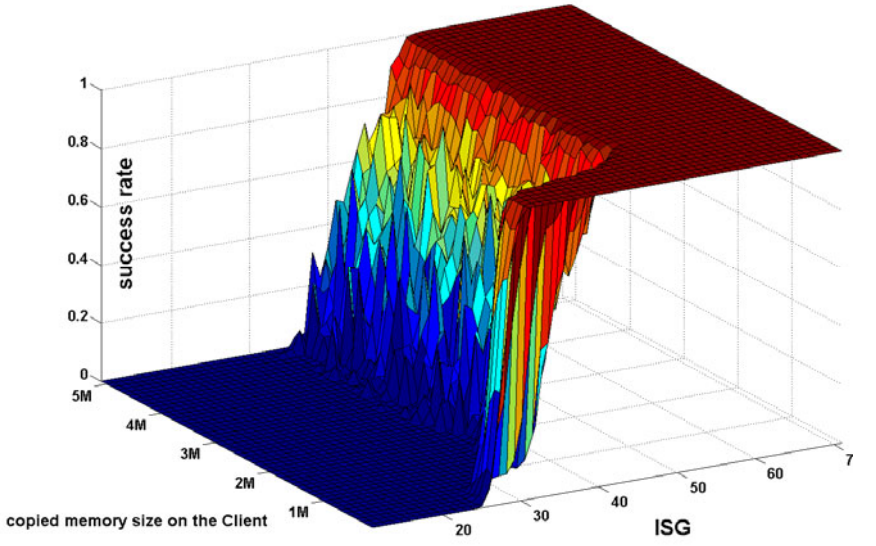


Fig. 2. The success rate of real-time transmission under various ISG and memory transfer size

As commonly known, the *inter slot gap* (ISG) which specifies the time between two consecutive slots is an important criterion to measure the performance of TDMA based network. The lager ISG implies that fewer data can be transmitted because there are no data transmission behaviors during the ISG. However, if the ISG is set too small, nodes may have no enough time to process the received data thereby failing

the transmission task in their assigned slots. Essentially, the ISG has to be set probably, and it should large enough to allow every node have sufficient time for completing local tasks (receiving and processing), and it should also as smaller as it can be for achieving the maximum network utilization. To this end, we measure the real-time performance of RTAI based Linux system under the configurable ISG and various memory transfer. The experimental set up is similar with the configuration in [5] and we will derive our adaptive real-time control scheme through analysis of the experimental results.

The setup of test scenario is demonstrated in Fig.1. We define the value of ISG which can guarantee the successful communication as *guaranteed success value* (GSV). The GSV is set as follow,

$$GSV = \begin{cases} ISG_l, & Mem < T \\ ISG_h, & Mem > T \end{cases} \tag{1.1}$$

Where T is a threshold and when the transferred memory exceeds this threshold, a breaking jump of ISG can be observed (see Fig. 2). Hence, the ISG can be configured adaptively according to the varying of memory transfer size. This adaptive configuration obviously can enhance the channel utilization and data throughput.

4 Conclusion

An adaptive configuration of real-time communication scheme is proposed in this paper. The dynamic configuration is realized through analyzing the success rate of real-time transmission under various ISG and memory transfer size. The proposed real-time communication scheme is applied to control the mobile robot with vision system and manipulator. Through the application of this scheme, the vision system though is memory greedy, can communicate with robotic control system in real time.

References

1. Zhou, K., Richtsfeld, A., Zillich, M., Vincze, M.: Coherent Spatial Abstraction and Stereo Line Detection for Robotic Visual Attention. In: 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2011) (September 2011)
2. Zhou, K., Richtsfeld, A., Zillich, M., Vincze, M., Vrecko, A., Skocaj, D.: Visual Information Abstraction for Interactive Robot Learning. In: The 15th International Conference on Advanced Robotics (ICAR 2011), Tallinn, Estonia (June 2011)
3. Kiszka, J.: The real-time driver model and first applications. In: The 7th Real-Time Linux Workshop, Lille, France (2005)
4. Zhou, K., Zillich, M., Vincze, M.: Multi-model Fitting Using Particle Swarm Optimization for 3D Perception in Robot Vision. In: 2010 IEEE International Conference on Robotics and Biomimetics (ROBIO 2010), Tianjin, China (2010)
5. Palli, G., Biagiotti, L., Melchiorri, C.: An Open Source Distributed Platform for the Control of the PUMA 560 Manipulator. In: The 9th Real Time Linux Workshop, Linz, Austria (2007)
6. Zhou, K., Breinbauer, B., Rausch, T.: Violations of Real Time Communication Constraints caused by Memory Transfers exceeding CPU Cache Limits in RTAI and RTnet. In: 5th IEEE International Conference on Industrial Informatics (INDIN 2007), Vienna, Austria (2007)

Design of High-Speed Image Transfer Interface Based on Link Port

Wang Wei, Wang Chunping, and Fu Qiang

Ordnance Engineering College, Shijiazhuang, 050003

Abstract. In current image processing system, the image processing speed is restricted because the data transmission capacity can not meet the requirements of real-time, so a high-speed image transfer interface based on link port is designed to solve the problem in this paper. By researching the link port communication protocol of the TS201, this paper actualized the interface on FPGA, and it made a hardware implementation. The results show that: the FPGA-based link port interface could work in the clock frequency of 200MHz, and could accurately communicate with the TS201 link port in a high speed. The realization of the FPGA-based link port interface provides a good connection mode for high-speed image data transmission between TS201 and FPGA, with a wide range of applications.

Keywords: link port, FPGA, high-speed image transmission, TS201.

1 Introduction

With the rapid development of digital signal processing technology, FPGA + DSP has become a common combination in digital image processing. This combination not only meet the requirements of the system performance, but also has a great cost advantage compared with the traditional application specific integrated circuit (ASIC). FPGA devices are widely used in various occasions because of its high integration, powerful, field-programmable, etc. AD Corporation introduced a new generation of TigerSHARC series processors TS201, with up to 600 MHz operating speed and 4 high-speed data communication link ports. These high-performance processors have been widely used in communications, military, aerospace, medical and other fields. At present, the amount of data in digital image processing is increasing and the increasing of signal processing means, making the requirements of speed in real-time signal processing of large amounts of data transmission is ever increasing, the traditional RS422, RS485 and other protocols have been difficult to meet the real-time demand. Therefore, high-speed data transmission has become a bottleneck in extending the FPGA + DSP applications scope. The TS201's link port has high transmission frequency, high noise rejection capability, small electromagnetic interference, low power consumption, etc. Design a link port interface circuit on FPGA to implement link port communication between FPGA and DSP, is an effective way to improve the data transmission speed.

The paper is organized as follows. Section 2 introduces the link port communication protocol of the TS201. Section 3 actualizes the interface on FPGA. Some simulation and test results are presented in Section 4. The conclusions are reported in Section 5.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 137–142.
springerlink.com

2 Link Port Protocol

Link port is a high-speed communication channel based on LVDS (low voltage differential signaling) technology, designed for the TS201 by AD Corporation. LVDS has higher transmission speed, longer transmission distance which makes link port has the many advantages: high transmission frequency, high noise rejection capability, small electromagnetic interference, low power consumption, etc. TS201 has 4 link ports; each port has two independent channels that can operate simultaneously. The link port communicates through a 1or 4-bit data bus per direction, using four control signals: lxclkoutp / n, lxacki, lxclkinp / n and lxacko; lxbcmpo and lxbcmpi signals are used to signal that the current buffer transmission has been completed. In applications, TS201's link port can work in 4-bit transfer mode or 1-bit serial transmission mode by setting the relevant register.

There are some general rules that apply to the timing in the link port protocol:

1) First data (1-bit or 4-bit) is transferred on the rising edge of the link port clock;
2) Last data (1-bit or 4-bit) is transferred on the falling edge of the link port clock;
3) When link port is idle lxclkoutp signal is driven low;
4) The minimum granularity of the transfer is a quad word; it is to say 128 bits. A quad word is transferred in 16 clock cycles (4-bit mode) or 64 clocks (1-bit mode).
5) Each quad word transmission follows the rule that low bit is transmitted first. Link port transmission timing diagram is shown in Figure 1 and Figure 2.

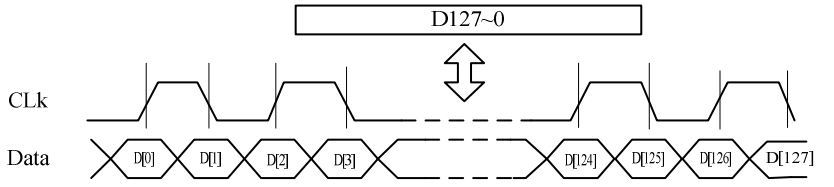


Fig. 1. 1-bit mode of data transmit

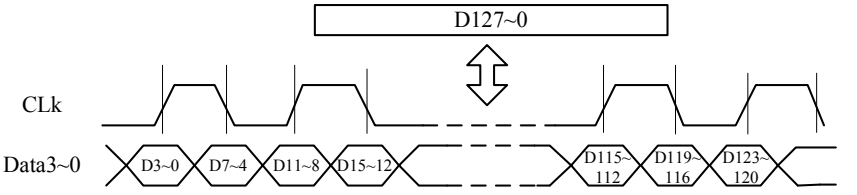


Fig. 2. 4-bit mode of data transmit

When the receive buffer is empty, transmit channel detected lxacki is high, it begin sending data. The first data is effective before the rising edge of the first lxclkoutp and the last data element is transmitted before last clock falling edge. lxclkoutp is driven low when the link is idle.

3 Link Port Actualize on FPGA

Corresponding to TS201's link port with independent receive and transmit channels, FPGA need to use different transmit circuit and receive circuit for communicating with the TS201. To improve the design quality and efficiency, this paper used a modular design approach to complete link port transmission.

3.1 The Transmitter Design

The transmitter consists of the core module, the clock configuration module, control module and FIFO module.

The core module is responsible for serial and parallel data conversion and data DDR transmit and receive, this module provides the core function of link port. The core module needs three clock inputs: clk, clk270 and clk4. The clock clk and clk270 are at the same frequency of link port clock, and the phase between clk and clk270 is 270°. Frequency of clk4 is 1/4 times clk frequency, as internal reference clock of the core module.

The clock configuration module provides the clock inputs for the core module. FIFO module is responsible for data buffering, this paper introduces the IP core of Quartus II, and set FIFO depth is 32 bits. The role of the control module has two, one is to control the front FIFO read request signal, and the other is to control the back buffer FIFO write request signal.

The transmitter works as follow: after initialization, the image data in the FPGA's SDRAM are written into the transmitter front FIFO, until the FIFO is filled, and when acki signal is valid, these data will be written into the core module to convert data, according to link port protocol, then the converted data are transmitted to DSP through link port. The flow chart is shown in Figure 3.

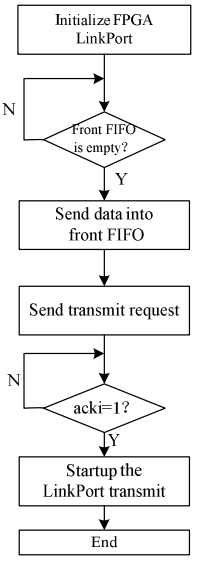


Fig. 3. The transmitter flow chart

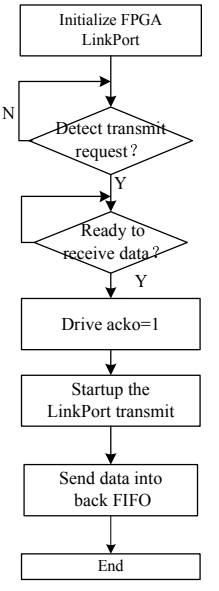


Fig. 4. The receiver flow chart

3.2 The Receiver Design

The receiver consists of the core module, the clock configuration module, control module and FIFO module, too. The core module provides the core function of link port, and other modules are responsible for clock, control logic, data buffer, etc.

When DSP send data to FPGA through link port, data are written into the receiver buffer FIFO, then the core module converted these serial data to parallel data, and write the parallel data to the back FIFO. The receiver flow chart is shown in Figure 4.

The receiver core module use rdclk as a local clock which is asynchronous with link port clock inclk, but its frequency must be between 2/5-2/3 times inclk frequency, generally select 1/2.

This design makes use of IP core of Quartus II as back FIFO. The write request signal of back FIFO is controlled by the control module. Write clock and read clock of back FIFO are separating without synchronization.

4 Results

TS201 link port clock can be configured as 1/1 core clock, 1/1.5 core clock, 1/2 core clock, 1/4 core clock ,so the frequency of TS201 link port clock can up to 600MHz. But FPGA link port maximum transmission frequency is only 200MHz because of the FPGA timing. In this paper, we test the FPGA receiver and transmitter link port at the clock frequency of 200MHz. Simulation use 1-bit mode. Figure 5 shows the transmitter timing simulation; Figure 6 shows the timing simulation of the receiver. The timing diagram shows: the receiver and the transmitter can accomplish communication correctly, meet link port protocol well, and transfer functions can be realized.

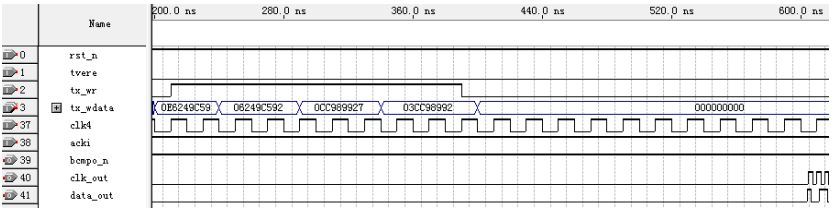


Fig. 5(a). The transmitter emulation in 1-bit mode (data write to FIFO)

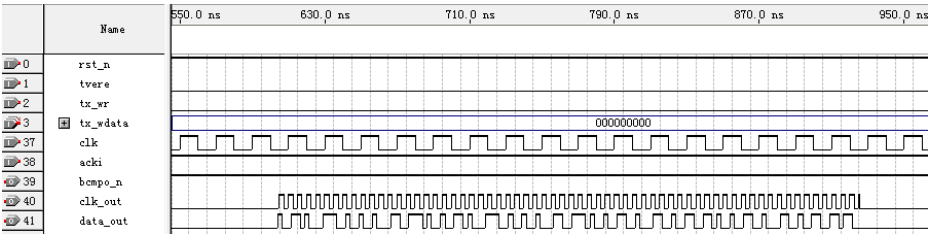


Fig. 5(b). The transmitter emulation in 1-bit mode (link port transmit data)

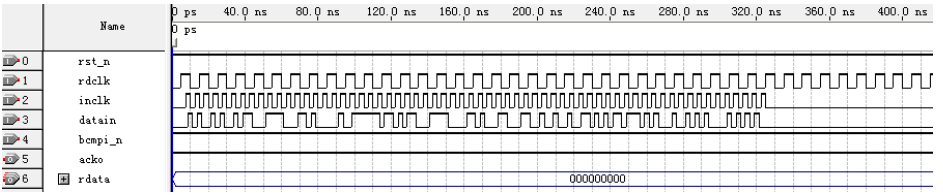


Fig. 6(a). The receiver emulation in1-bit mode (link port receive data)

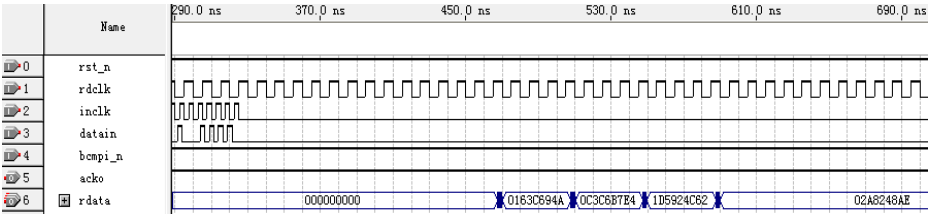


Fig. 6(b). The receiver emulation in 1-bit mode (data read from FIFO)

Based on the simulation, the design of high-speed image transmission interface is test on a hardware platform. The hardware platform consists of TS201 development board and FPGA processing board that links with link port, the platform is shown in Figure 7. Authentication method: TS201 sends an image (600 * 459) from its external memory to the FPGA SDRAM though link port, then FPGA send the image back to TS201 SDRAM though link port. The result of the transmission is shown in Figure 8. The testing image has 256 gray-scale, link port clock frequency is 200 MHz, so the transmit speed is 400MBit / s, the interface could send and receive an image at about 5.5ms, could meet up to 100 frame/ s image transmission task.

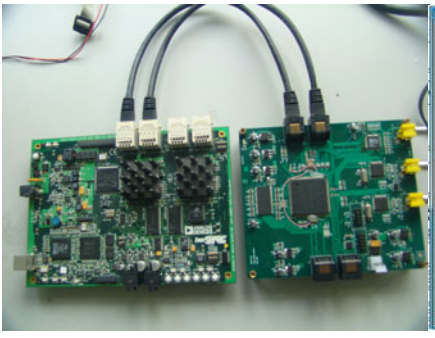


Fig. 7. Hardware platform

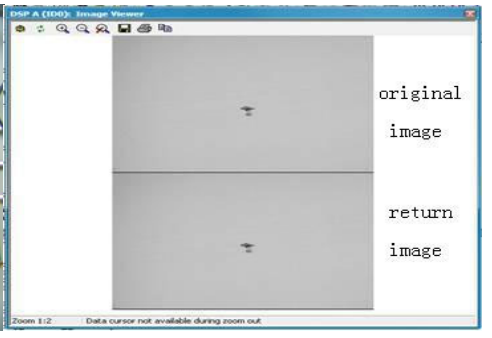


Fig. 8. The results of experiment

5 Conclusions

This paper introduced TS201 link port protocol in detail, designed FPGA-based link port transmit and receive modules, and completed the simulation and hardware

interface test, actualized the high-speed image transfer interface finally. The results show that the design implemented link port communication between TS201 and FPGA, the one-way data transfer speed up to 400MBit / s, data transfer is stable and error-free, solved a digital image processing contradictions of transmission speed and real-time. The realization of the FPGA-based link port interface makes transmission speed meet the real-time demand, with a certain application.

References

1. Liu, S., Luo, Y.: ADSP TS20XS TigerSHARC® Processor Theory and Application Design, pp. 184–341. Publishing House of Electronics Industry (2007)
2. Link-Port Reference Design, Altera Corporation, pp. l–10 (2003)
3. Cyclone II Device Handbook, Altera Corporation (2005)
4. ADSP-TS201 TigerSHARC Processor Hardware Reference, Analog Devices, Inc. (2004)

A Facial Modeling and Animation System Based on Single Image

Zhang Mandun[1], Qiao Yan[1,2], and Wang Yangsheng[2]

[1] School of Computer and Science and Engineering,
Hebei University of Technology,
Tianjin 300401
[2.] Digital Interactive Media Laboratory, Institute of Automation,
Chinese Academy of Sciences,
Beijing, 100190
zhangmandun@126.com, someoneqy@sina.com,
yangsheng.wang@ia.ac.cn

Abstract. In this paper, a facial modeling and animation system based on single image is proposed to generate the natural modeling and animation. First, a Mophable Model has been established offline. Then in the facial modeling stage, the feature points on the image have been positioned and the facial model has been generated. Last in the expression animation stage, a set of animation datum have been produced by MAYA, using expression cloning method to drive the facial model.

Keywords: facial modeling, facial animation, expressional synthesis.

1 Introduction

Realistic facial modeling and animation is an important branch of Computer Graphics field. Because of its broad application prospects, it attracts widespread concerns and great interest for a growing number of researchers. It focuses on how to imitate the real movements and how to generate the corresponding expression of the face. In this paper, we proposed a method to generate the natural facial modeling and animation based on single image.

Compared with the human face modeling algorithm based on more than one picture, Single image face modeling method has many advantages. For example, human-computer interface is simple and it is more convenient for researchers to collect the photos. Also complexity of the program is lower and it has a broader applicability. In the facial modeling methods with single image, Mophable Model[1] which Blanz proposed is one of the most effective way.

In facial animation methods, Parke[2] used shape parameters and expression parameters to describe the human face and control the shape and movement of different parts on the face. And then MPEG-4 standards are proposed to defined a 3D facial animation format[3]. The animation methods based on motion capture track the markers which are on the face of the performers, calculate the 3D motion information and then drive 3D human face model.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 143–150.
springerlink.com

2 System Flow

Recently demanding of quality is increasing. Traditional image-based synthesis technology has been difficult to satisfy this demand. But it is easy to deal with the dynamic speed problem with 3D grid.

In this paper, we proposed a facial modeling and animation system based on single front image to generate the natural modeling and animation. System flow is shown in figure 1. First, a Mophable Model is established offline. We analyze real 3D human face model by principal component analysis method (PCA) to get the shape and texture component of the face. Then in the facial modeling stage, we position the location by Adboost method and the feature points on the image by active appearance models method(AAM) and generated the facial model. Last in the expression animation stage, we produce a set of animation data by MAYA. Using expression cloning method [4] [5], we drive the facial model.

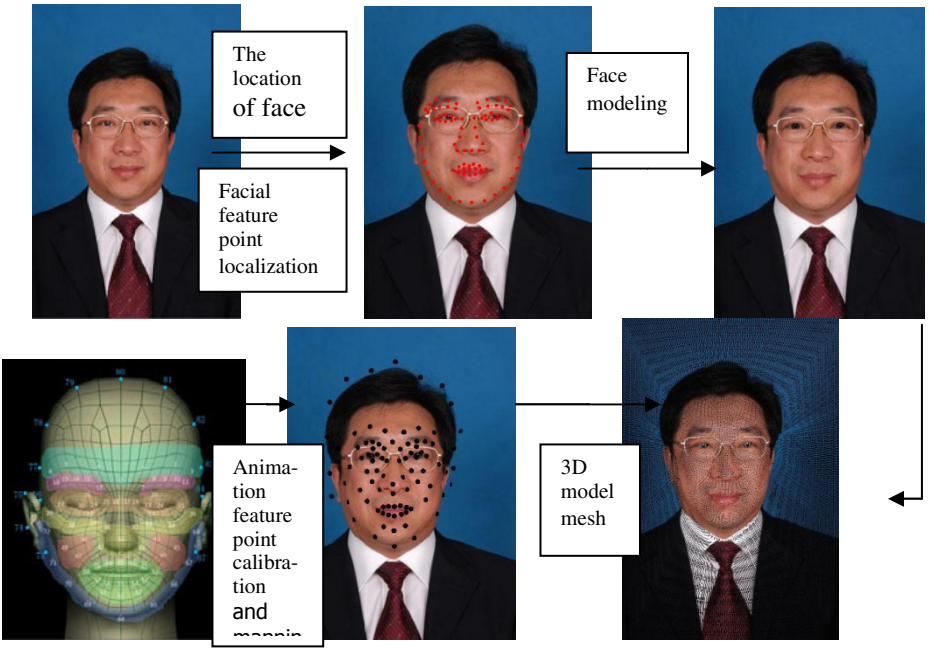


Fig. 1. System flow chart

3 Face Modeling by Single Picture

Face modeling by single picture has simple human-computer interface and it is more convenient for researchers to collect the photos. Also complexity of the program is lower. But single provided a very limited amount of information, which the image information of a face in profile is incomplete.

3.1 Mophable Model Offline

A Mophable Model has been established through analyzing real 3D human face model by principal component analysis method (PCA). In this paper, 3D scanner which called Konica Minolta Vivid910 is used to collect 3D face samples, including 38 male and 12 female. Figure 2a shows the point cloud data of three perspectives. Figure 2b shows the mesh after treatment.

Fig. 2a. The point cloud data **Fig. 2b.** The mesh

Fig. 2. Scan model

Face models after pretreatment have different grid structure. Before principal component analysis, samples require the same topology. We use grid deformation by radial basis function and vector projection to achieve the model after regularity processing. During the scanning process, the samples have different head posture. So the rotation, scaling transformation between the samples should be removed. We select a model as a reference for other models. Other samples demand alignment with the standard model. The matrix of the key point coordinates of reference model is ***Xs***. The matrix of the key point coordinates of target model is ***Xd***. The two matrix can use a similar transformation to describe.

After alignment, a Mophable Model is generated by PCA.

$$\mathbf{x} = \overline{\mathbf{x}} + \sum_{i=1}^{m} b_i \mathbf{x}_i \tag{3.1}$$

In the above formula, $\overline{\mathbf{x}} \in R^{3N}$ is the average face model. Xi is eigenvectors correspond to the first m largest eigenvalues. Any facial shape can be achieved with a linear combination of an average face and basis vectors.

3.2 Personalized Face Model Reconstruction

When entering a face picture, Adboost algorithm [6] is used to compute face location; Improved Active Appearance Model (AAM) face feature location algorithm[7] is used to find the key points' locations. Each image calibrates 87 key points. Interactive tools can be used to adjust the location of key points. The key point on an image have shown in figure 3a. Then on the average 3D model, we also marked the corresponding 87 key points in hand. The key point on average model have shown in figure 3b.

(a) The key point positioning results

(b) The key point on average model

Fig. 3. Corresponding key points between Image and model

Assume that $\mathbf{p}_i = (x_i, y_i, z_i)^T$ is the ith key point of face Model. In this paper, the weak perspective projection is used to project the key points on the 3D face model onto the 2D plane. We assume that first there is only rotation transformation. After projecting the points onto the plane, the transformation of scaling and translation can align the points between the model and image. Assume that $\overline{\mathbf{p}}_i = (\overline{x}_i, \overline{y}_i)^T$ are coordinates of the projection points on the plane; S is scaling factor; $\mathbf{t} = (t_x, t_y)^T$ the amount of translation; Wi is an matrix of 3*3N which used to select the ith key point on the model.

3D human face model and its key points are transformed as:

$$\overline{\mathbf{p}}_i = s\begin{bmatrix}1,0,0\\0,1,0\end{bmatrix}\mathbf{R}\cdot W_i\cdot\mathbf{x}+\mathbf{t} \tag{3.2}$$

Assume that rotation matrixes that the model rounds the x-axis, y-axis and z-axis are $\mathbf{R}_\alpha$, $\mathbf{R}_\beta$ and $\mathbf{R}_\gamma$, and the rotation angles are α, β and γ. Rotation transformation of the face model can be expressed as:

$$\begin{aligned}\mathbf{R} &= \mathbf{R}_\gamma\mathbf{R}_\beta\mathbf{R}_\alpha\\ &= \begin{bmatrix}\cos\gamma & \sin\gamma & 0\\ -\sin\gamma & \cos\gamma & 0\\ 0 & 0 & 1\end{bmatrix}\cdot\begin{bmatrix}\cos\beta & 0 & -\sin\beta\\ 0 & 1 & 0\\ \sin\beta & 0 & \cos\beta\end{bmatrix}\cdot\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\alpha & \sin\alpha\\ 0 & -\sin\alpha & \cos\alpha\end{bmatrix}\\ &= \begin{bmatrix}\cos\gamma\cos\beta & \cos\alpha\sin\gamma+\sin\alpha\sin\beta\cos\gamma & \sin\alpha\sin\gamma-\cos\alpha\sin\beta\cos\gamma\\ -\sin\gamma\cos\beta & \cos\alpha\cos\gamma-\sin\alpha\sin\beta\sin\gamma & \sin\alpha\cos\gamma+\cos\alpha\sin\beta\sin\gamma\\ \sin\beta & -\sin\alpha\cos\beta & \cos\alpha\cos\beta\end{bmatrix}\end{aligned} \tag{3.3}$$

As the weak perspective projection, the depth does not affect XY coordinates.

$$\begin{aligned}\overline{x}_i &= s\left(x_i\left(\cos\gamma\cos\beta\right)+y_i\left(\cos\alpha\sin\gamma+\sin\alpha\sin\beta\cos\gamma\right)\right)+\\ &\quad z_i\left(\sin\alpha\sin\gamma-\cos\alpha\sin\beta\cos\gamma\right)+t_x\\ \overline{y}_i &= s\left(x_i\left(-\sin\gamma\cos\beta\right)+y_i\left(\cos\alpha\cos\gamma-\sin\alpha\sin\beta\sin\gamma\right)\right)+\\ &\quad z_i\left(\sin\alpha\cos\gamma+\cos\alpha\sin\beta\sin\gamma\right)+t_y\end{aligned} \tag{3.4}$$

Formula (3.4) establishes links between a 3D model and an image.

Assume that $\mathbf{q}_i=\left(x_i,y_i\right)^T$ is the ith key point of the image. Face modeling is to find the best posture parameters $s,\mathbf{R},\mathbf{t}$ and PCA factors b_i until minimizing the aberration between projection points from model and the key point of the image. So the cost function should be established.

3.3 Texture Mapping

This article only collected 50 individuals' facial samples which can't represent the rich texture changes. We choose the input image as a texture mapping of the 3D model.

In figure 4, after estimating the 3D face model and transformation parameters, the projection coordinates on the 2D plane can be computed by formula (3.4). And then directly we use these coordinates as texture coordinates of the vertices of the model. The method is fast but model aspect exists texture stretching phenomenon.

Fig. 4. Face modeling results

4 3D Facial Expression Animation

In this paper, the human face model has been driven by key points sparse data mapping method [8].

4.1 Animation Data Mapping

Due to experimental conditions, this article can't collect fine real expression data. Therefore, we use MAYA8.5 produced a 20s facial animation data. Each frame is the coordinate information of the 88 feature point from the source model. First we should select a face model in MAYA8.5 and then make facial animation controllers. Through the controllers different expressions can be generated. Finally in the model 88 key points are chose for animation and output of these points movement data are saved for the animation files.

The key points construct a sparse grid. This paper attempts to direct mapping method. After data scaling, the motion data adds directly to the target face model.

$$\overline{\mathbf{v}}_d^k = \mathbf{S} \cdot \overline{\mathbf{v}}_s^k \tag{4.1}$$

$\overline{\mathbf{v}}_d^k$ is the pure expression data of target model in the Kth frame. $\boldsymbol{S}$ is the scaling matrix.

Direct mapping method in this experiment obtained good results. But because the animation data is not the movement of true face, expression of smiling is not real enough.

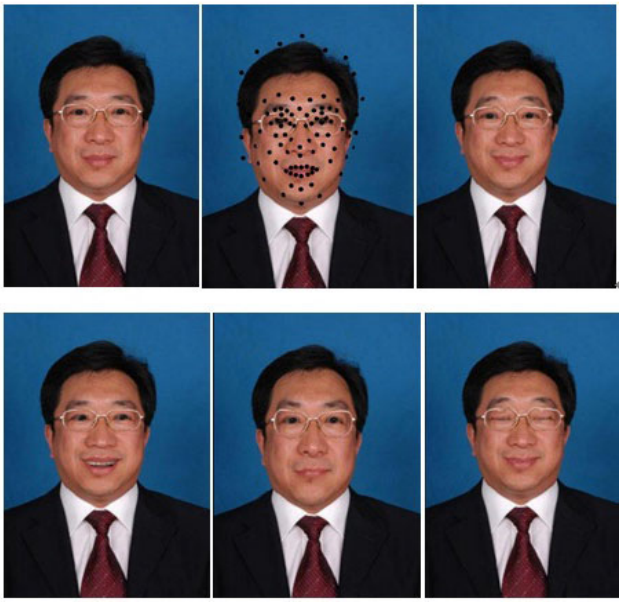


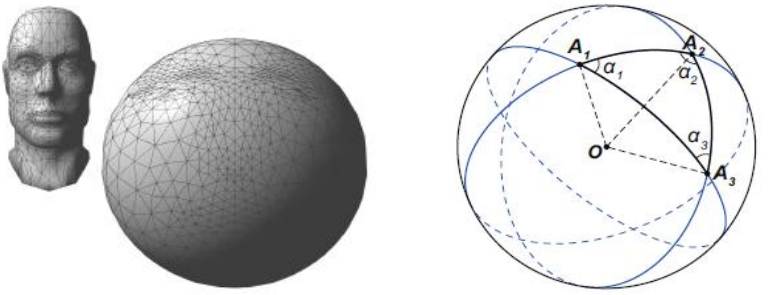
Fig. 5. Animation data mapping

4.2 Animation Data Interpolation

The animation data only gives the motion information of the key points. Based on the spherical parameterization data interpolation technique [9] [10], we solve out the movement data of remaining vertex. Spherical parameterization algorithm transforms the target model into spherical parameter domain; identifies interpolation coefficient

that key points to the common vertex and thus obtains interpolation grid deformation animation.

First the target model should use spherical parameterization. Then calculate the barycentric coordinates of the vertex. According to the barycentric coordinates, we obtain vertex motion vector by interpolating. Figure 6a shows spherical parametric effect. Figure 6b gives grid after spherical parametric and triangular surface changes into a spherical triangle.



(a) Spherical parameterization results (b) Spherical triangle

Fig. 6. Spherical parameterization

The formula (4.2) shows that any point P within an equilateral triangle can be expressed as a weighted average of three vertices. These weights are the barycentric coordinates of the points.

$$P = c_1 A_1 + c_2 A_2 + c_3 A_3 \tag{4.2}$$

Barycentric coordinates are calculated using the spherical triangle area.

$$c_i = \frac{e_i}{e_1 + e_2 + e_3} \tag{4.3}$$

e_i is the triangle area which is composed of the point P and the ith edges. After obtaining the barycentric coordinates, the motion vectors of point P can be calculated by substituting the movement of key points for coordinates in formula (4.2).

5 The Experimental Results Summary

Throughout the process of facial modeling and animation, only calibration of feature points requires a few human interactions. The remaining are accomplished automatically.

This paper presents a face modeling and animation approach based on single image. The operation of this system is simple and quick. The system has high degree of automation and can generate more realistic and natural facial animation.

Acknowledgements. Supported by Youth Foundation of Higher Education Scientific Research of Hebei Province 'Research of fast personalized virtual human modeling or human-computer interaction' (Grant No.2010228).

References

1. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. In: Proceedings of ACM SIGGRAPH, pp. 187–194 (1999)
2. Parke, F.I.: Computer generated animation of faces. In: Proceedings of the ACM Annual Conference, pp. 451–457 (1972)
3. ISO/IEC 14496-2:2001. Coding of Audio-Visual Objects - Part 2: Visual, 2nd edn. (2001)
4. Noh, J.Y., Neumann, U.: Expression cloning. In: Proceedings of SIGGRAPH 2001, pp. 277–288 (2001)
5. Sumner, R.W., Popovic, J.: Deformation transfer for triangle meshes. ACM Trans. Graph 23(3), 399–405 (2004)
6. Wang, S.: A survey of face feature location. Institute of Automation. Chinese Academy of Sciences (2007)
7. Wang, S., Wang, Y., Chen, X.: Weighted active appearance models. In: Master's thesis of International Conference on Intelligent Computing, pp.1295-1304 (2007)
8. Patel, A., Smith, W.A.P.: 3D morphable face models revisited. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1327–1334 (2009)
9. Yao, J., Wang, Y., Ding, B., Li, J.: Facial animation mapping based on spherical parameterization. China Journal of Image and Graphics 14(7) (2009)
10. Gotsman, C., Gu, X., Sheffer, A.: Fundamentals of Spherical Parameterization for 3D Meshes. ACM Trans. Gragh. 22(3), 358–363 (2003)

Approximate Controllability for the Semilinear Fuzzy Integrodifferential Equations in n-Dimensional Fuzzy Vector Space

Ja Hong Koo[1], Yong Gyun Lee[2], Young Chel Kwun[3], and Jin Han Park[4,*]

[1] Department of Fiber Plastic Design, Dong-A University, Busan 604-714, Korea
[2] Department of Chemistry, Dong-A University, Busan 604-714, Korea
[3] Department of Mathematics, Dong-A University, Busan 604-714, Korea
[4] Department of Applied Mathematics, Pukyong National University, Busan 608-737, Korea
jihpark@pknu.ac.kr

Abstract. In this paper, we study the ε-approximate controllability for the semilinear fuzzy integrodifferential control system in fuzzy vector space. This is an extension of result of Kwun et al.[4] to n-dimensional fuzzy vector space.

Keywords: approximate controllability, semilinear, fuzzy integrodifferential equation, fuzzy vector space.

1 Introduction

The exact controllability has limited applicability. Approximately controllable system are more abundant. For the background of the approximate controllable, see ([1], [2],[5], [6], [7]). But the aboved cited results of the approximate controllability for the functional differential did not consider the fuzzy differential equations. Recently, Kwun et. al[4] studied approximate controllability for semilinear fuzzy integrodifferential equation in one-dimension fuzzy vector space.

In this paper, we study the approximate controllability for the following semilinear fuzzy integrodifferential equations in n-dimensional fuzzy vector space:

$$\frac{dx_i(t)}{dt} = A_i[x_i(t) + \int_0^t G(t-s)x_i(s)ds] + f(x_i(t)) + B_i u_i(t), 0 < t < T, \tag{1}$$

$$x_i(0) = x_{0_i},\ i = 1,2,...,n, \tag{2}$$

where the state $x_i(t)$, takes values in $X_i (\subset E_N^i)$ and the control $u(\cdot)$ is in another space $Y_i (\subset E_N^i)$. B_i is a bounded linear operator from Y_i to X_i. $A_i : [0,T] \to X_i$ is fuzzy coefficients, E_N^i is the set of all upper semicontinuously convex fuzzy numbers on R. $f : X_i \to X_i$ is nonlinear regular fuzzy function.

* Corresponding author.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 151–154.
springerlink.com

$G(t)$ is $n \times n$ continuous matrix such that $\dfrac{dG(t)x_i}{dt}$ is continuous for $x_i \in X_i$ and $t \in [0,T]$ with $\|G(t)\| \le k$, $k > 0$, $x_{0_i} \in C([0,T]: X_i)$ is initial value.

2 Existence and Uniqueness of Fuzzy Solution

In this section we consider the existence and uniqueness of fuzzy solutions for the equations (1)-(2)($u \equiv 0$). We define $A = (A_1,...,A_n)$, $x = (x_1,...,x_n)$, $f = (f_1,...,f_n)$ and $x_0 = (x_{0_1},...,x_{0_n})$. Then $A, x, f, x_0 \in X \subset (E_N^i)^n$. Where $(E_N^i)^n = E_N^1 \times ... \times E_N^n$ is n -dimensional fuzzy vector space.(See [8]) Instead of the equations (1)-(2)($u \equiv 0$), we consider the following fuzzy integrodifferential equations in $(E_N^i)^n$.

$$\frac{dx(t)}{dt} = A[x(t) + \int_0^t G(t-s)x(s)ds] + f(x(t)), 0 < t < T, \tag{3}$$

$$x(0) = x_0, \ i = 1,2,...,n, \tag{4}$$

With fuzzy coefficient $A: [0,T] \to (E_N^i)^n$, initial value $x_0 \in (E_N^i)^n$. Given nonlinear regular fuzzy functions $f: (E_N^i)^n \to (E_N^i)^n$ satisfies global Lipschitz condition, i.e., there exists finite constants $k > 0$ such that $d_L([f(x(s))]^\alpha, [f(y(s))]^\alpha) \le k d_L([x(s)]^\alpha, [y(s)]^\alpha)$ for all $x, y \in (E_N^i)^n$.

Definition 1. [3] The complete metric D_L on $(E_N^i)^n$ is defined by $D_L(u,v) = \sup_{0<\alpha<1} d_L([u]^\alpha, [v]^\alpha) = \sup_{0<\alpha\le 1} \max_{1\le i\le n} \{|u_{il}^\alpha - v_{il}^\alpha|, |u_{ir}^\alpha - v_{ir}^\alpha|\}$ for any $u, v \ (E_N^i)^n$.

Definition 2. Let $u, v \in C([0,T]: (E_N^i)^n)$. $H_1(u,v) = \sup_{0\le t\le T} D_L(u(t), v(t))$.

For the sequel, we need the following assumption:

(H1) $S(t)$ is a fuzzy number satisfying, for $y \in (E_N^i)^n$, $S'(t)y \in C^1$ ([0,T]: $(E_N^i)^n$) $\cap C([0,T]: (E_N^i)^n)$, $\dfrac{dS(t)y}{dt} = A[S(t)y + \int_0^t G(t-s)S(s)yds]$, where $[S(t)]^\alpha = \prod_{i=1}^n [S_i(t)]^\alpha = \prod_{i=1}^n [S_{il}^\alpha(t), S_{ir}^\alpha(t)]$, $S(0) = I$ and $S_{ij}^\alpha(t)$ ($j = l, r$) is continuous with $|S_{ij}^\alpha(t)| \le c$, $c > 0$, for all $t \in [0,T]$.

(H2) $ckT < 1$.

If x is an integral solution of (1)-(2)($u \equiv 0$), then x is given by

$$x(t) = S(t)x_0 + \int_0^t S(t-s) f(x(s))ds \tag{5}$$

Theorem 1. Let $T > 0$, and hypotheses (H1)-(H2) hold. Then for every $x_0 (\in C([0,T]:X))$, equation (1)-(2)($u \equiv 0$) has a unique solution $x \in C([0,T]:X)$.

3 Approximate Controllability

In this section, we study approximate controllability for the equations (1)-(2). We consider a unique solution of equations (1)-(2), for each $u \in Y \subset (E_N^i)^n$;

$$x(t;u) = S(t)x_0 + \int_0^t S(t-s)(f(x(s)) + Bu(s))ds, \tag{6}$$

And we also define the continuous linear operator $\tilde{S}$ from X to $C([0,T]:X)$ by $(\tilde{S}p)(t) = \int_0^t S(t-s)p(s)ds, \quad p \in X, \ 0 \le t \le T$. We define a solution mapping W from Y to $C([0,T]:X)$ by $(Wu)(t) = x(t:u)$. The reachable set of nonlinear system are used to be compared to the reachable sets of its corresponding linear system ($f \equiv 0$ in (6)). Put

$$K_T(0) = \{z(T) \in X; z(T) = S(T)x_0 + \int_0^T S(T-s)(Bu)(s)ds, u \in Y\}$$

and the reachable set $K_T(f)$ in X by $K_T(f) = \{x(T;u) \in X; x(T;u) =$

$$S(T)x_0 + \int_0^T S(T-s)(f(x(s)) + (Bu)(s))ds, u \in Y\}.$$

Lemma 1. Let $u, x_0 \in X$. Then under hypotheses (H1)-(H2), the solution mapping $(Wu)(t) = x(t:u)$ of (6) satisfies

$$H_1(x, \chi_{\{0\}}) \le (cH_1(x_0, \chi_{\{0\}}) + cTH_1(B, \chi_{\{0\}})H_1(u, \chi_{\{0\}}))\exp(ckT).$$

Lemma 2. . Let $u_1, u_2 \in Y$. Then under hypotheses (H1)-(H2), the solution mapping $(Wu)(t) = x(t:u)$ of (6) satisfies

$$H_1(x(t;u_1), x(t;u_2)) \le cT\exp(ckT)H_1(Bu_1, Bu_2).$$

Definition 3. The system (6) is called approximately controllable on $[0,T]$ if $\overline{K_T(f)} = X$, that is for any given $\varepsilon > 0$ and $\xi_T \in X$ there exists some control $v \in Y$ such that $H_1(\xi_T - S(T)x_0, \tilde{S}f(x(s;v)) + \tilde{S}Bv) < \varepsilon$.

We assume the following hypotheses ; For every arbitrary given $\varepsilon > 0$ and $p \in X$, there exists some $u \in Y$ such that

(H3) $H_1(\tilde{S}p, \tilde{S}Bu) < \varepsilon$, **(H4)** $H_1(Bu, \chi_{\{0\}}) \le q_1 H_1(p, \chi_{\{0\}})$ where q_1 is a positive constant independent of p, **(H5)** q_1 satisfies $\frac{q_1 kcT}{1-kcT} < 1$,

(H6) $\{S(t); t \in R^+\}$ is compact.

Lemma 3. Under hypotheses (H3)-(H6), we have $\overline{K_T(0)} = X$.

Lemma 4. Let $x(t:u)$ and suppose that (H1)-(H6) hold. Then for $T > 0$ there exists a constant $0 < M < 1$ such that

$$H_1(f(x(t;v_1)), f(x(t;v_2))) \le \frac{M}{1-M} H_1(Bv_1, Bv_2).$$

Theorem 2. Under hypotheses (H1)-(H6), $\overline{K_T(0)} = \overline{K_T(f)}$ proved that $T < 1/(q_1+1)kc$.

References

1. George, R.K.: Approximate controllability of semilinear systems using integral contractors. Numerical Functional Analysis and Optimization 16, 127–138 (1995)
2. Kwun, Y.C., Park, D.G., Kang, W.K.: Approximate controllability for semilinear delay integrodifferential equations with nonlocal initial value. Indian Journal Pure Application and Mathematics 31, 1607–1617 (2000)
3. Kwun, Y.C., Kim, J.S., Park, M.J., Park, J.H.: Nonlocal controllability for the semilinear fuzzy integrodifferential equations in n-dimensional fuzzy vector space. Advances in Difference Equations 2009, Article ID 734090, 16pages (2009)
4. Kwun, Y.C., Kim, J.S., Park, M.J., Park, J.H.: ε-Approximate controllability for the semilinear fuzzy integrodifferential equations. J. Comp. Anal. Appl. 12, 1171–1179 (2011)
5. Naito, K.: Approximate controllability for trajectories of semilinear control system. J. Optimization Theory and Applications 60, 57–65 (1989)
6. Park, J.Y., Han, H.K., Kwun, Y.C.: Approximate controllability of second order integrodifferential systems. Indian Journal Pure Application and Mathematic. 29, 941–950 (1998)
7. Zhou, H.X.: Approximate controllability for a class of semilinear abstract equations. Siam J. Control and Optimization 21, 551–565 (1983)
8. Wang, G., Li, Y., Wen, C.: On fuzzy n-cell number and n-dimension fuzzy vectors. Fuzzy Sets and Systems 158, 71–84 (2007)

A Novel Chemical Plume Tracing Method Using a Mobile Sensor Network without Anemometers

Yu-Xiu Wu, Qing-Hao Meng, Yong Zhang, and Ming Zeng

School of Electrical Engineering and Automation,
Tianjin University, 300072 Tianjin, China
yuxiu_wu@163.com, linsio@126.com,
{qh_meng,zengming}@tju.edu.cn

Abstract. Chemical plume tracing (CPT) in time-varying airflow environments is addressed. A novel CPT method using a mobile sensor network (MSN) with fixed topology is proposed. The motion direction of the MSN is determined by the concentration detection-event probability. Unlike the existing multi-robot CPT methods, the global position of each mobile node is not needed in the fixed-topology MSN. In addition, anemometers are not used in the proposed CPT method. Furthermore, in the proposed MSN method, a sensor node measures chemical concentration and then communicates with the other nodes using binary value. Simulations in a large-scale advection–diffusion airflow environments validate the convergence of the proposed MSN based CPT method.

Keywords: Chemical plume tracing, Mobile sensor network, Fixed topology MSN.

1 Introduction

Olfaction has been widely used by animals for foraging, courtship, communication and evading predators. Inspired by animals' tracing odor abilities, in the 1990s, researchers started to use mobile robots equipped with gas sensors and other sensors to imitate animals' behaviors of odor source localization [1]. Such study is called chemical plume tracing (CPT) [2]. The research on the CPT could be applied in areas such as judging toxic or harmful gas leakage locations, searching for survivors in collapsed buildings, humanitarian de-mining and thwarting terrorist attacks. Owing to the turbulent and fluctuant characteristics of gas transport, tracing the resultant patchy plume down to its source is thus not a trivial task.

In turbulent flows, the concentration of a chemical substance cannot always be measured by a chemical sensor located at the downwind directions of the chemical source because of complex fluctuating structure of the chemical plume [3]. When a chemical patch arrives it is detected as a burst with a complicated small-scale structure, as local concentration fluctuates strongly while the patch is passing by [4]. Therefore, the instantaneous concentration value might not have direct relation with its measurement location in the time-varying chemical plume. In other words, it is hard to get the real diffusion model in the time-varying airflow environments.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 155–162.
springerlink.com

To realize the CPT task, multi-robot strategy has been proposed by several researchers. Zarzhitsky and colleagues [5] proposed an artificial physics-based multi-robot formation method to realize the CPT via simulation, where the fluxotaxis tracing strategy [6] was adopted to navigate the robot toward the emitter source. A collaborative spiral surge (SS) strategy for distributed multiple robots was proposed by Hayes [7]. The cooperation in the SS strategy is that the upwind surging robot attracts all robots downwind or with no plume information to surge in the direction of it. The particle swarm optimization algorithm was also improved by several researchers for the CPT [8-9].

To our knowledge, only a few multi-robot based CPT researches in which anemometers were not used have been reported. Marjovi et al. [10] proposed a robot swarming navigation algorithm for finding the odor source in an unknown environment using only the chemical sensors, and most robots moved toward the high concentration areas. The experiments in an indoor artificial airflow environment demonstrated the feasibility.

In our CPT method, anemometers are not used, and only the relative position information between mobile nodes is needed. The sensor nodes communicate the binary concentration information via wireless mode. On the basis of the fact that the maximal amplitude of the concentration within a chemical patch decreases away from the chemical source and the coverage area of a patch increases [3, 4], a MSN motion strategy for CPT is proposed in this paper. The mobile nodes within the MSN move synchronously toward the direction determined using detection-event probability. Simulations in a large-scale advection–diffusion airflow environments [11] validate the convergence of the proposed MSN based CPT method.

2 CPT without Wind Information

2.1 CPT Strategy

Figure 1 shows a circular-topology MSN with six nodes in an instantaneous chemical plume, where the plume is copied from the advection–diffusion plume model [11]. The dots with different radii stand for the chemical patches, and hollow small circles distributed along the circumference of the dashed circle represent the mobile nodes. The intersection points between the dashed circle and the straight line L being perpendicular to the wind direction are presented using A and B. The line L divides the dashed circle into two parts, and C and D stand for two points in the upper and lower half circumferences, respectively.

Given a concentration threshold C_λ , if the concentration measured by the MSN node i is higher than C_λ , then it is thought that the detection event occurs in the node i.

Lemma 1: Suppose MSN is divided into two parts by the straight line L which is perpendicular to wind direction, and the topology graph intersects with the two plume boundaries (See Fig. 1). Then we have

$$P(O_{L_{up}}) < P(O_{L_{down}}) . \tag{1}$$

Where $P(O_{L_{up}})$ and $P(O_{L_{down}})$ stand for the detection-event probabilities of the MSN parts being close to and far from the chemical source, respectively.

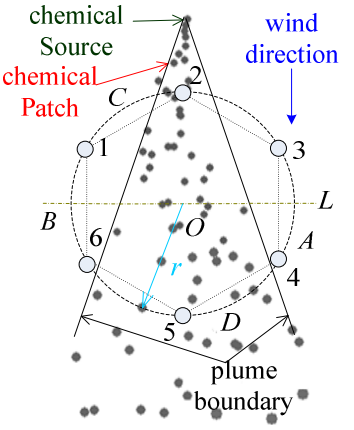


Fig. 1. A MSN with fixed circular topology in an instantaneous chemical plume. The chemical source releases chemical molecules continuously and the release rate is constant. The wind transports chemical patches far from the source. The coverage area of the chemical patches increases and the average concentration decreases with an increase in the distance. The circular center and radius of the MSN topology are denoted by *O* and *r*, respectively. The mobile nodes are distributed uniformly along the circumference.

Using the Divergence Theorem [12]

$$\int_W \nabla \cdot (c\vec{V}) dW = \oint_S (c\vec{V}) \cdot d\vec{S} , \tag{2}$$

where c stands for the chemical concentration, $\vec{V}$ is the wind vector, W represents the control area, S stands for the enclose curve of the control area, $c\vec{V}$ is the chemical mass flux. Considering the conservation of mass in the dashed-line circle in Fig.1, we get

$$\oint_{S_1} (c\vec{V}) \cdot d\vec{S}_1 = \int_{W_1} \nabla \cdot (c\vec{V}) dW_1 = \int_{W_2} \nabla \cdot (c\vec{V}) dW_2 = \oint_{S_2} (c\vec{V}) \cdot d\vec{S}_2 = 0 . \tag{3}$$

Where W_1 and W_2 stand for the areas enclosed by S_1 and S_2, respectively. S_1 is composed by the line segment AB (denoted as L_1) and the arc ACB (denoted as L_2); S_2 is composed by the line segment BA (denoted as L_3) and the arc BDA (denoted as L_4). Suppose the wind is constant, then

$$\int_{L_2} c dL_2 = \int_{L_4} c dL_4 . \tag{4}$$

From references [3, 4] we know that the average concentration decreases along the downwind direction. In addition, L_2 and L_4 are symmetrical about L, and L_2= L_4. So if the average value of $c \neq 0$, the average value of c at the point of L_2 is higher than that at the symmetrical point of L_4. So the non-zero concentration part in L_4 is longer than that in L_2. Then, lemma 1 is proved.

According to the lemma 1, the parts with low and high detection-event probabilities probably locate in the upwind and downwind areas, respectively. To search for the odor source, the MSN should avoid moving downwind. A motion direction is constructed for the MSN to trace plume. On one hand, the constructed motion direction can lead the MSN to intersect with the both of plume boundaries (see Fig.1); on the other hand, by following the constructed direction, the MSN can move toward the part which has lower detection-event occurrence probability (i.e., the upwind direction).

Figure 2 shows the schematic diagram of the constructed motion direction for the MSN. The three filled small circles (denoted as i, j, k) in Fig. 2(a) stand for the mobile nodes at which detection events occur. In Fig. 2(b), the symbols $\vec{F}_i$, $\vec{F}_j$ and $\vec{F}_k$ stand for three unit vectors along the directions from the center O to the nodes i, j and k, respectively. The symbol $\vec{F}$ is the composite vector of $\vec{F}_i$, $\vec{F}_j$ and $\vec{F}_k$. In Fig. 2(c), the dashed arrow $\vec{F}_w$ indicates a unit vector of upwind direction, and the symbol $\vec{F}'$ is the composite vector of $\vec{F}_w$ and $\vec{F}/|\vec{F}|$. In Fig. 2(d), the vectors $\vec{F}_i$ and $\vec{F}_k$ are in the same straight line N, and the straight line L is perpendicular to N. The symbol $\vec{F}''$ indicates the composite vector of $\vec{F}$ and $\vec{F}_i/|\vec{F}_i|$.

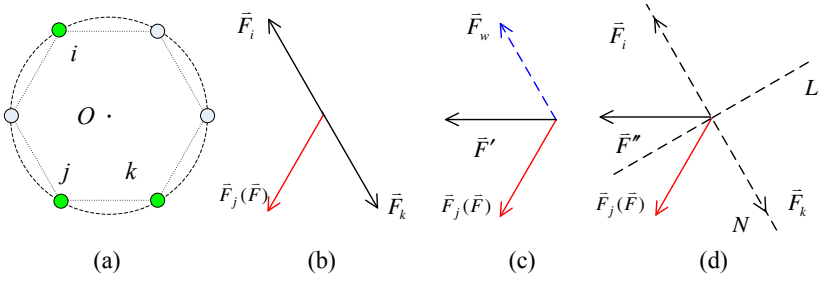


Fig. 2. MSN motion direction for CPT

In Fig. 2(b), the composite vector $\vec{F}$ can be expressed as follows,

$$\vec{F} = \vec{F}_j = \vec{F}_i + \vec{F}_j + \vec{F}_k . \tag{5}$$

From the lemma 1 we know that the vector $\vec{F}$ points to the downwind direction. If the MSN moves toward the direction of the vector $\vec{F}'$, it is quite possible that the MSN moves toward the chemical source and keeps in the chemical plume. The wind

information indicated by the vector $\vec{F}_w$, however, cannot be obtained. As a matter of experiences, if the detection events occur at the two nodes being on the same diameter of the MSN topology, we can judge that the wind direction is probably parallel with the diameter. On the basis of the lemma 1, we can further judge that the direction of $\vec{F}_i$ in Fig. 2(d) stands for the upwind direction, i.e.,

$$\vec{F}_w // \vec{F}_i . \tag{6}$$

So the composite vector $\vec{F}''$ in Fig. 2(d) can be expressed as follows

$$\vec{F}'' = \vec{F} \Big/ \left|\vec{F}\right| + \vec{F}_i \Big/ \left|\vec{F}_i\right| . \tag{7}$$

2.2 CPT Algorithm

The MSN is considered to be a distributed system in this research. The mobile nodes are distributed uniformly along the circumference in the beginning, and the velocity of each node is constant. Without considering the error of velocity, the MSN can be treated as a rigid body. The chemical concentration is measured by every sensor node at different location. When a detection event happens, a sensor node saves and sends the information to the other nodes. So every node can get the detection-event information in one control period. The communication information between MSN nodes is binary, and the amount of information is less than n (bit) (n is the number of sensor nodes). The MSN based CPT algorithm is listed as follows,

```
Algorithm: MSN-CPT
    while the center of MSN is not close to source  do
        if the MSN is within plume
          then execute change_node_orientation()
        else execute sweep()
        end if
    end while
    Strategy: change_node_orientation ()
        if the detection events happens at two MSN nodes
         being on the same diameter of the MSN topology
          then calculate straight line L and save it
        else
          straight line L = saved straight line L
        end if
        calculate F'', the direction of F'' is the node
         motion direction
```

The `sweep()` function in the above CPT algorithm is a coverage method for finding the plume, and the sweeping trajectory in this research is similar to that described in [13].

3 Simulation and Results

Figure 3 shows the advection–diffusion plume model developed by Farrell et al. [11]. The blue arrows denote the speed and direction of wind, where the lengths and directions of arrows indicate the wind speeds and wind directions, respectively. The small hollow circles locating at the upper right side stand for the mobile nodes. The red arrows connecting to the nodes indicate the motion direction of the MSN. The meandering curve composed by lots of grey dots stands for the chemical plume. In the simulations, the search region is a rectangle defined by $x \in [0,100]$ m and $y \in [-50,50]$ m. The chemical source is placed at (20, 0) m and the start location of the MSN is (90, -30) m. The speed of the MSN node is set to 0.5 m/s. If the simulation time is more than 600 s, the simulation is terminated and the CPT task is thought to be failed. If the distance from MSN center to the chemical source is less than 2 m and detection event occurs, the simulation is thought to be successful. One hundred simulations were implemented for each parameter. The success rate is defined as the number of successful simulations divided by 100, and the average success time is defined as the average time of all the successful simulations.

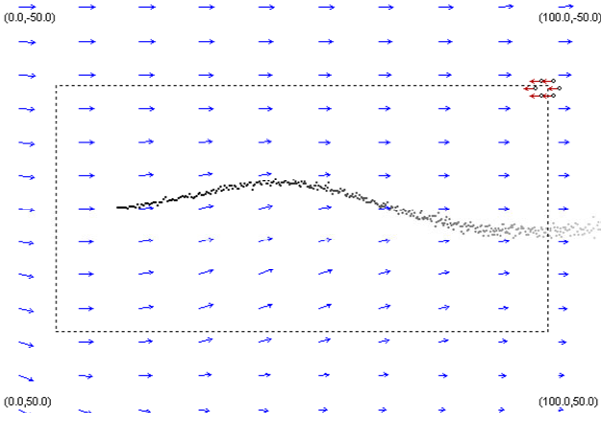
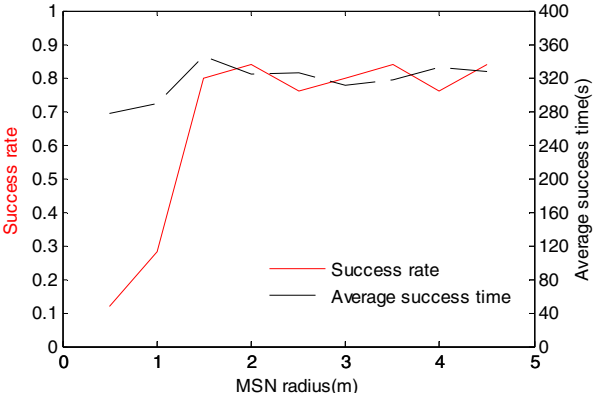


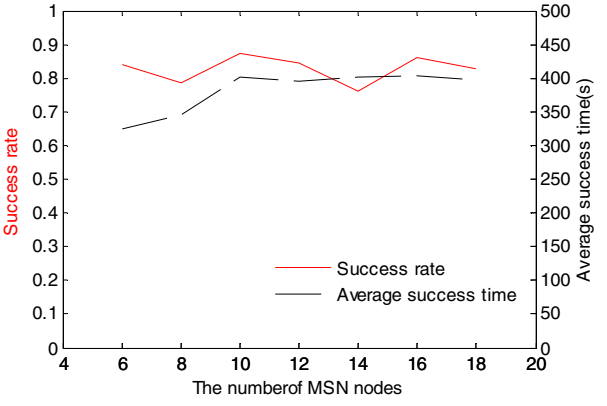
Fig. 3. The simulation environment

The MSN topology radius and the number of MSN nodes are the main factors that influence the algorithm performance. Simulations showed that the success rate was always lower than 20% when the number of MSN nodes was less than six. Therefore, the minimal number of MSN nodes was set to six in our simulations. Two situations were simulated. The first one was that the MSN topology radius was varied from 0.5 m to 4.5 m with an interval of 0.5 m while keeping the number of MSN nodes fixed (six in this simulation). The second situation was that the number of MSN nodes changed from six to eighteen with an interval of 2 while keeping the topology radius fixed (2 m in this simulation). Figure 4 (a) and (b) show the simulation results of the two situations.

From Fig. 4 (a) it is seen that when the MSN radius is longer than 1.5 m the success rate and average success time tend to be stable around 80.6% and 345s, respectively. The success rate is low when the MSN radius is shorter than 1.5 m, this is because when the MSN diameter is smaller than the plume width, all the nodes are located in the plume, which does not meet the assumption of the lemma1. Figure 4 (b) shows the effect of the number of MSN nodes on the success rate and the average success time. When the number is more than 8, the success rate and the average success time tend to be stable around 83.3% and 400 s, respectively. The results shown in Fig. 4 tell us that the performances of the MSN based CPT tend to be stable when the MSN radius is longer than 1.5 m and the number of nodes are more than 8. In other words, increasing the number of nodes (more than 8) and MSN radius (longer than 1.5 m) are not necessary for improving the CPT performances.



(a) The results of MSN based CPT with different topology radii



(b) The results of CPT using MSN with different number of nodes

Fig. 4. The results of the MSN based CPT

4 Conclusion

A fixed-topology mobile sensor network without anemometers is proposed for chemical plume tracing in time-varying airflow environments. Simulations using a 100 m-by-100 m advection–diffusion plume model show that the odor source could be found within 400 s with a success rate of more than 80% when the number of the MSN is more than eight and the MSN topological radius is longer than 1.5 m. Next we will consider the MSN nodes formation problem and implement the proposed CPT method using the real-hardware platform.

Acknowledgement. This work is supported by China '863' Program (No. 2007AA04Z219), the NSFC projects (No.60875053, 60802051), Tianjin NSF project (09JCYBJC02100) and the Program for New Century Excellent Talent in University.

References

1. Ishida, H., Suetsugu, K., Nakamoto, T., et al.: Study of Autonomous Mobile Sensing System for Localization of Odor Source Using Gas Sensors and Anemometric Sensors. Sensor. Actuat. A-Phys. 45(2), 153–157 (1994)
2. Russell, R.A.: Tracking chemical plumes in constrained environments. Robotica 19(4), 451–458 (2001)
3. Murlis, J., Elkinton, J.S., CardéR, T.: Odor Plumes and How Insects Use Them. Annu. Rev. Entomol. 37, 505–532 (1992)
4. Balkovsky, E., Shraiman, B.I.: Olfactory search at high Reynolds number. Proc. Natl Acad. Sci. USA 99, 12589–12593 (2002)
5. Zarzhitsky, D., Spears, D.F.: Swarm approach to chemical source localization. In: IEEE International Conference on Systems, Man and Cybernetics, pp. 1435–1440. IEEE Press, Waikoloa (2005)
6. Zarzhitsky, D., Spears, D.F., Thayer, D., Spears, W.M.: Agent-Based Chemical Plume Tracing Using Fluid Dynamics. In: Hinchey, M.G., Rash, J.L., Truszkowski, W.F., Rouff, C.A. (eds.) FAABS 2004. LNCS (LNAI), vol. 3228, pp. 146–160. Springer, Heidelberg (2004)
7. Hayes, A.T., Martinoli, A., Goodman, R.M.: Distributed odor source localization. IEEE. SENS. J. 2(3), 260–271 (2002)
8. Jatmiko, W., Sekiyama, K., Fukuda, T.: A PSO based mobile robot for odor source localization in dynamic advection-diffusion with obstacles environment: theory, simulation and measurement. IEEE Computational Intelligence Magazine 2(2), 37–51 (2007)
9. Li, F., Meng, Q.H., Bai, S., et al.: Probability-PSO algorithm for multi-robot based odor source localization in ventilated indoor environments. In: International Conference on Robotics and Automation, pp. 1206–1215. Springer, Wuhan (2008)
10. Marjovi, A., Nunes, J., Sousa, P., et al.: An olfactory-based robot swarm navigation method. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4958–4963. IEEE Press, Anchorage (2010)
11. Farrell, J.A., Murlis, J., Long, X., et al.: Filament-based atmospheric dispersion model to achieve short time-scale structure of odor plumes. Environ. Fluid. Mech. 2(1), 143–169 (2002)
12. Hughes-Hallett, D., Gleason, A.M., McCallum, W.M., et al.: Calculus: Single and Multivariable. John Wiley and Sons (1998)
13. Lilienthal, A.J., Duckett, T.: Building gas concentration gridmaps with a mobile robot. Robot. Auton. Syst. 48(1), 3–16 (2004)

Exploration of Spatial Pipeline Computation for Heuristic Access Coarse-Grained Reconfigurable Cells*

Xinning Liu, Wei Ge, and Yue Du

National ASIC System Engineering Research Center,
Southeast University, Nanjing,
210096, People's Republic of China
{xinning.liu,duiker,duyue032}@seu.edu.cn

Abstract. Spatial pipelined approach can be an effective technique for efficient improvement of coarse-grained reconfigurable cells. Take advantage of the characteristics, data-driven and "shadow variable", of Dynamic Giant Reconfigurable Cells (DGRC). This paper shows the exploration of spatial computation approach. Experimental results show that the proposed spatial pipeline approach can achieve the highest 5x times improvement in performance and reduce the configuration resources.

Keywords: spatial pipeline computation, heuristic access, coarse-grained reconfigurable, program optimization.

1 Introduction

The coarse-grained reconfigurable architectures have the potential to gain its benefit from spatial model of computing compared to traditionally microprocessor's temporal fetch-and-execute model[1-4]. Spatial computation allows outputs from one *reconfigurable cell* (RC) of the reconfigurable array to be directly connected to the inputs of another RC. Temporal computation (i.e. the output from a processing element is always stored in an intermediate register) limits the performance[5]. Furthermore, the coarse-grained reconfigurable architectures' instruction accounts for several normal instructions when compared to a VLIW processor or RISC processor, thus the number of instructions decreases.

Since loops are typically the most processor intensive part of stream multimedia applications, optimization of inner loops is of special relevance[5]. Natural loops in the control flow graph are recognized whether they originate from FOR loops, WHILE loops or even backwards GOTO statements in the C source code. All loops are treated as WHILE loops (having data-dependent exits)[6].

To solve these high data-parallel and computation intensive applications above, we propose *Dynamic Giant Reconfigurable Cells* (DGRC) - a novel CGRA architecture.

* This work was sponsored by the National Scientific Foundation of China (Grant No. 61006029) and Jiangsu Scientific Foundation (Grant No. BK2010165).

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 163–170.
springerlink.com

DGRC integrates a large number of tiny RCs into a signal matrix. Extending the RC functional characteristics, DGRC has the capacities of heuristic access and configuration switch. Combined with high data throughput LRF, the DGRC can fulfill efficient data streaming applications and a powerful parallel computing. And it provides a new solution in the field of scientific computing and streaming media applications.

The rest of the paper is organized as follows. Section 2 presents target architecture and its main features. Section 3 describes the basic spatial computation on the DGRC. Section 4 presents the improved spatial pipeline computation approach. Section 5 presents the experimental results with discussions. Section 6 concludes the paper and presents future work.

2 Target Architecture

In order to support the spatial computation on the coarse-grained reconfigurable cells, the DGRC need to satisfy two functional requirements: (1) the context and data must be controlled independently; (2) the computation data can be parallel access.

We adapt the DGRC as the target architecture platform to explore the spatial calculation. Compared with the traditional CGRAs, the DGRC is unique for its achieving heuristic self-control by initiative data access and automatic context switching.

A simplified micro-system of the DGRC is shown in Fig.1. It contains a 4x4 orthogonal reconfigurable matrix equipped with 16 evenly located basic Reconfigurable Cells (RC). Typically, the Context Register File (CRF) is used for the storage of the running time contexts information, while the Global Register File (GRF) and the Local Register File (LRF) are used for storing the global data, the local data and temporary variables respectively. The DGRC establishes a communication with the hybrid register files via the mixed matrix cross-bus.

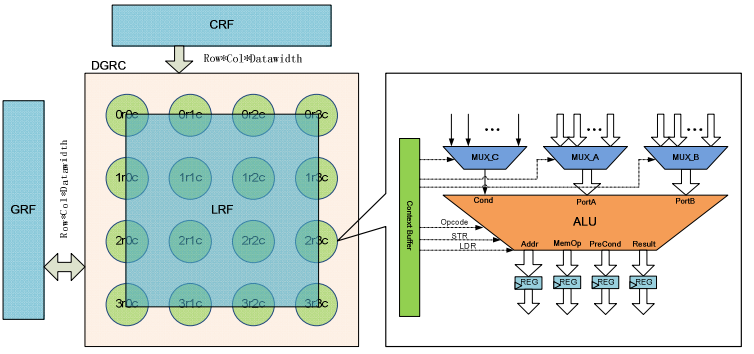


Fig. 1. A simplified micro-system of the DGRC

The RC is constructed with data multiplexer, ALU (fixed-point operations) and access logic as shown in **Fig.1**. Data multiplexer can receive data from nearest RCs and immediate operand. It also accepts positive and negative condition flag from nearest RCs. ALU module takes charge of basic arithmetic and logical operations, such as add, shift, compare, etc. The ALU can do the half data-width multiple, that means if the data width is 32-bit, the 16x16 multiplier is instantiated in the ALU. The predicated signal like the FUs mentioned in the ADRES[7], not only supports the branch operation, but also indicates the status of the arithmetic and access operation. And each RC has the capacity to load and store data at the size of byte, half word, word, which can achieve the self data access and context switching operation.

Adopting the interconnection matrix, each RC constitutes a different interconnect structure with the surrounding eight RCs by configuring static parameters. RCs at the edge of the matrix, which have no input connection, are set to zero by default. Separate input and output data paths of RC allow data to spread from top to down or vice versa. One flatten single-layer can simplify the mapping implementation.

LRF could be full access, which not only greatly simplifies the programming model and the variables mapping, but also reduces the communication latency. Reading operation can be applied to the entire LRF by all of the RCs at the same time. However, writing operation must be mutually exclusive to the same register of the LRF. The LRF looks like a "Register Cloud" floating on the reconfigurable matrix as shown in **Fig.2,** and each RC in the array can access it out-of-order.

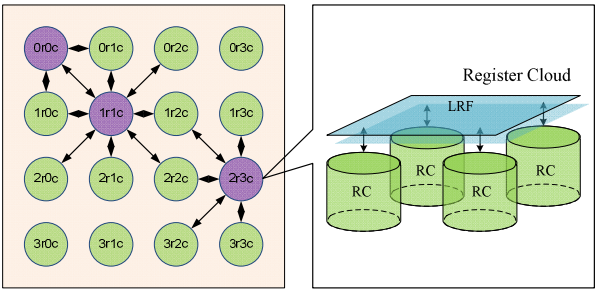


Fig. 2. Illustration of RC interconnection and register cloud

3 Basic Spatial Computation

We have manually mapped several algorithms on the DGRC in traditional way. The loop cases as FOR loops, WHILE loops and backwards GOTO statements in C source code all rewrite to WHILE loops (having data-dependent exits).

As is shown in **Fig.3**, the basic mapping process on the DGRC is as follows:

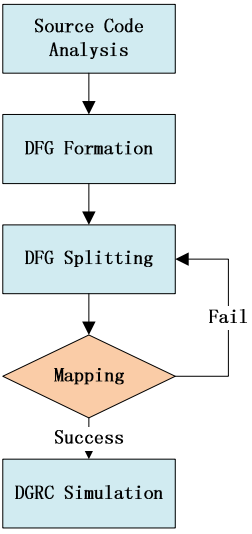


Fig. 3. Basic mapping process

Source Code Analysis: Although the IF-ELSE statements are well supported by DGRC, we focus on the FOR loops and WHILE loops, which can get more parallel acceleration. The branch predictions and computation can be treated as a set of logic and arithmetic operations.

DFG Formation: A data flow graph (DFG) is constructed upon our analysis. As the DFG is built, all control dependence is converted to data dependence through the predicated signal, which determines the corresponding RC to be active or not. Load and Store nodes are introduced to indicate global inputs and outputs of the DFG.

DFG Splitting: The original DFG is often too large to be mapped on the RC array, so we need to split the large DFG into a few Sub-DFGs. We adopt the method, which generate least temporal variables, as the standard of the partition. After the partition, new temporal variables should be allocated into register files. CP (Context Pointer) operation, which has been added at the end of the current context, switches to another context.

Mapping: If the Sub-DFG mapping into the RC array is limited by the resource of DGRC, we go back to DFG splitting step until the new Sub-DFG is placed on the RC array.

We take n_real_update algorithm, which get from benchmark dspstone, as an example.

```
TYPE main()
{                                                    (1)
static TYPE A[N], B[N], C[N], D[N] ;
……
```

```
    for (i = 0 ; i < N ; i++)
      *p_d++  = *p_c++ + *p_a++ * *p_b++ ;
  ......
    return(0)  ;
  }
```

First, we draw the DFG(Data Flow Gragh) of the algorithm.

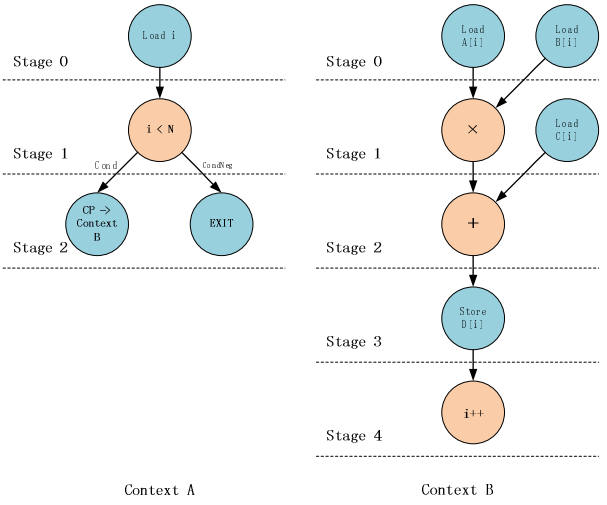


Fig. 4. Original mapping result

According to the basic spatial computation process, the original mapping result of n_real_update is shown in Fig.4. The branch of the loop is mapped in the first context. If the loop condition is true, jumps to the second context, else exit the loop. The second context executes the loop body once and modifies the loop variable "i", and then jumps back to the first context. The codes after the for-loop are mapped in the following contexts which are not shown in the figure.

The original contexts make the output of the RC directly connect to the inputs of another one, reducing the register fetch operation, but the classic spatial mapping is not pipelined and concurrency only exists in the RCs which are on the same stage. What's more, context switching occurs at each iteration of the loop, which is too frequent to maintain a high performance. The original mapping results cannot fully exploit the potential of the DGRC.

4 Improvement Pipeline Mapping

In order to accelerate the computation, we improve the mapping method in the following ways.

Since the computation potential cannot be reached by the classic mapping, we try to pipeline the execution of the RCs to achieve a higher performance.

We analyze the inner-loop and intra-loop dependence and then try to exploit the concurrency between different iterations as much as possible. As DGRC is a data-driven model instead of a clock-trigger model, we can do this easily. Each RC is active at the Condition of input signal and enable status being both valid (Table 1), which provide flexible choose to enhance the mapping flow.

Table 1. Condition input field.

Name	Hex	Description
ALR	0x3	Always run
OSR	0x2	One shot run, then Condition run
CNR	0x1	Condition run
NVR	0x0	Never run

As Fig.5. shows, first, we merge the original two contexts together to avoid contexts switching.

Then we try to make full use of the condition bit of DGRC. We configure the condition execution mode of the RC in stage 0 to be OSR, and others to be CNR, so the first RC computes at the first cycle and outputs result and condition bit to its following RCs. Thus other RCs will get their input and corresponding condition bit every cycle, then RCs in different stages could be active.

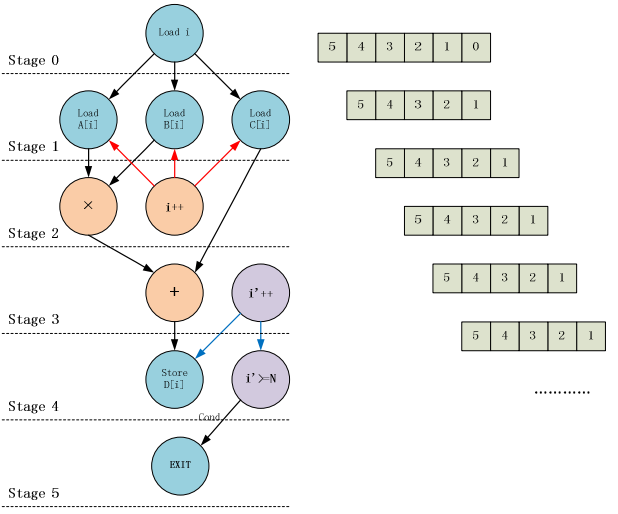


Fig. 5. Optimized pipeline mapping result

Data dependence is a great limitation for pipelining. We introduce “shadow variable”, which is the duplication of the true variable, to reduce the reflection of data dependence for the same variable. The “shadow variable” i' controls the STORE operation independently. So the pipeline computation in the DGRC becomes possible and we can make the array fully pipelined.

5 Experiment Result

A cycle-accurate C++ simulator of DGRC was proposed to perform our experiment. The detailed parameters of the target DGRC are listed in Table 2.

Table 2. Architectural parameters of the target DGRC architecture

Architecture Parameter	Number
System dimension	6 by 6
Connect of RCs	8NN
Data Width	16 bits
Address Width	16 bits

The Target DGRC has 6 rows and 6 columns with 36 RCs. Each RC directly connects with nearest 8 neighbor RCs, and the data width of RC's input and output was set to 16 bits.

We apply three algorithms taken from dspstone benchmark to estimate our work by two different mapping methods. The experiment results are listed in Table 2.

Table 3. Experiment results.

Algorithm	Approach	Context Number	Exec Cycles	IPC	Initial Interval (II)
FIR	basic	2	116	1.595	9
	pipeline	1	22	9.136	1
n_real_update	basic	2	115	2.04	7
	pipeline	1	21	13.14	1
matrix1	basic	2	811	2.508	8
	pipeline	2	331	7.257	1

We take the number of execution cycles, instructions per cycle (IPC) and Initial Interval (II) as the parameters for the estimation of our improved pipeline mapping method. Since we apply spatial pipelining, the II of each case is significantly reduced, thus considerable concurrency is obtained. As we can draw from the **Table 3**, with the proposed spatial pipelined mapping method, a 5.27x times improvement in performance is achieved for FIR, as well as 5.48x times for n_real_update and 2.45x times for matrix1.

5 Conclusions

This paper proposed spatial pipelining computation based on the DGRC, which effectively improve the abilities of high data-parallel and computation intensive applications. The experimental results show that the proposed spatial pipeline

approach can achieve remarkable improvement (2.45~5.48) in performance at the benefit of reducing configuration context.

Future work will focus on the automatic mapping and optimization of spatial pipeline algorithm.

References

1. DeHon, A.: The density advantage of configurable computing. Computer 33(4), 41–49 (2000)
2. André, D., Wawrzynek, J.: Reconfigurable computing: what, why, and implications for design automation. ACM, New Orleans (1999)
3. Yoon, J.W., Shrivastava, A., Sanghyun, P., Minwook, A., Yunheung, P.: A Graph Drawing Based Spatial Mapping Algorithm for Coarse-Grained Reconfigurable Architectures. IEEE Transactions on Very Large Scale Integration (VLSI) Systems 17(11), 1565–1578 (2009)
4. Minwook, A., Yoon, J.W., Yunheung, P., Yoonjin, K., Kiemb, M., Kiyoung, C.: A spatial mapping algorithm for heterogeneous coarse-grained reconfigurable architectures. In: Proceedings of Design, Automation and Test in Europe, DATE 2006, p. 6 (2006)
5. Barat, F., Jayapala, M., Op de Beeck, P., Deconinck, G.: Software pipelining for coarse-grained reconfigurable instruction set processors. In: Design Automation Conference, 2002. Proceedings of ASP-DAC 2002. 7th Asia and South Pacific and the 15th International Conference on VLSI Design. Proceedings, pp. 338–344 (2002)
6. Callahan, T.J., Wawrzynek, J.: Adapting software pipelining for reconfigurable computing. ACM, San Jose (2000)
7. Mei, B., Vernalde, S., Verkest, D., De Man, H., Lauwereins, R.: ADRES: An Architecture with Tightly Coupled VLIW Processor and Coarse-Grained Reconfigurable Matrix. Springer, Berlin (2003)

Controllability for the Fuzzy Differential Equations Driven by Liu Process

Yong Gyun Lee[1], Ja Hong Koo[2], Young Chel Kwun[3], and Jin Han Park[4,*]

[1] Department of Chemistry, Dong-A University, Busan 604-714, Korea
[2] Department of Fiber Plastic Design, Dong-A University, Busan 604-714, Korea
[3] Department of Mathematics, Dong-A University, Busan 604-714, Korea
[4] Department of Applied Mathematics, Pukyong National University, Busan 608-737, Korea
jihpark@pknu.ac.kr

Abstract. In this paper, we study the controllability for the fuzzy differential equations driven by Liu process. This is an extension of result of Kwun et al.[6] to fuzzy differential equations driven by Liu process with control term.

Keywords: controllability, fuzzy differential equations driven by Liu process, fuzzy number, credibility space.

1 Introduction

For the background of the controllability of fuzzy system, see ([3],[6], [10],[11]). And controllability of stochastic system, see ([1],[2],[5]). But the aboved cited results of the controllability for the stochastic functional differential did not consider the random fuzzy differential equations. In 2006, Liu ([8]) studied credibility theory. Recently, Liu([9]) studied existence of fuzzy solutions for the fuzzy differential equations driven by Liu process. In this paper, we study the controllability for the following fuzzy differential equations driven by Liu process:

$$dx(t,\theta) = Ax(t,\theta)dt + f(x(t,\theta))dC(t) + Bu(t)dt, t \in [0,T] \tag{1}$$

$$x(0) = x_0 \in E_N, \tag{2}$$

where the state $x(t,\theta)$ takes values in $X(\subset E_N)$ and another space $Y(\subset E_N)$. E_N is the set of all upper semicontinuously convex fuzzy numbers on R, (Θ, P, Cr) is credibility space, A is fuzzy coefficient, the state function $x:[0,T]\times(\Theta,P,Cr) \to X$ is a fuzzy process, $f: X \to X$ is nonlinear regular fuzzy function, $u:[0,T]\times(\Theta,P,Cr) \to Y$ is a control function, B is a linear bounded operator from Y to X. $C(t)$ is a standard Liu process, $x_0 \in E_N$ is initial value.

* Corresponding author.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 171–174.
springerlink.com

2 Existence of Solutions

We will not fully describe details which use already in the other published papers for lack of space(see [4],[5],[7],[8],[9]). Those details will be used without special mentioning as known facts. On the other hands, we will describe full contents which define and use in this paper at first time.

In this section we consider the existence and uniqueness of solutions for the equations(1)-(2)($u \equiv 0$). We consider the following fuzzy differential equations driven by Liu process.

$$dx_t = Ax_t dt + f(x_t)dC_t, t \in [0,T], \tag{3}$$

$$x(0) = x_0 \in E_N, \tag{4}$$

where the state $x_t = x(t,\theta)$ takes values in $X(\subset E_N)$. A is fuzzy coefficient, initial value $x_0 \in (E_N)$. $f : X \to X$ is nonlinear regular fuzzy function, C_t is a standard Liu process.

Lemma 2.1. [9] Let g be a function of two variables and let a_t be an integrable uncertain process. Then the uncertain differential equation $dX_t = a_t X_t dt + g(t, X_t)dC_t$ has a solution $X_t = Y_t^{-1} Z_t$ where $Y_t = \exp(-\int_0^t a_s ds)$ and Z_t is the solution of uncertain differential equation $dZ_t = Y_t g(t, Y_t^{-1} Z_t)dC_t$ with initial value $Z_0 = X_0$.

Lemma 2.2. For $x(0) = x_0$, if x_t is solution of the equation (3)-(4), then x_t is given by $x_t = S(t)x_0 + \int_0^t S(t-s)f(x_s)dC_s, t \in [0,T]$, where $S(t)$ is continuous with $S(0) = I$, $|S(t)| \le c, c > 0$, for all $t \in [0,T]$.

Lemma 2.3. [4] Let $f(t)$ be continuous fuzzy process, the following inequality of fuzzy integral holds $\left|\int_c^d f(t)dC_t\right| \le K\int_c^d |f(t)|dt$, where $K = K(\theta)$.

For the sequel, we need the following assumption:

(H1) For $x_t, y_t \in C([0,T]\times(\Theta, P, C_r), X), t \in [0,T]$, there exists positive number m such that $d_L([f(x_t)]^\alpha, [f(y_t)]^\alpha) \le m d_L([x_t]^\alpha, [y_t]^\alpha)$.

(H2) $2cmKT < 1$.

By Lemma2.2, we know that equation (3)-(4) has solution x_t. Thus in theorem 2.1, we show that uniqueness of solution for equation (3)-(4).

Theorem 2.1. If hypotheses (H1) and (H2) are hold. For every $x_0 \in E_N$, then the equation (3)-(4) has unique solution $x_t \in C([0,T]\times(\Theta, P, C_r), X)$.

3 Controllability

In this section, we study controllability for the equations (1)-(2). We consider a unique solution of equations (1)-(2), for each $u \in Y(\subset E_N)$;

$$x_t = S(t)x_0 + \int_0^t S(t-s)f(x_s)dC_s + \int_0^t S(t-s)Bu_s ds, \tag{5}$$

$$x(0) = x_0 \in E_N, \tag{6}$$

where $S(t)$ is continuous with $S(0) = I$, $|S(t)| \le c$, $c > 0$, for all $t \in [0,T]$.

Definition 3.1. The equation (1)-(2) is called controllable on $[0,T]$, if for every $x_0 \in E_N$ there exists a control $u_t \in Y$ such that the solution x_t of (1)-(2) satisfies $x_T = x^1 \in X$, a.s. θ (i.e., $[x_T]^\alpha = [x^1]^\alpha$).

Define the fuzzy mapping $G : \tilde{P}(R) \to X$ by

$$G^\alpha(v) = \begin{cases} \int_0^T S^\alpha(T-s)Bv_s ds, v \in \overline{\Gamma}_u, \\ 0, \text{otherwise}, \end{cases}$$

where $\tilde{P}(R)$ is a nonempty fuzzy subset of R and $\overline{\Gamma}_u$ is closure of support u. Then there exists $G_i^\alpha (i = l, r)$ such that

$$G_l^\alpha(v_l) = \int_0^T S_l^\alpha(T-s)B(v_s)_l ds, (v_s)_l \in [(u_s)_l^\alpha, (u_s)^1],$$

$$G_r^\alpha(v_r) = \int_0^T S_r^\alpha(T-s)B(v_s)_r ds, (v_s)_r \in [(u_s)^1, (u_s)_r^\alpha],$$

We assume that G_l^α, G_r^α are bijective mappings. We can introduce α-level set of u_s is $[u_s]^\alpha = [(u_s)_l^\alpha, (u_s)_r^\alpha]$

$= [(G_l^\alpha)^{-1}\{(x^1)_l^\alpha - S_l^\alpha(T)(x_0)_l^\alpha - \int_0^T S_l^\alpha(T-s)f_l^\alpha((x_s)_l^\alpha)dC_s\}, (G_r^\alpha)^{-1}\{$

$(x^1)_r^\alpha - S_r^\alpha(T)(x_0)_r^\alpha - \int_0^T S_r^\alpha(T-s)f_r^\alpha((x_s)_r^\alpha)dC_s\}]$. Then substitute this expression into the equation (5)-(6) yields α-level set of x_T. $[x_T]^\alpha =$

$$[S_l^\alpha(T)(x_0)_l^\alpha + \int_0^T S_l^\alpha(T-s)f_l^\alpha((x_s)_l^\alpha)dC_s + G_l^\alpha(G_l^\alpha)^{-1}\{(x^1)_l^\alpha$$

$$- S_l^\alpha(T)(x_0)_l^\alpha - \int_0^T S_l^\alpha(T-s)f_l^\alpha((x_s)_l^\alpha)dC_s\},$$

$$S_r^\alpha(T)(x_0)_r^\alpha + \int_0^T S_r^\alpha(T-s)f_r^\alpha((x_s)_r^\alpha)dC_s + G_r^\alpha(G_r^\alpha)^{-1}\{(x^1)_r^\alpha$$

$$- S_r^{\alpha}(T)(x_0)_r^{\alpha} - \int_0^T S_r^{\alpha}(T-s) f_r^{\alpha}((x_s)_r^{\alpha}) dC_s \}]$$

$= [(x^1)_l^{\alpha}, (x^1)_r^{\alpha}] = [x^1]^{\alpha}$. Hence this control u_t satisfies $x_T = x^1$, a.s. θ.

We now set $\Phi x_t = S(t)x_0 + \int_0^t S(t-s) f(x_s) dC_s + \int_0^t S(t-s) BG^{-1}\{x^1$

$- S(T)x_0 - \int_0^T S(T-\tau) dC_\tau\} ds$, where fuzzy mapping G^{-1} satisfies above statements. Notice that $\Phi x(T) = x^1$, which mean that the control u_t steers the (5)-(6) from the origin to x^1 in time T provided we can obtain a fixed point of the operator Φ.

(H3) Assume that the linear system of (5)-(6) ($f \equiv 0$) is controllable.

Theorem 3.1. If Lemma 2.3 and hypotheses (H1)-(H3) are satisfied, then the equation (1)-(2) is controllable on $[0,T]$.

References

1. Arapostathis, A., George, R.K., Ghosh, M.K.: On the controllability of a class of nonlinear stochastic system. System & Control Letters 44, 25–34 (2001)
2. Balasubramaniam, P.: Controllability of quasilinear stochastic evolution equations in hilbert spaces. J. Appl. Math. and Stochastic Analysis 14, 151–159 (2001)
3. Dinh, P.N., Lam, Q.D.: On the stability and controllability of fuzzy control set differential equations. In: 4th International Workshop on REC 2010, pp. 293–307 (2010)
4. Fei, W.: Uniqueness of solutions to fuzzy differential equations driven by Liu's process with non-Lipschitz coefficients. In: International Conference on Fuzzy Systems and Knowledge Discovery 2009, pp. 565–569 (2009)
5. Feng, Y.: Convergence theorems for fuzzy random variables and fuzzy martingales. Fuzzy Sets and Systems 103, 435–441 (1999)
6. Kwun, Y.C., Kim, J.S., Park, M.J., Park, J.H.: Nonlocal controllability for the semilinear fuzzy integrodifferential equations in n-dimensional fuzzy vector space. Advances in Difference Equations 2009, Article ID 734090, 16pages (2009)
7. Liu, B., Liu, Y.K.: Expected value of fuzzy variable and fuzzy expected value models. IEEE Transactions on Fuzzy Systems 10, 445–450 (2002)
8. Liu, B.: A survey of credibility theory. Fuzzy Optim. Decis. Making 5, 387–408 (2006)
9. Liu, B.: Fuzzy process, hybrid process and uncertain process. Journal of Uncertain Systems 2, 3–16 (2008)
10. Park, J.H., Park, J.S., Kwun, Y.C.: Controllability for the Semilinear Fuzzy Integrodifferential Equations with Nonlocal Conditions. In: Wang, L., Jiao, L., Shi, G., Li, X., Liu, J. (eds.) FSKD 2006. LNCS (LNAI), vol. 4223, pp. 221–230. Springer, Heidelberg (2006)
11. Park, J.H., Park, J.S., Ahn, Y.C., Kwun, Y.C.: Controllability for the impulsive semilinear fuzzy integrodifferential equations. Advances in Soft Computing 40, 704–713 (2007)
12. Wang, G., Li, Y., Wen, C.: On fuzzy n-cell number and n-dimension fuzzy vectors. Fuzzy Sets and Systems 158, 71–84 (2007)

Research on Active Anti-rolling Methods Using Moving Mass System

Rongwu Yang, Liang Shen, and Quan Zhou

Shanghai Jiao Tong University, postgraduate, 1954 Huashan Road,
Xuhui Distrist Shanghai, China
{yrongwu,doriszhouy}@gmail.com, shenliang@seastel.com

Abstract. The active moving mass system is used to reduce the undesirable rolling motion in the present paper. First, dynamic differential equations of the ship roll with moving mass system are built. After that, the mathematical models of different control methods, including PID feedback control and model predictive control, have been built. The simulations of different anti-rolling methods are numerically achieved based on Simulink. Finally, a model test is conducted in S60 model. The results are presented and the rolling amplitude can be reduced up to 70% in irregular wave.

Keywords: moving mass, PID feedback control, model predictive control, anti-rolling.

1 Introduction

Ships at sea are inevitably influenced by many factors, such as sea waves and winds. These factors generate undesirable motions such as roll. These motions could be extremely fierce when ships meet tough sea conditions, and seriously endanger life or interfere with normal work at sea. Therefore, it is very important for ships to damp its rolling angle by using anti-rolling devices.

There are various kinds of anti-rolling devices. Based on the working principle, these devices can be divided into gravity, hydrodynamic and gyroscope types. Base on the control method, these can be divided into passive, semi-active and active types. There are several kinds of anti-rolling devices currently in use, including gyroscope, bilge keel, anti-rolling tank, active fin stabilizer, rudder roll stabilization, moving mass system and jet flap[1][2].

The active moving mass system is used to reduce the undesirable rolling motion in the present paper. It generates a reducing moment by moving the mass from side to side and changing the weight of the port and starboard of ship. In the 1880s, Kremer [2] designed the first passive moving mass system. In the 1930s, Seimens[3] invented an active moving mass system. Because of the unsteadiness of the performance due to poor accuracy of the mechanical feedback system and huge noise it made, the moving mass system was not considered as a promising anti-rolling means for a long time. However, with the development of control technology and manufacturing technology and considering this method's advantages including simpler structure and less cost, moving mass system has become increasingly advantageous in small- and

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 175–184.
springerlink.com

medium-sized vessels. The Sirehna company[4] is conducting research on moving mass system and has launched a full-scale test.

In the present paper, the dynamic differential equations of the vessel with moving mass system are built based on the analysis of Conolly Equation and the dynamic model of moving mass system. Then the mathematical models of different control methods, including PID feedback control and model predictive control, are analyzed and built. The simulations of different control methods are numerically achieved based on Simulink. After that, a model test is conducted and the result is presented.

2 The Dynamic and Control Principles of Moving Mass System

The dynamic principle of the moving mass system consists of roll dynamics (Conolly Equation) and moving mass dynamics; the control principle consists of traditional PID feedback control and modern model predictive control.

2.1 The Dynamic Differential Equations of the Ship Roll

The swing around the x-axis along the longitudinal direction of ship is called roll. Based on d'Alembert principle, the equilibrium condition of ship is the sum of external moment $\sum M = 0$, so

$$I'_{xx}\ddot{\phi} + 2N\dot{\phi} + W\left|\dot{\phi}\right|\dot{\phi} + Dh\phi = Dh\alpha_m . \tag{1}$$

If roll angle is small, the nonlinear term can be ignored. So

$$I'_{xx}\ddot{\phi} + 2N\dot{\phi} + Dh\phi = Dh\alpha_m . \tag{2}$$

Finally the dynamic differential equations of the vessel roll can be expressed as:

$$\ddot{\phi} + 2\nu\dot{\phi} + \omega_\phi^2\phi = \omega_\phi^2\alpha_m, (2\nu = 2N / I'_{xx}; \omega_\phi^2 = Dh / I'_{xx}) . \tag{3}$$

Where Φ is the roll angle; D is the displacement of ship; h is the initial metacentric height; N and W are damping torque coefficients; I'_{xx} is the sum of ship inertia moment and additional inertia moment; υ is damping coefficient, which indicates the damping and inertia's influence on roll damping; ω_Φ is roll natural frequency, which is the roll frequency of ship in calm water.

If there are anti-rolling devices on the ship, it is necessary to add control moment M_C to equation (2), then,

$$I'_{xx}\ddot{\phi} + 2N\dot{\phi} + Dh\phi = Dh\alpha_m + M_C . \tag{4}$$

If an active anti-rolling system makes $M_C = -Dh\alpha_m$, then the ship will stop rolling because of damping. If the control moment was divided as follow:

$$M_C = A\phi + B\dot{\phi} + C\ddot{\phi} . \tag{5}$$

Then the roll equations can be written as:

$$(I'_{xx}+C)\ddot{\phi}+(2N+B)\dot{\phi}+(Dh+A)\phi=Dh\alpha_m. \quad (6)$$

So the anti-rolling system is virtually increasing the inertia moment (From I'_{xx} up to $I'_{xx}+C$), damping (From $2N$ up to $2N+B$) and initial stability (From Dh up to $Dh+A$). Table 1 shows the effects using different control strategies.

Table 1. The effects using different control strategies

Control strategy	Effects	Advantage	Disadvantage
Angle control	Increase restoring moment, decrease roll period	Feedback signal is accurate, effective on low frequency band wave	Enhance roll in high frequency band wave
Angular velocity control	Increase roll damping	effective on all frequency band wave, especially on resonance region	Have little anti-rolling effects on low and high frequency band wave
Angular acceleration control	Increase inertia moment, increase roll period	effective on high frequency band wave	Enhance roll in high frequency band wave and difficult to obtain accurate signal from instrument
Proportional control	Decrease wave moment	have the same anti-rolling effects on all frequency band wave	difficult to obtain accurate signal from instrument and not the optimal solution

Fig. 1 shows the roll RAO curves using different control strategies, where the horizontal axis is frequency of wave input and the vertical axis is the roll RAO of ship. The control target of anti-rolling device is to keep the ship stable by viewing the wave moment as noise and reducing the noise:

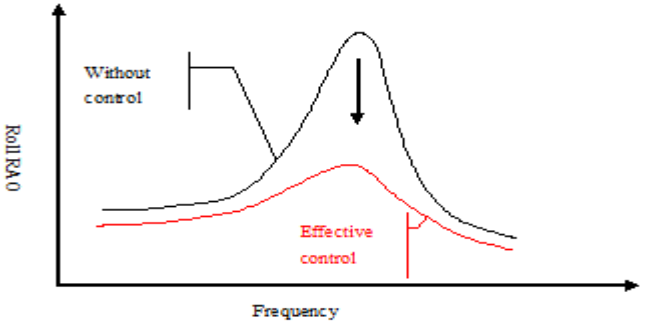


Fig. 1. The roll RAO curves change under the effective control

2.2 The Dynamic Model of Moving Mass Device

Fig. 2 shows the 3d model of the moving mass system. In this system[5], the rail of the moving mass is placed across the deck, and the mass is moving along the transverse direction of ship. The distance between the moving mass and ship gravity center is Zm. So the 3d model can be simplified as 2d model just considering the single-degree-of-freedom roll. Fig. 3 shows the 2d force analysis of the moving mass system. is the position of the mass; F is the force of actuator exerting on the mass; Mc is the control moment; is roll angle; a is the acceleration of the moving mass. From the analysis of force balance on the mass along the transverse direction of ship,

$$F - mg\sin(\phi) - ma = 0. \tag{7}$$

$$a = F/m - g\sin(\phi) = \ddot{d}(t). \tag{8}$$

The reaction of the mass generates the control moment Mc, which consists of moments coming from reaction F′ (relative to F) and N (relative to normal component of gravity mgcos(ϕ)).

$$M_{F'} = m\ddot{d}(t)z_m + mg\sin(\phi)z_m. \tag{9}$$

$$M_N = -mg\cos(\phi)d(t). \tag{10}$$

Minus indicates that the moment is in the opposite direction with roll angle. Finally,

$$M_C = m\ddot{d}(t)z_m + mg\sin(\phi)z_m - mg\cos(\phi)d(t). \tag{11}$$

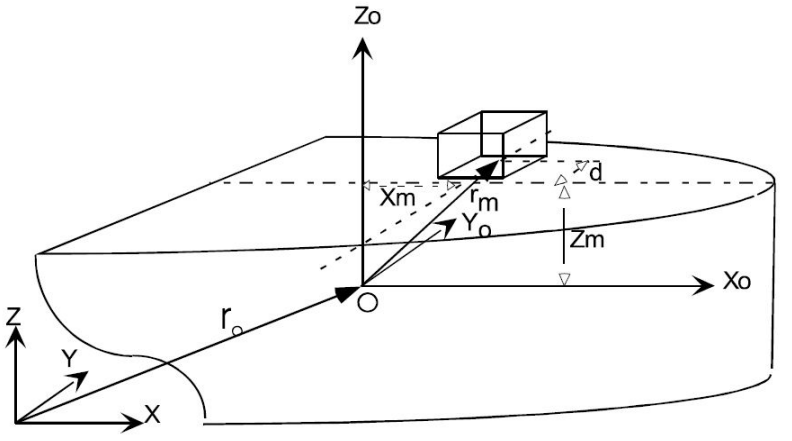


Fig. 2. Sketch of the moving mass system

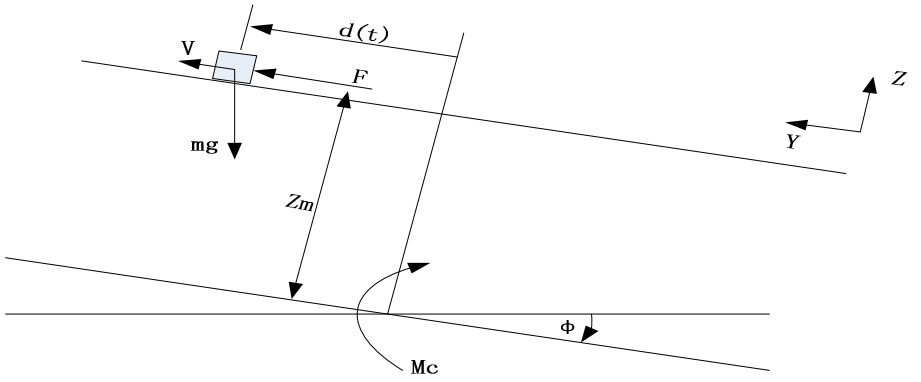


Fig. 3. Force analysis of the moving mass system

2.3 PID Feedback Control

Fig. 4 is the control block diagram using PID feedback control. The working principle is as follow: the controller accepts the roll signal and generates the control signal $d_c(t)$. $\Delta d(t)$ is position difference between the command position and real position. And motor accepts the position difference signal and generates the force $F(t)$ to push the mass moving to real position $d(t)$. Meanwhile the control moment M_c can be determined by formula (11).

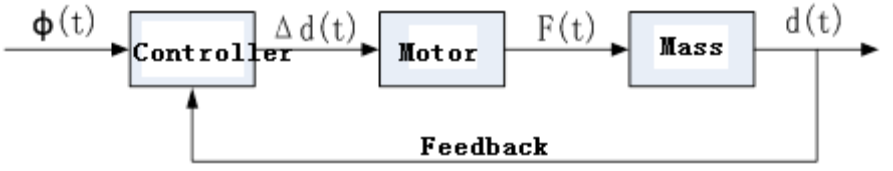


Fig. 4. The control block diagram using PID feedback control

According to formula (5) and (11), the control signal $d_c(t)$ can be expressed as follow:

$$d_c(t) = a\phi + b\dot{\phi} + c\ddot{\phi}. \tag{12}$$

$$d_c(t) = K_i \int \dot{\phi} dt + K_p \dot{\phi} + K_d \frac{d\dot{\phi}}{dt}. \tag{13}$$

According to Table 1, the angular velocity control strategy was chosen because it has anti-rolling effects on all frequency bands, especially on resonance region.

So $K_i = K_d = 0$, K_i is angle control coefficient; K_p is angular velocity control coefficient; K_d is angular acceleration control coefficient.

$$d_c(t) = K_p\dot{\phi}. \tag{14}$$

2.4 Predictive Control

Adopting angular velocity as feedback in PID feedback control can lead to obvious noise disturbance and uncontrollable vibration. Meanwhile, it is inconvenient to put the constraints in the model. So the present paper introduces model predictive control (MPC) to improve the control effects.

Base on the model prediction, receding horizon and feedback correction, MPC has advantages as follow: the control effect is noticeable; the robustness is fine; model accuracy is less demanded; it is easy to put the constraints in the model.

MPC algorithm is achieved as follow: (1) the model of controlled plant is built and the step response is obtained; (2) the control Increment $\Delta u(k)$ is worked out based on optimal control after the weights, prediction and control horizons are chosen; (3) the prediction error was computed and updated based on receding horizon principle; (4) step 2 and 3 are repeated to achieve the online control.

3 Simulation Model in Simulink

The simulation is achieved after the wave model, ship rolling model, actuator model and controller model have been built in Simulink.

3.1 Actuator Simulation Model

The actuator consists of a servo motor and a moving mass.

Fig. 5 shows the model built in Simulink. The PID module is used to simulate servo motor, and the mass is represented as a gain module and a double integration module. The input signal of the actuator is the position command from controller, and its output signal is the real position of the mass.

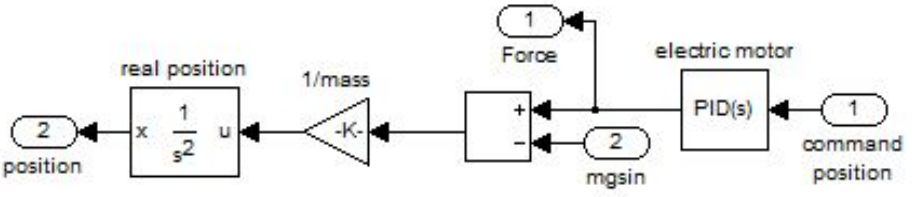


Fig. 5. The simulation model of actuator in Simulink

3.1 Controller Simulation Model

1) PID feedback controller

Based on equation (13), PID module can be used to simulate PID feedback controller. According to the angular velocity control strategy in Table 1, it uses angular velocity as feedback, so it needs a derivative module to change the roll signal to angular velocity signal.

2) Model predictive controller

Model predictive controller is built based on Model Predictive Control Toolbox in Matlab. The controlled plant (actuator and roll model) must be a linear model. The transfer function of the plant can be obtained using Mason formula.

Because the gravity center of mass is higher than that of the ship, the controlled plant has non-negative zeros. So the controlled plant is non-minimum phase system and undershoot will occur in step response. With the same set of motor parameters, the undershoot in MPC controller is smaller than in PID feedback controller.

4 The Results of Simulation and Test

A model test is conducted in a towing tank to verify the anti-rolling effect. The spectrum of irregular waves was JONSWAP spectrum. The overall length of the S60 model is 3m; molded breadth is 0.39m; draught is 0.16m. The ship model is in beam wave and the speed is 0m/s.

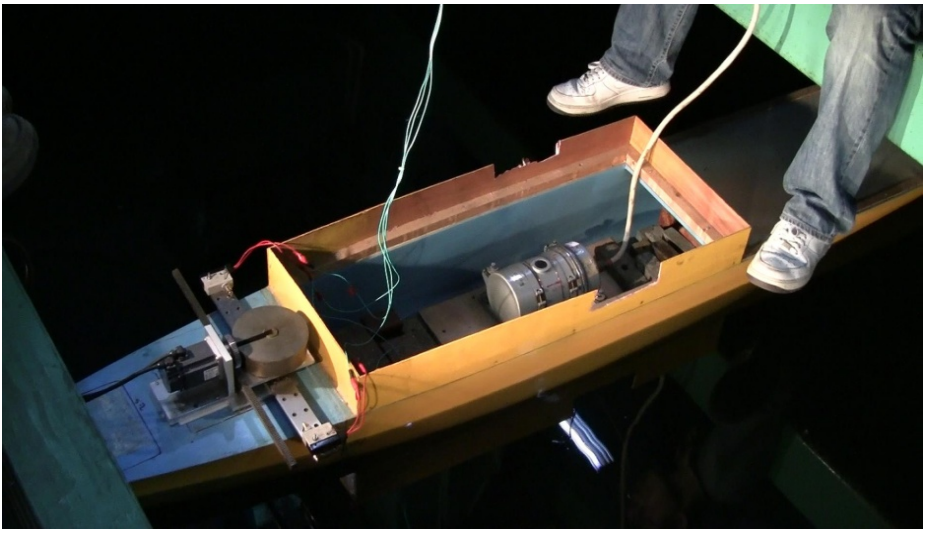


Fig. 6. The test of S60 model

Fig. 6 shows the model test. The natural period of model ship is 1.36s, so the wave in Table 2 was selected.

Table 2. The parameters of the irregular wave in simulation and test

Sea state	Peak period/s	Period range/s	Significant wave height /cm
4	1.19	[0.48 1.76]	3.7
5	1.27	[0.52 1.87]	5.4

After the parameters of ship model, motor and wave are selected, the simulation and test can be conducted.

4.1 PID Feedback Control

Fig. 7 shows the time-travel curves of roll without control and under PID control.

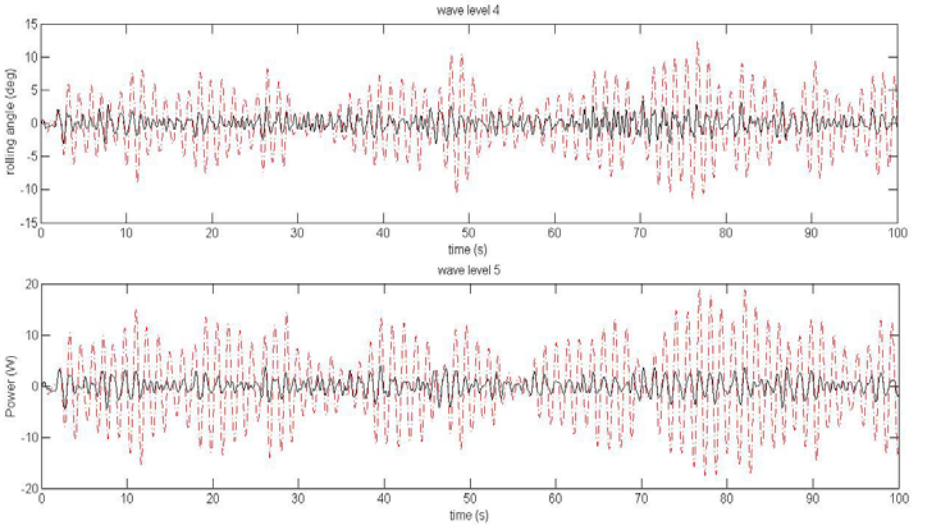


Fig. 7. The time-travel curves of roll under PID feedback control in simulation

Table 3 shows the test results of roll without control and under PID control; Table 4 shows the displacement of mass and the output power of motor in simulation.

Table 3. The test results of roll without control and under PID control

Sea state	The significant value of roll/deg		The max value of roll/deg	
	Without control	PID feedback control	Without control	PID feedback contro
4	6.48	4.09	12.49	3.26
5	8.7	4.72	18.74	3.65

Table 4. The displacement of mass and the output power of motor in simulation

Sea state	The displacement of mass/cm		The output power of motor /w	
	The significant value	The max value	The significant value	The max value
4	1.85	4.53	1.98	9.84
5	1.9	3.94	1.05	3.49

4.2 Model Predictive Control (MPC)

Fig. 8 shows the time-travel curves of roll without control and under MPC.

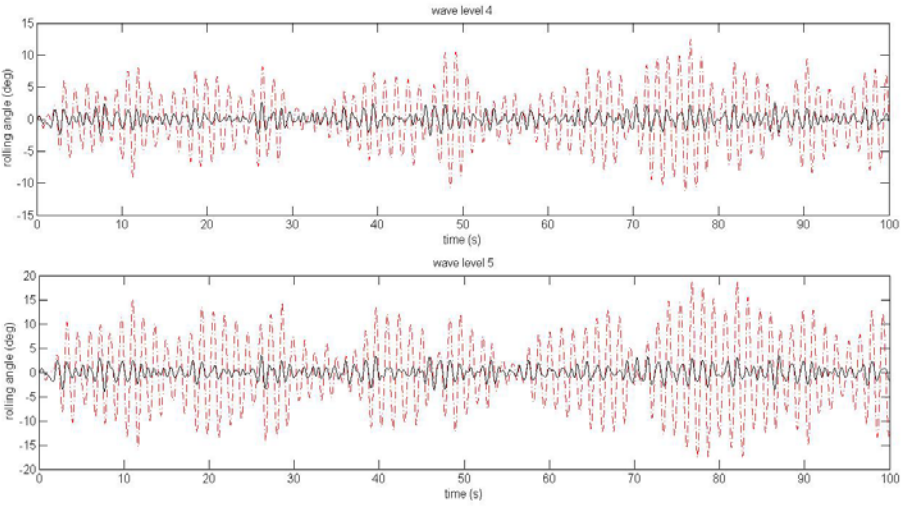


Fig. 8. The time-travel curves of roll without control and under MPC

Table 5 shows the test results of roll without control and under MPC, and Table 6 shows the displacement of mass and the output power of motor in simulation.

As prediction is introduced, with the constraint condition taken into account and the control effect optimized, MPC reduces the motion frequency and amplitude of moving mass, reduces the output power of motor, and maintains good anti-rolling performance.

Table 5. The test results of roll without control and under MPC

Sea state	The significant value of roll/deg		The max value of roll/deg	
	Without control	MPC	Without control	MPC
4	6.48	3.67	12.49	2.61
5	8.7	4.43	18.74	3.15

Table 6. The displacement of mass and the output power of motor in simulation

Sea state	The displacement of mass/cm		The output power of motor /w	
	The significant value	The max value	The significant value	The max value
4	1.45	2.1	0.59	1.28
5	2.7	4.04	2.03	4.05

5 Conclusions

1) The active moving mass system has good anti-rolling performance on the ship. It can reduce the roll angle up to 70%. Because of the constraints of weight and actuator, it is suitable for small and medium ship.
2) PID feedback control is effective for moving mass system. It can greatly reduce the roll greatly in resonance region. Therefore the strategy to damp roll by increasing the damping is effective. It has noticeable anti-rolling effects on resonance region, but little on low and high frequency band. Meanwhile adopting angular velocity as feedback in PID feedback control can lead to obvious noise disturbance and uncontrollable vibration.
3) Model predictive control (MPC) is also very effective. As the prediction is introduced, with the constraint condition taken into account and the control effect optimized, MPC reduces the motion frequency and amplitude of moving mass, reduces the output power of motor, and maintains good anti-rolling performance.
4) Although MPC control system has better undershoot and stability, it is difficult to analyze stability, especially with constraints.

References

1. Bhattacharyya, R.: Dynamics of Marine Vehicles, pp. 364–401. Wiley, America (1978)
2. Perez, T.: Ship Motion Control, pp. 116, 119, 120. Springer, Germany (2005)
3. Sheng, Z., Liu, Y.: Ship Theory, pp. 378–380. Shanghai Jiao Tong University Press, Shanghai (2008)
4. SIREHNA, DCNS center of excellence in hydrodynamics, http://www.sirehna.com
5. Treakle III, T.W.: A Time-Domain Numerical Study of Passive and Active Anti-roll Tanks to Reduce Ship Motions, pp. 3, 11–14. Faculty of the Virginia Polytechnic Institute and State University (1998)

Research and Development of Portable Data Acquisition System Based on the OMAP-L138

Qing Zhu and Feiyun Xu

Research Center of Conditon Monitoring and Fault Diagnosis,
Southeast University, Jiangning District, 211189, Nanjing, Jiangsu, China
zqing_2009@163.com,fyxu@seu.edu.cn

Abstract. The OMAP-L138 processor of TI has the characteristics of high-performance, small size and low-power consumption, it is widely applied to the development of portable device. In order to obtain the best performance of microprocessor, communication between ARM side and DSP side becomes the key problem of the application of system development. The development method of portable data acquisition system based on the OMAP-L138 platform is studied in this paper. Dual-core communication program based on DSP/BIOS Link is designed in paper makes the collected data can be transmitted to the ARM side and the data can be used to real-time display of waveform and a new solution for developing portable data acquisition analysis system is put forward in this paper.

Keywords: DSP/BIOS Link, OMAP-L138, dual-core communication, portable data acquisition system.

1 Introduction

With the development of embedded computer technology, contradiction between high performance and small size, low power consumption of microprocessor increases prominently. Open multimedia application platform (OMAP) launched by TI offers a good solution for the problem[1]. Therefore, the authors adopt the OMAP-L138 application processor in the portable data acquisition system. The processor seamlessly integrates ARM926EJ-S RISC core and TMS320C6748 DSP core[2]. By using the ARM9 core on ship, the development platform can support various of embedded operating systems that can realize the whole control of system. With the introduction of low power consumption fixed-point DSP core of C67x series, developers can make full use of its high intensity real-time processing ability to implement various signal processing algorithms, which improves execution speed and decreases the power consumption of system. In order to achieve multi-channel real-time data acquisition function of portable data acquisition system based on the OMAP-L138 platform, information and data transfer between dual-core becomes the key problem. There are two kinds of dual-core communication software commonly used: one kind is Gateway DSP that developed by Nokia and the other is the DSP/BIOS Link that developed by TI. This paper mainly studies the constructional method of DSP/BIOS Link on the OMAP-L138 platform and designs the program of dual-core communication based on

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 185–192.
springerlink.com

DSP/BIOS Link, and realizes the development of portable data acquisition system which is verified by experiment. All in all, a new solution for developing portable data acquisition analysis system is put forward in this paper.

2 Fundamentals of DSP/BIOS Link

DSP/BIOS Link abbreviates to DSPLink and includes versions of Windows environment and Linux environment. It is foundation software for the inter-processor communication across the GPP-DSP boundary. It eliminates the need for developers to develop such dual-core communication program and makes them concentrate on the development of the needed application. It provides a generic API that abstracts the characteristics of the physical link connecting GPP and DSP from the applications. Users can choose the desired API upon the demand of development.

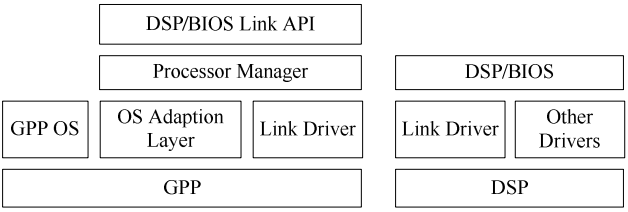


Fig. 1. Software structure of DSP/BIOS Link [3]

The software structure of DSP/BIOS Link can be seen in figure 1. On the GPP side, the OS Adaptation Layer encapsulates the generic OS services which are required by the other components of DSP/BIOS Link. This component exports a generic API that isolates the other components from the specifics of an OS. All other components use this API instead of direct OS calls, which makes DSP/BIOS Link convenient across different OS. The Link Driver encapsulates the low-level control operations on the physical link between the GPP and DSP. This module is responsible for controlling the execution of the DSP and data transfer by using defined protocol across the GPP-DSP boundary. The Processor Manager maintains book-keeping information for all components and allows different boot-loaders to be plugged into the system. Besides, it builds exposes the control operations provided by the Link Driver to the user through the API layer. DSP/BIOS Link API is the interface for all clients on the GPP side. On the DSP side, as one of the drivers in DSP/BIOS, the Link Driver specializes in communicating with the GPP over the physical link and the communication is done using the DSP/BIOS modules.

3 Construction of DSP/BIOS Link Development Platform

In this paper, DSP/BIOS Link adopts Linux environment version. Evaluation board needs to establish development platform before installation of DSP/BIOS Link. The work mainly includes customizing Linux kernel and establishing cross-compiling

environment [4]. In order to avoid frequent burning of Flash, the building process of the development platform can be designed as follows: Linux kernel can be downloaded via the TFTP service into the RAM and mounts the NFS file system to start. This method also can be used for availability testing of kernel and applications. After completing applications development, Linux kernel can be burned into Flash and mounts the jffs2 file system to start. All applications can be released on it.

Before compiling the source program of DSP/BIOS Link, it needs to set the environment variables and building properties of DSP/BIOS Link. Firstly, the installation path and add the path of script compiler to PATH differently upon the different types of shell used by Linux server when the environment variables of DSP/BIOS Link are set. So, it is needed to modify the DSPLINK and PATH in dsplinkenv script under different environment variables. Secondly, building properties of DSP/BIOS Link needs to configuration. DSP/BIOS Link provides interactive script named dsplinkcfg.pl that wrote by perl language. The script uses to configure building properties of DSP/BIOS Link. The appropriate configuration selected by the actual development platform. The options of configuration contain the processor's types, the OS of GPP, the OS of DSP and configuring properties of DSP/BIOS Link. Makefile's information of ARM side and DSP side will be got after completing all configuration work. Because of the installation path of different users installs differently, the path of cross-compile tool and embedded Linux kernel that burns into Flash in Makefile of two sides needs to be modified. All work makes the program of DSP/BIOS Link can be compiled via call the cross-compiler and DSP/BIOS Link platform can be built upon the kernel information of user's configuration.

A kernel file named dsplinkk.ko will be got after completing the installation of DSP/BIOS Link. In order to verify the correctness of the constructed platform, the boot-mode of Linux kernel needs to modify to mount NFS file system. DSP/BIOS Link module will be worked after creating an equipment node of dsplink in the file system and executing the insmod dsplinkk.ko instruction. If the platform constructs successfully, the serial console will print out information as shown in figure 2.

```
root@seed:/opt/dsplink# mknod /dev/dsplink c 230 0
root@seed:/opt/dsplink# insmod dsplinkk.ko
DSPLINK Module (1.65.00.03) created on Date: Nov 18 2010 Time: 08:30:04
```

Fig. 2. Result of DSP/BIOS Link module takes effect

4 Realization of Portable Data Acquisition System

Hardware frame design of portable data acquisition system as shown in figure 3. In order to reduce the burden of ARM core and assign tasks reasonably, the ADS8556 is controlled by DSP core. Software development of portable data acquisition system can be divided into DSP side and ARM side, and dual-core communication program based on DSP/BIOS Link plays a bridging role in the programs on both sides. Therefore, dual-core communication program of the system becomes the key of software development. Software designing on DSP side includes the design of data acquisition program and dual-core communication program, and this task can be worked on CCS. After opening the Samba service on Linux server, CCS can call

DSPLink header files and library files of DSP side via the service. The executable file of DSP side will be generated after linking and compiling. Software designing on ARM side includes the design of real-time display program and dual-core communication program. Because of the Qt/Embedded is widely used in developing embedded GUI, the development of real-time display program uses these tools and Qt creator. In order to make Qt creator can call library, head and API of DSPLink, it is needed to add LIBS += /…/dsplink.lib and QMAKE_CXXFLAGS +=(shell cat /…/API_defines.txt) (shell cat /…/API_includes.txt) in .pro file. So, the development of program in Linux system can be carried out in Qt creator entirely.

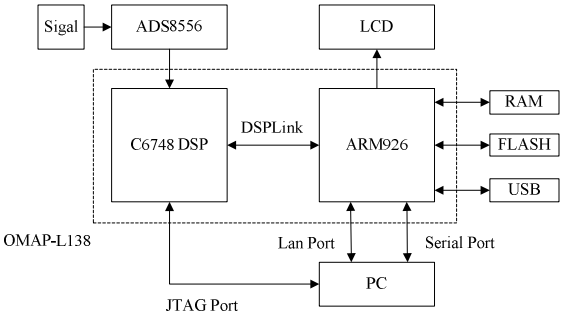


Fig. 3. Hardware frame design of portable data acquisition system

4.1 Program Designing of DSP Side

DSP/BIOS is a scalable real-time kernel. It is designed to be used by applications that require real-time scheduling and synchronization, host-to-target communication, or real-time instrumentation [5]. Introduction of DSP/BIOS make it more convenient for users to write multi-tasking applications. Considering the subsequent development of data acquisition system and the addition of various functions, application development of DSP side uses DSP/BIOS program. Handle of required tasks is created after initializing DSPLink module, the task thread will be executed on DSP. In order to make DSP exchange information with ARM, DSP needs to define an information structure named TSKDSPARM_TransferInfo to interact with ARM. The content of structure includes the message queue head, the parameters of the interactive information and the STOP flag, and these parameters will be sent with the message queue (MSGQ). By using the provided API of DSPLink, DSP operates the MSGQ will become very simple. MSGQ can be created with the usage of MSGQ_open() ,and named "DSPMSGQ" which is similar to its identifier. Depending on this identifier, ARM can use MSGQ_locate() to locate MSGQ of DSP side. ARM gets the MSGQ of DSP side only after location. Dual-core communication program is the core program of DSP side. Program design focuses on the sending sequence of MSGQ between ARM and DSP to ensure that a new set of collected data in dspAdBuffer can be read by ARM timely. Flow chart can be seen in figure 4. When DSP receives the MSGQ from ARM, it will determine whether to enter the data acquisition flow or not according to the STOP flag. If the STOP flag is no, the program will go into an infinite loop, so as to realize the continuous acquisition of

data, and take nSampleNum as the sampling points flag of system. When nSampleNum reaches a predetermined point, DSP will read the data from AD_BUF[6][ADLENGTH] into dspAdBuffer. After that, DSP will send MSGQ to ARM and inform the ARM to read the data into armAdBuffer. When DSP receives the MSGQ from ARM again, the next set of data acquisition will be started.

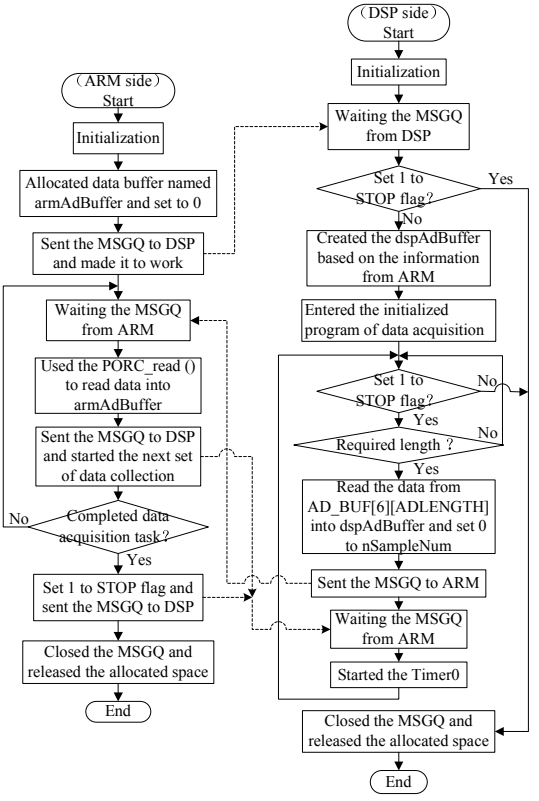


Fig. 4. Flow chart of dual-core communication

In the program on DSP side, if the STOP flag is not for 1, the program will go into data acquisition flow. The data acquisition flow combines with the flow of dual-core communication. In this system, data acquisition program is designed by regular sampling. ADS8556 starts to work after the board is powered on. External conversion clock of ADS8556 is input by CPLD with the frequency is 630 kHz. Thus, ADS8556 obtains its maximum sampling frequency. When ADS8556 begins to converse of sampling value, the BUSY flag will be set to 1 and the flag will be set to 0 after accomplishing. This event will trigger GPIO Bank2 to produce an EDMA event which synchronizes EDMA to move data of 6 channel's 16 bit sampling value into BIN_AD_BUF [6] and data acquisition will be realized. Timer0 is used to change the sampling frequency. After setting a required timer cycle, Timer0 will trigger timer interrupt periodically. When the program enters the timer interrupt service routine,

EDMA event enable and GPIO Bank2 EDMA event enable will be opened. This event makes the BUSY flag of ADS8556 take effect and the sampling frequency is changed. And after this, GPIO Bank2 triggers the EDMA to work and sampling value is transferred by EDMA. Flow chart of data acquisition program as shown in figure 5 and figure 6.

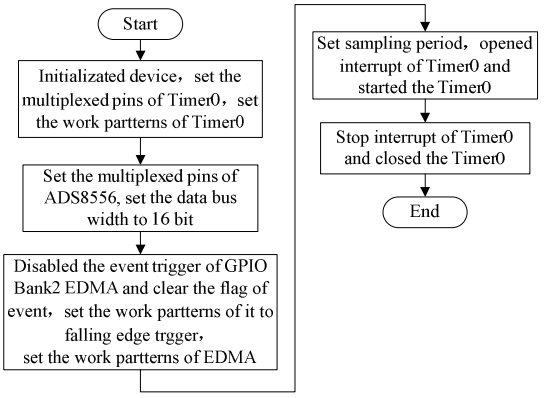


Fig. 5. Initialization flow chart of AD

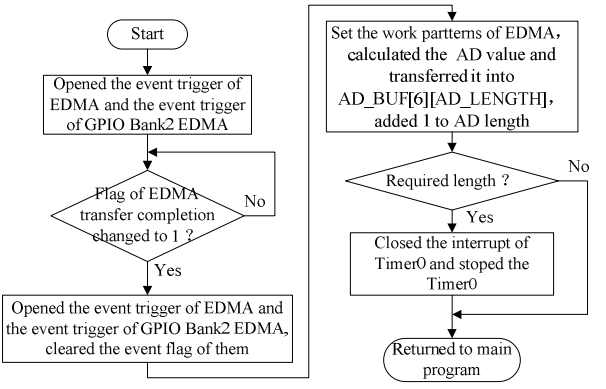


Fig. 6. Flow chart of timer interrupt service routine

4.2 Program Designing of ARM Side

Program designing on ARM side mainly includes the design of the custom drawing control and dual-core communication program. In order to achieve the data constantly reading from DSP and real-time displaying, ARM uses the multithreaded programming technology based on the Qt. In main thread, function of the custom drawing control includes the drawing of coordinates and curve and the renewal of the drawing. The date of drawing curve is stored in type of QVector<QPointF>.

The curve drawing function is used to translate the recorded coordination of QPointF to coordination of control which is stored in multi-line variable. The drawPolyline() is called and used to draw all curves of the multi-line variable. In order to avoid the flashing problem in real-time drawing, double buffering drawing technology is used in the drawing refresh function. The function saves pictures in the variable of QPixmap type and calls the update() to refresh all controls. The painteEvent () will copy all pictures to control and the real-time refresh of pictures is realized.

The running function of ARMReadDSPDataThread thread is used to implement dual-core communication function between ARM and DSP, and the flow chart shown as figure 4. In order to send MSGQ to DSP after DSP awaked by ARM, MSGQ is also needed to create on ARM side and allocated by the MSGQ_locate ().The dual-core communication program on ARM side is responsible for sending MSGQ to DSP and making it collect date. After receiving the MSGQ from DSP, ARM uses the PROC_read () to read the data from dspAdBuffer into armAdBuffer. When the task of data acquisition completes, ARM will set the STOP flag as 1 and send the MSGQ to DSP to make it stop working. SetDataToQPointFThread thread is responsible for reading the date from armAdBuffer into DATA[6], and setting the refresh frequency of drawing as 1second. So, the purpose of real-time drawing is reached.

5 Test and Conclusion

The system uses signal generator and oscilloscope to verify the correctness. The signal generator produces a sinusoidal signal with 20 volt peak-to-peak value, 50 Hz frequency and the data acquisition system synchronously collects the data of 6 channels with 1 kHz sampling frequency and 1024 points of data length. When DSP finished a set of data acquisition, the ARM reads the data and displays the waveform with 1second refresh frequency. As shown in figure 7, the refresh frequency of drawing is set 1 second and the comparison result of the display waveform and the waveform of the oscilloscope can be seen. The test result demonstrates the efficiency of the proposed approach.

With the rapid development of embedded computer technology and advanced microprocessors, significant changes also have taken place in embedded operation systems. The portable instruments based on the embedded system with the characteristics of highly integrated, low power consumption and flexibly usage have great application values. This paper studies the development methodology of portable data acquisition system. The build method of DSP/BIOS Link development platform on the OMAP-L138 is studied and the dual-core communication program based on DSP/BIOS Link is designed in this paper. Multi-channel synchronous data acquisition program and waveform real-time display program is also designed in this paper. Finally, the system is tested. The test result has verified the practicability of the design. Thus, a new solution for developing portable data acquisition system is realized and a valuable reference for developers who use the OMAP-L138 processor first time to solve the problem of dual-core communication.

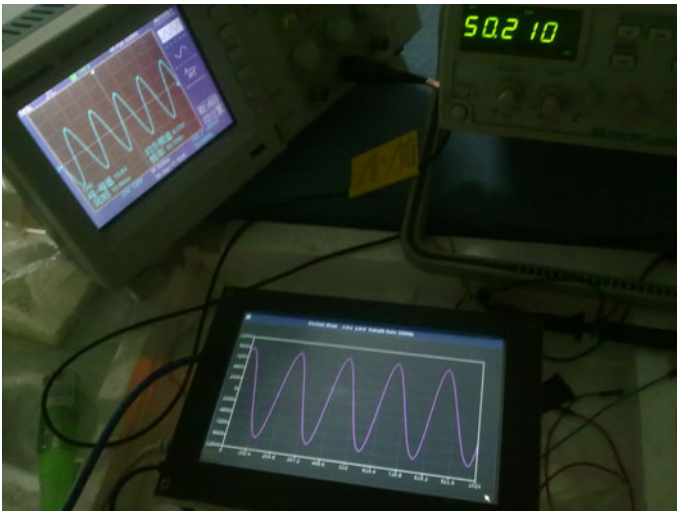


Fig. 7. Experimental chart of data acquisition system

References

1. Peng, Q., Yang, L., Pan, Y.: Open Multimedia Application Platform-The Principle and Application of OMAP Processor, pp. 1–3. Publish House of Electronics Industry, Beijing (2005)
2. Texas Instruments. OMAP-L138 Applications Processor System Reference Guide (2010), http://focus.ti.com/lit/ug/sprugm7d/sprugm7d.pdf
3. Texas Instruments. DSP/BIOS Link User's Guide, (2010), http://software-dl.ti.com/dsps/dsps_public_sw/sdo_sb/targetcontent/DSPLink/1_63/index_FDS.html
4. Texas Instruments. DSP/BIOS Link Installation Guide, (2010), http://software-dl.ti.com/dsps/dsps_public_sw/sdo_sb/targetcontent/DSPLink/1_63/index_FDS.html
5. Texas Instruments. TMS320 DSP/BIOS v5.41 User's Guide, (2009), http://focus.ti.com.cn/cn/general/docs/litabsmultiplefilelist.tsp?literatureNumber=spru423h

An Application of Generalized Interval-Valued Intuitionistic Fuzzy Soft Sets in a Decision Making Problem

Young Chel Kwun[1], Jin Han Park[2,*], Ja Hong Koo[3], and Yong Kyun Lee[4]

[1] Department of Mathematics, Dong-A University, Busan 604-714, Korea
[2] Department of Applied Mathematics, Pukyong National University, Busan 608-737, Korea
[3] Department of Fiber Plastic Design, Dong-A University, Busan 604-714, Korea
[4] Department of Chemistry, Dong-A University, Busan 604-714, Korea
jihpark@pknu.ac.kr

Abstract. We present a method of object recognition from an imprecise multiobserver data, which extends the work of Roy and Maji [11] to generalized interval-valued intuitionistic fuzzy soft set theory. The method involves the construction of Comparison Table from a generalized intuitionistic fuzzy soft set in a parametric sense for decision making.

Keywords: Generalized interval-valued intuitionistic fuzzy soft sets, Comparison Table, object recognition.

1 Introduction

Molodtsov [7] introduced soft sets as a mathematical tool for dealing with uncertainties which is free from the above-mentioned difficulties. Since the soft set theory offers mathematical tool for dealing with uncertain, fuzzy and not clearly defined objects, it has a rich potential for applications to problems in real life situation. He showed how soft set theory is free from the parameterization inadequacy syndrome of fuzzy set theory, rough set theory, probability theory and game theory. With this point of view, the classical soft sets have been extended to fuzzy soft sets [9,10], intuitionistic fuzzy soft sets [11,12], vague soft sets [13] and interval-valued intuitionistic fuzzy soft sets [14], respectively. In this paper, we present some results on an application of generalized interval-valued intuitionistic fuzzy soft sets in decision making problem. The problem of object recognition has received paramount importance in recent times. The recognition problem may be viewed as a multiobserver decision making problem, which the final identification of the object is based on the set of inputs from different observers who provide the overall object characterization in terms of diverse sets of parameters. A generalized interval-valued intuitionistic fuzzy soft set theoretic approach to solution of the above decision making problem is presented.

* Corresponding author.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 193–198.
springerlink.com

2 Generalized Interval-Valued Intuitionistic Fuzzy Soft Sets in Decision Making

In this section we present generalized interval-valued intuitionistic fuzzy soft set and some results of it.

Definition 1. A generalized interval-valued intuitionistic fuzzy set A on nonempty set X is an object of the form $A = \{\langle x, \mu_A(x), \gamma_A(x)\rangle : x \in X\}$, where $\mu_A : X \to L$ and $\gamma_A : X \to L$ (L denotes the set of all interval numbers over $[0,1]$) with the property $\underline{0} \le \mu_A(x) \le \gamma(x)^c \le \underline{1}$ for any $x \in X$ and denote $\mu_A(x) = [\mu_A(x)^-, \mu_A(x)^+]$ and $\gamma_A(x) = [\gamma_A(x)^-, \gamma_A(x)^+]$. Here, $\underline{0} = [0,0]$ and $\underline{1} = [1,1]$, and we define operations as follows: $a \le b \Leftrightarrow a^- \le b^-$ and $a^+ \le b^+$ for any $a = [a^-, a^+]$, $b = [b^-, b^+] \in L$ and $a^c = [1-a^+, 1-a^-]$ for any $a \in L$. Denote by $GIIF(X)$ denotes the set of all generalized interval-valued intuitionistic fuzzy sets of X.

Let $U = \{h_1, h_2, \cdots, h_n\}$ be the set of n objects, which may be characterized by a set of parameters $\{A_1, A_2, \cdots, A_i\}$. The parameter space E may be written as $E \supseteq A_1 \cup A_2 \cup \cdots \cup A_i$. Let each parameter set A_i represents the i th class of parameters and the elements of A_i represents a specific property set. Here we assume that these property sets may be viewed as generalized interval-valued intuitionistic fuzzy sets.

Definition 2. Let $GIIF(U)$ denotes the set of all generalized interval-valued intuitionistic fuzzy sets of U. Let $A_i \subseteq E$. A pair $\langle F_i, A_i\rangle$ is a generalized interval valued intuitionistic fuzzy soft set (**GIIF** soft set for short) over U, where F_i is a mapping given by $F_i : A_i \to GIIF(U)$.

Definition 3. Let $\langle F, A\rangle$ and $\langle G, B\rangle$ be two **GIIF** soft sets over U. Then $\langle F, A\rangle$ is said to be a **GIIF** soft subset of $\langle G, B\rangle$, denoted by $\langle F, A\rangle \tilde{\subset} \langle G, B\rangle$, if

(1) $A \subseteq B$;

(2) for any $\varepsilon \in A$, $F(\varepsilon)$ is a generalized interval-valued intuitionistic fuzzy subset of $G(\varepsilon)$, that is, for all $x \in U$ and $\varepsilon \in A$, $\mu_{F(\varepsilon)}(x) \le \mu_{G(\varepsilon)}(x)$ and $\gamma_{F(\varepsilon)}(x) \ge \gamma_{G(\varepsilon)}(x)$, i.e., $\mu^-_{F(\varepsilon)}(x) \le \mu^-_{G(\varepsilon)}(x)$, $\mu^+_{F(\varepsilon)}(x) \le \mu^+_{G(\varepsilon)}(x)$, $\gamma^-_{F(\varepsilon)}(x) \ge \gamma^-_{G(\varepsilon)}(x)$ and $\gamma^+_{F(\varepsilon)}(x) \ge \gamma^+_{G(\varepsilon)}(x)$.

Definition 4. Let $\langle F, A\rangle$ and $\langle G, B\rangle$ be two **GIIF** soft sets over a universe U. Then "$\langle F, A\rangle$ and $\langle G, B\rangle$ "is a **GIIF** soft set, denoted by $\langle F, A\rangle \wedge \langle G, B\rangle$, is defined by $\langle F, A\rangle \wedge \langle G, B\rangle = \langle H, A \times B\rangle$,where $H(\alpha, \beta) = F(\alpha) \cap G(\beta)$ for any $(\alpha, \beta) \in A \times B$, that is $H(\alpha, \beta)(x) = \langle[\min\{\mu^-_{F(\alpha)}(x), \mu^-_{G(\beta)}(x)\}, \min\{\mu^+_{F(\alpha)}(x), \mu^+_{G(\beta)}(x)\}], [\max\{\gamma^-_{F(\alpha)}(x), \gamma^-_{G(\beta)}(x)\}, \max\{\gamma^+_{F(\alpha)}(x), \gamma^+_{G(\beta)}(x)\}]\rangle$, for all $(\alpha, \beta) \in A \times B$ and $x \in U$.

2.1 Comparison Table

The Comparison Table of **GIIF** soft set $\langle F, A\rangle$ is a square table in which number of rows and number of columns are equal, rows and columns are labeled by the object names $h_1, h_2, \cdots, h_n$ of the universe set U, and the entries c_{ij} $(i, j = 1,2,\cdots,n)$ is the number of parameters satisfying $\mu_{ik} \geq \mu_{jk}$ and $\gamma_{ik} \leq \gamma_{jk}$, where $\mu_{ik} = [\mu_{ik}^-, \mu_{ik}^+]$ and $\mu_{jk} = [\mu_{jk}^-, \mu_{jk}^+]$ are, respectively, the interval degrees of membership of h_i and h_i in $F(e_k)$ for all k, and $\gamma_{ik} = [\gamma_{ik}^-, \gamma_{ik}^+]$ and $\gamma_{jk} = [\gamma_{jk}^-, \gamma_{jk}^+]$ are, respectively, the interval degrees of non-membership of h_i and h_i in $F(e_k)$ for all k, where k is the number of parameters presented in a **GIIF** soft set. Clearly, $0 \leq c_{ij} \leq k$ for any i, j, where k is the number of parameters in A. Thus, c_{ij} indicates a numerical measure which h_i dominates h_j in c_{ij} number of parameters out of k parameters.

2.2 Row Sum, Column Sum and Score of an Object

- The row sum r_i of an object h_i is calculated by $r_i = \sum_{j=1}^{n} c_{ij}$. Clearly, r_i indicates the total number of parameters in which h_i dominates all the members of U.
- The column sum t_j of an object h_j is calculated by $t_j = \sum_{i=1}^{n} c_{ij}$. The integers t_j indicate the total number of parameters in which h_j is dominates by all the numbers of U.
- The score s_i of an object h_i is calculated by $s_i = r_i - t_i$.

2.3 Algorithm

The problem here is to choose an object from the set of given objects with respect to a set of choice parameters P. We now present an algorithm for identification of an object, based on multiobservers input data characterized by color, size and surface texture features.

1. Input the **GIIF** fuzzy soft sets $\langle F, A\rangle$, $\langle G, B\rangle$ and $\langle H, C\rangle$.
2. Input the parameter set P as observed by the observers.
3. Compute the corresponding resultant **GIIF** soft set $\langle S, P\rangle$ from the **GIIF** soft sets $\langle F, A\rangle$, $\langle G, B\rangle$ and $\langle H, C\rangle$ and place it in tabular form.
4. Construct the Comparison Table of the **GIIF** soft set $\langle H, C\rangle$ and compute row sum r_i and column sum t_i of h_i for all i.
5. Compute the score of h_i for all i.
6. The decision is s_k if $s_k = \max_i s_i$.
7. If k has more than one value then any one of h_k may be chosen.

3 An Application in a Decision Making Problem

Consider the problem of selecting the most suitable object from the set of objects with respect to a set of choice parameters. Let $U = \{h_1, h_2, h_3, h_4, h_5, h_6\}$ be the set of

objects having different colors, sizes and surface texture features. Let E = {blackish, dark brown, yellowish, reddish, small, very small, average, large, very large, course, moderately course, fine, extra fine} be the set of parameters. Let A, B and C be three subsets of E such that A = {blackish, dark brown, reddish, yellowish} represents the color space B = {small, average, large, very large} represents the size of the object, and C = {course, moderately course, fine, extra fine} represents the surface texture granularity.

Assuming that the **GIIF** soft set $\langle F, A\rangle$ describes the 'objects having color space', the **GIIF** soft set $\langle G, B\rangle$ describes the 'objects having size' and the **GIIF** soft set $\langle H, C\rangle$ describes the 'texture feature of the objects surface'. The problem is identify an unknown object from multiobservers generalized interval-valued intuitionistic fuzzy data, specified by different observers, in terms of **GIIF** soft sets $\langle F, A\rangle$, $\langle G, B\rangle$ and $\langle H, C\rangle$. The tabular representations of **GIIF** soft sets $\langle F, A\rangle$, $\langle G, B\rangle$ and $\langle H, C\rangle$ are shown in Tables 1-3.

Table 1. Tabular representation of the **GIIF** soft set $\langle F, A\rangle$

U	blackish (a_1)	dark brown (a_2)	yellowish (a_3)	reddish (a_4)
h_1	⟨[0.2,0.5],[0.4,0.7]⟩	⟨[0.2,0.3],[0.4,0.6]⟩	⟨[0.5,0.6],[0.4,0.5]⟩	⟨[0.4,0.5],[0.2,0.3]⟩
h_2	⟨[0.2,0.4],[0.3,0.7]⟩	⟨[0.5,0.6],[0.3,0.4]⟩	⟨[0.1,0.4],[0.3,0.8]⟩	⟨[0.2,0.5],[0.5,0.6]⟩
h_3	⟨[0.2,0.3],[0.4,0.5]⟩	⟨[0.2,0.3],[0.5,0.6]⟩	⟨[0.1,0.4],[0.4,0.8]⟩	⟨[0.2,0.5],[0.4,0.7]⟩
h_4	⟨[0.1,0.4],[0.3,0.8]⟩	⟨[0.2,0.5],[0.2,0.7]⟩	⟨[0.3,0.4],[0.4,0.5]⟩	⟨[0.1,0.4],[0.3,0.8]⟩
h_5	⟨[0.2,0.5],[0.4,0.7]⟩	⟨[0.2,0.4],[0.3,0.6]⟩	⟨[0.2,0.3],[0.4,0.6]⟩	⟨[0.2,0.3],[0.4,0.5]⟩
h_6	⟨[0.1,0.6],[0.3,0.9]⟩	⟨[0.1,0.7],[0.2,0.8]⟩	⟨[0.2,0.3],[0.4,0.6]⟩	⟨[0.1,0.4],[0.3,0.8]⟩

Table 2. Tabular representation of the **GIIF** soft set $\langle G, B\rangle$

U	large (b_1)	very large (b_2)	small (b_3)	average (b_4)
h_1	⟨[0.1,0.4],[0.4,0.8]⟩	⟨[0.1,0.4],[0.2,0.9]⟩	⟨[0.1,0.3],[0.2,0.3]⟩	⟨[0.2,0.5],[0.5,0.6]⟩
h_2	⟨[0.1,0.4],[0.3,0.8]⟩	⟨[0.2,0.5],[0.3,0.6]⟩	⟨[0.2,0.4],[0.3,0.7]⟩	⟨[0.2,0.3],[0.5,0.7]⟩
h_3	⟨[0.2,0.3],[0.4,0.6]⟩	⟨[0.2,0.3],[0.4,0.5]⟩	⟨[0.3,0.4],[0.4,0.5]⟩	⟨[0.2,0.5],[0.4,0.7]⟩
h_4	⟨[0.1,0.3],[0.2,0.7]⟩	⟨[0.1,0.4],[0.6,0.8]⟩	⟨[0.1,0.7],[0.2,0.8]⟩	⟨[0.2,0.3],[0.5,0.7]⟩
h_5	⟨[0.1,0.4],[0.2,0.9]⟩	⟨[0.1,0.7],[0.1,0.9]⟩	⟨[0.1,0.4],[0.2,0.5]⟩	⟨[0.2,0.3],[0.4,0.6]⟩
h_6	⟨[0.2,0.4],[0.3,0.7]⟩	⟨[0.1,0.4],[0.2,0.9]⟩	⟨[0.1,0.4],[0.3,0.8]⟩	⟨[0.2,0.3],[0.4,0.5]⟩

Table 3. Tabular representation of the **GIIF** soft set $\langle H, C\rangle$

U	course (c_1)	moderately course (c_2)	fine (c_3)	extra fine (c_4)
h_1	⟨[0.1,0.4],[0.3,0.8]⟩	⟨[0.1,0.4],[0.4,0.8]⟩	⟨[0.1,0.6],[0.1,0.8]⟩	⟨[0.1,0.4],[0.2,0.3]⟩
h_2	⟨[0.2,0.5],[0.3,0.6]⟩	⟨[0.3,0.4],[0.5,0.5]⟩	⟨[0.2,0.5],[0.4,0.7]⟩	⟨[0.3,0.4],[0.5,0.6]⟩
h_3	⟨[0.2,0.3],[0.4,0.5]⟩	⟨[0.2,0.3],[0.4,0.6]⟩	⟨[0.2,0.4],[0.3,0.7]⟩	⟨[0.1,0.2],[0.4,0.5]⟩
h_4	⟨[0.2,0.3],[0.5,0.7]⟩	⟨[0.2,0.4],[0.5,0.6]⟩	⟨[0.4,0.5],[0.5,0.6]⟩	⟨[0.2,0.4],[0.3,0.7]⟩
h_5	⟨[0.2,0.3],[0.4,0.6]⟩	⟨[0.2,0.3],[0.4,0.6]⟩	⟨[0.2,0.3],[0.4,0.5]⟩	⟨[0.2,0.3],[0.4,0.6]⟩
h_6	⟨[0.2,0.3],[0.3,0.8]⟩	⟨[0.2,0.5],[0.4,0.7]⟩	⟨[0.2,0.3],[0.7,0.8]⟩	⟨[0.1,0.4],[0.3,0.8]⟩

Let us now see how the algorithm may be used to solve original problem. Consider the **GIIF** soft sets $\langle F, A\rangle$, $\langle G, B\rangle$ and $\langle H, C\rangle$ as defined above. Suppose that $P = \{(a_1 \wedge b_1) \wedge c_1, (a_1 \wedge b_4) \wedge c_3, (a_2 \wedge b_1) \wedge c_2, (a_3 \wedge b_3) \wedge c_3\}$ be the set of choice parameters of an observer. On the basis of this parameter we have to take the decision from the availability set U. The tabular representation of **GIIF** soft set $\langle S, P\rangle$ will be as. (Table 4).

The Comparison Table of the above resultant **GIIF** soft set is as below (Table 5).

Table 4. Tabular representation of the resultant **GIIF** soft set $\langle S, P\rangle$

U	$(a_1 \wedge b_1) \wedge c_1$	$(a_1 \wedge b_4) \wedge c_3$	$(a_2 \wedge b_1) \wedge c_2$	$(a_3 \wedge b_3) \wedge c_3$
h_1	⟨[0.1,0.4],[0.4,0.8]⟩	⟨[0.1,0.5],[0.5,0.8]⟩	⟨[0.1,0.3],[0.4,0.8]⟩	⟨[0.1,0.3],[0.4,0.8]⟩
h_2	⟨[0.1,0.4],[0.3,0.8]⟩	⟨[0.2,0.3],[0.5,0.7]⟩	⟨[0.1,0.4],[0.5,0.8]⟩	⟨[0.1,0.4],[0.4,0.8]⟩
h_3	⟨[0.2,0.3],[0.4,0.5]⟩	⟨[0.2,0.3],[0.4,0.7]⟩	⟨[0.2,0.3],[0.5,0.6]⟩	⟨[0.1,0.4],[0.4,0.8]⟩
h_4	⟨[0.1,0.3],[0.5,0.8]⟩	⟨[0.1,0.3],[0.5,0.8]⟩	⟨[0.1,0.3],[0.5,0.7]⟩	⟨[0.1,0.4],[0.5,0.8]⟩
h_5	⟨[0.1,0.3],[0.4,0.9]⟩	⟨[0.2,0.3],[0.4,0.7]⟩	⟨[0.1,0.3],[0.4,0.9]⟩	⟨[0.1,0.3],[0.4,0.6]⟩
h_6	⟨[0.1,0.3],[0.3,0.9]⟩	⟨[0.1,0.3],[0.7,0.9]⟩	⟨[0.1,0.4],[0.4,0.8]⟩	⟨[0.1,0.3],[0.7,0.8]⟩

Table 5. Comparison Table of the resultant **GIIF** soft set $\langle S, P\rangle$

	h_1	h_2	h_3	h_4	h_5	h_6
h_1	4	0	0	2	2	2
h_2	2	4	1	3	1	3
h_3	1	2	4	4	2	2
h_4	0	0	0	4	0	2
h_5	1	1	1	1	4	2
h_6	1	1	0	0	2	4

Next we compute the row sum (r_i), column sum (t_i), and the score (s_i) for each (h_i) as shown below (Table 6). From the above score table, it is clear that the maximum score is 9, scored by h_3 and the decision is in favour of selecting h_3.

Table 6. The row sum, column sum and score of h_1

	row sum (r_i)	column sum (t_i)	score (s_i)
h_1	10	9	1
h_2	14	8	6
h_3	15	6	9
h_4	6	14	-8
h_5	10	11	-1
h_6	8	15	-7

4 Conclusions

In this paper, we give an application of **GIIF** soft theory in object recognition problem. The recognition strategy is based on multiobserver input parameter data set. The algorithm involves the construction of Comparison Table from the resultant **GIIF** soft set and the final decision is taken based on the maximum score computed from the Comparison Table (Tables 4 and 5). To extend our work, further research could be done to study the issues on the parameterization reduction of the **GIIF** soft sets, and to explore the applications of using the **GIIF** soft set approach to solve real world problems such as decision making, forecasting and data analysis.

References

1. Varadhan, S.R.S.: Probability Theory. American Mathematical Society (2001)
2. Zadeh, L.A.: Fuzzy sets. Inform Control 8, 338–353 (1965)
3. Zadeh, L.A.: Is there a need for fuzzy logic. Information Sciences 178, 2751–2779 (2008)
4. Atanassov, K.: Intuitionistic fuzzy sets. Fuzzy Sets and Systems 20, 87–96 (1986)
5. Gau, W.L., Buehrer, D.J.: Vague sets. IEEE Transactions on Systems, Man and Cybernetics 23, 610–614 (1993)
6. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers (1991)
7. Molodtsov, D.: Soft set theory - first results. Computers and Mathematics with Applications 37, 19–31 (1999)
8. Maji, P.K., Biswas, R., Roy, A.R.: Soft set theory. Computers and Mathematics with Applications 45, 555–562 (2003)
9. Maji, P.K., Biswas, R., Roy, A.R.: Fuzzy soft sets. Journal of Fuzzy Mathematics 9, 589–602 (2001)
10. Roy, A.R., Maji, P.K.: A fuzzy soft set theoretic approach to decision making problems. Journal of Computational and Applied Mathematics 203, 412–418 (2007)
11. Maji, P.K., Biswas, R., Roy, A.R.: Intuitionistic fuzzy soft sets. Journal of Fuzzy Mathematics 9, 677–692 (2001)
12. Maji, P.K., Roy, A.R., Biswas, R.: On intuitionistic fuzzy soft sets. Journal of Fuzzy Mathematics 12, 669–683 (2004)
13. Xu, W., Ma, J., Wang, S., Hao, G.: Vague soft sets and their properties. Computers and Mathematics with Applications 59, 787–794 (2010)
14. Min, W.K.: Interval-valued intuitionistic fuzzy soft sets. Journal of Fuzzy Logic and Intelligent Systems 18, 316–322 (2008)

Magnetic Anomaly Points of Magnetic Field Achieved by Simulation and Experimental Validation in High-Precision Magnetic Test*

Hui Li, Chong Kang, Weiming Cheng, and Yue Zhang

College of Science, Harbin Engineering University, Harbin, China
kangchong@hrbeu.edu.cn, ann_lilylove@hotmail.com

Abstract. For high precision measurement of geomagnetic field, the induced magnetic field calculations for ferromagnetic objects have certain theoretical guidance meaning when degaussing. The influence of induced magnetic fields in ferromagnetic objects in a weak magnetic field was studied for this paper. Based on far magnetic field of magnetic dipoles, three magnetic dipoles are used to simulate the ferromagnetic objects; the formula of magnetic induction in space any point can be obtained. The influence of ferromagnetic objects on space magnetic field was approached theoretically as well as experimentally. Finally, calculation for a nail example showed that the established model is effective in practical engineering.

Keywords: High-precision magnetic survey, Magnetized magnetic field, Magnetic dipoles.

1 Introduction

The geomagnetic field is weak magnetic field; ferromagnetic substances produce an additional magnetic field around under the action, this magnetic field becomes interference field then affects next measurement. The change of the internal state of magnetic medium on the magnetic field is called magnetized, according to the magnetic characteristics of magnetized medium, magnetic medium can be divided into two categories: The first kind is called paramagnetic medium and diamagnetic medium, the magnetic characteristics of them are very weak, they belong to weak magnetic materials; The second type is ferromagnetism, it can produce strong magnetized field, they are strong magnetic material. Ferromagnetic substances are ferromagnetism, because of the effects of earth's magnetic field, produced by the ferromagnetic substances buried in the ground called magnetic anomaly [1]. In order to improve the magnetic protection of naval ships in practical application, many countries navy reduced the magnetic anomaly around navy ships by installing degaussing system, but the induced magnetic field of navy ships must be used to guide the design, installation and adjustment of degaussing system[2]. The analysis and calculation of the magnetic field of the ferromagnetic magnetic induction can also be used to guide elimination for the fixed magnetic field of ferromagnetic substances.

* Fund project: National Natural Science Foundation of China (61174192); The Central University Scientific Research Foundation Special Business Expenses (HEUCFL20101110).

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 199–206.
springerlink.com

This article establishes three dipole models of the nail detected in the magnetic field testing, and deduced the value of the magnetic induction in space any point, and use "Mathematic" to simulate, finally do verification through experiments.

2 The Magnetic Abnormal Points Analysis and Processing

In geomagnetic measurements, there are a lot of dramatic changes of magnetic field appear in geomagnetic isograms rendered by the software surfer (Fig.1), in order to avoid the influence of magnetized magnetic field on testing region, we found the geographical position of the dramatic changes of magnetic field, then two ferromagnetic substances (rusty nails) are dogged out, and carried out the measurement again, Fig.2 show the magnetic field distribution , On this basis it is found that the magnetic field distribution in the scope of testing region change by the anomaly points. The magnetic anomaly is caused by rusty nail.

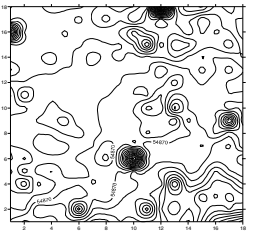


Fig. 1. The geomagnetic isograms with the area

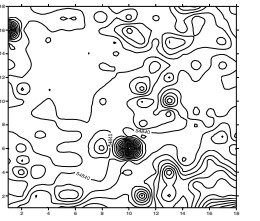


Fig. 2. The geomagnetic isograms after ruled out two abnormal point

3 Mathematical Model

When the volume of target magnets far smaller than the detection distance, this object can be used as magnetic dipole. In simple terms, magnetic dipole is a circular current.

As shown in Fig.3, we establish right-hand space right angle coordinate, the centre of the magnetic dipole is the origin, the direction of the magnetic vector is the positive direction of Z axis, and the direction of mutually perpendicular X axis and Y axis can be arbitrary definition. M is the arbitrary point in space, and its ball coordinates is $M(r,\phi_0,\theta_0)$, r is the distance from the centre to M, ϕ_0 is zenith angle, θ_0 is the angle between the tangential plane of M and the tangential plane of the X axis. Obviously, existing the relationship:

$$\begin{cases} x = r\sin\phi_0\cos\theta_0 \\ y = r\sin\phi_0\sin\theta_0 \\ z = r\cos\phi_0 \end{cases} \quad \text{Among them } 0 \le r < +\infty \quad 0 \le \phi_0 < \pi \quad 0 \le \theta_0 < 2\pi \tag{1}$$

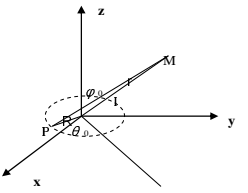


Fig. 3. Magnetic dipole and coordinate system

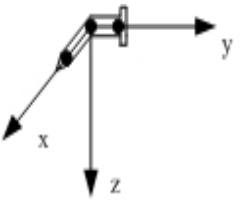


Fig. 4. Coordinate system of the nail

We establish coordinate for the nail approximately bended to right angle, consider the direction of the head of the nail as scale, called horizontal, and consider the direction of pointing to the head of the nail as the positive direction; consider the direction of the head of the nail as scale, called longitudinal, and consider the direction of pointing to the tail of the nail as the positive direction; the location of origin is the center of the right angle. We consider vertical plane as scale, called vertical direction, and consider the direction of down as positive direction. As shown in Fig.4, we establish the model for the nail using three magnetic dipoles; the three magnetic dipoles are respectively located at the origin, the X axis and the Y axis.

Classical mathematical expressions for magnetic dipole [3-4]

$$\vec{B}(\vec{m},\vec{r}) = \frac{\mu_0}{4\pi r^5}[3\vec{r}(\vec{r}\cdot\vec{m}) - r^2\vec{m}] \tag{2}$$

Type $\vec{B}$ is magnetic induction of magnetic dipoles , $\vec{r}$ is the distance from testing point to magnetic, $\vec{m}$ is magnetic moment, r is the scalar form of $\vec{r}$, μ_0 is vacuum magnetic permeability of magnetic field, size for $4\pi\times10^{-7}$ H/m, in the air, generally use the approximate value for the magnetic permeability.

Based on the type (2) magnetic dipoles model, deformation:

$$\vec{B}=\frac{\mu_0}{4\pi}\left[\frac{-\vec{m}}{r^3}+\frac{3(\vec{m}\vec{r})\vec{r}}{r^5}\right]=\frac{\mu_0}{4\pi}\left[\frac{-\vec{m}}{r^3}+\frac{3mr\cos\phi_0\vec{r}}{r^5}\right]=\frac{\mu_0}{4\pi}\left[\frac{-\vec{m}r^2+3mr\cos\phi_0\vec{r}}{r^5}\right] \tag{3}$$

A vector for the magnetic induction $\vec{B}=B_x\vec{i}+B_y\vec{j}+B_z\vec{k}$, then

$$\begin{cases} B_x = \dfrac{\mu_0}{4\pi}\left[\dfrac{3mr^2\cos\phi_0\sin\phi_0\cos\theta_0}{r^5}\right] \\ B_y = \dfrac{\mu_0}{4\pi}\left[\dfrac{3mr^2\cos\phi_0\sin\phi_0\sin\theta_0}{r^5}\right] \\ B_z = \dfrac{\mu_0}{4\pi}\left[\dfrac{3mr^2\cos^2\phi_0 - mr^2}{r^5}\right] \end{cases} \tag{4}$$

This type is three component expression of any point M (r,ϕ_0,θ_0) magnetic induction.

Clear type (4) up and translated into right angle coordinate system:

$$\begin{cases} B_x = \dfrac{3\mu_0 mxz}{4\pi r^5} \\ B_y = \dfrac{3\mu_0 myz}{4\pi r^5} \\ B_z = \dfrac{\mu_0 m(2z^2 - x^2 - y^2)}{4\pi r^5} \end{cases} \tag{5}$$

The size of magnetized magnetic field for the magnetic dipoles whose center is the origin in M point, we can get it from (5)

$$\begin{cases} B_{x1} = \dfrac{3\mu_0 mx \cdot z}{4\pi (\sqrt{x^2 + y^2 + z^2})^5} \\ B_{y1} = \dfrac{3\mu_0 my \cdot x}{4\pi (\sqrt{x^2 + y^2 + z^2})^5} \\ B_{z1} = \dfrac{m(2x^2 - z^2 - y^2)}{4\pi (\sqrt{x^2 + y^2 + z^2})^5} \end{cases} \tag{6}$$

Similarly, the center is located in $(x_0,0,0)$ place, magnetic field of magnetic dipole far field:

$$\begin{cases} B_{x2} = \dfrac{3\mu_0 m (x - x_0) \cdot z}{4\pi (\sqrt{(x - x_0)^2 + y^2 + z^2})^5} \\ B_{y2} = \dfrac{3\mu_0 m (x - x_0) \cdot y}{4\pi (\sqrt{(x - x_0)^2 + y^2 + z^2})^5} \\ B_{z2} = \dfrac{\mu_0 m[2(x - x_0)^2 - z^2 - y^2]}{4\pi (\sqrt{(x - x_0)^2 + y^2 + z^2})^5} \end{cases} \tag{7}$$

Center is located in $(0,\ y_0,0)$, magnetized magnetic field for magnetic dipole in the far field is:

$$\begin{cases} B_{x3} = \dfrac{3\mu_0 mx \cdot z}{4\pi (\sqrt{(x^2 + (y - y_0)^2 + z^2})^5} \\ B_{y3} = \dfrac{3\mu_0 mx(y - y_0)}{4\pi (\sqrt{(x^2 + (y - y_0)^2 + z^2})^5} \\ B_{z3} = \dfrac{\mu_0 m[2x^2 - z^2 - (y - y_0)^2]}{4\pi (\sqrt{(y - y_0)^2 + x^2 + z^2})^5} \end{cases} \tag{8}$$

The total magnetized magnetic field intensity for the point M :

$$\begin{cases} B_x = B_{x1} + B_{x2} + B_{x3} = \dfrac{3\mu_0 mxz}{4\pi(\sqrt{x^2+y^2+z^2})^5} + \dfrac{3\mu_0 m(x-x_0)z}{4\pi(\sqrt{(x-x_0)^2+y^2+z^2})^5} + \dfrac{3\mu_0 mxz}{4\pi(\sqrt{x^2+(y-y_0)^2+z^2})^5} \\ B_y = B_{y1} + B_{y2} + B_{y3} = \dfrac{3\mu_0 mxy}{4\pi(\sqrt{x^2+y^2+z^2})^5} + \dfrac{3\mu_0 m(x-x_0)y}{4\pi(\sqrt{(x-x_0)^2+y^2+z^2})^5} + \dfrac{3\mu_0 mx(y-y_0)}{4\pi(\sqrt{x^2+(y-y_0)^2+z^2})^5} \\ B_z = B_{z1} + B_{z2} + B_{z3} = \dfrac{m(2x^2-z^2-y^2)}{4\pi(\sqrt{x^2+y^2+z^2})^5} + \dfrac{\mu_0 m[2(x-x_0)^2-z^2-y^2]}{4\pi(\sqrt{(x-x_0)^2+y^2+z^2})^5} + \dfrac{\mu_0 m[2x^2-z^2-(y-y_0)^2]}{4\pi(\sqrt{x^2+(y-y_0)^2+z^2})^5} \end{cases} \quad (9)$$

4 The Simulation

We set the depth z of model to buried depth, and applied software "Mathematica" to simulate by the figure we had got of the burial nail, through the simulation curve we can see equivalence value maps with the changing trends of the simulation parameters, had done a lot of experiments. The simulation curve as follows:

Setting the parameter to the value as follows: $z = 0.5; a = 0.3; b = 0.4; c = 0.6; \mu = 1; y_0 = 0.7; x_0 = 0.8$

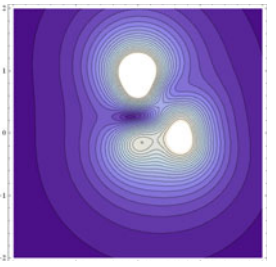


Fig. 5. When z=0.5, the curve that simulates the magnetic field intensity of the nail

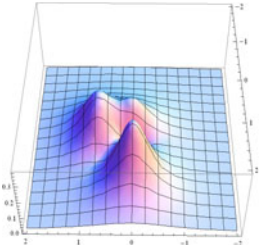


Fig. 6. When z=0.5,the 3d map that simulates the magnetic field changes of the nail

In order to know identical degree of magnetic field isograms and the model of the magnetic dipoles in the different depths, we simulated different buried depth z as following:

Setting the parameter to the value as follows: $z = 1; a = 0.3; b = 0.4; c = 0.6; \mu = 1; y_0 = 0.7; x_0 = 0.8$

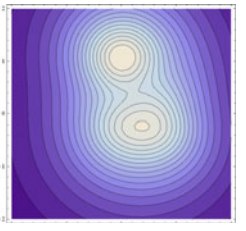


Fig. 7. When z=1,the curve that simulates the magnetic field intensity of the nail

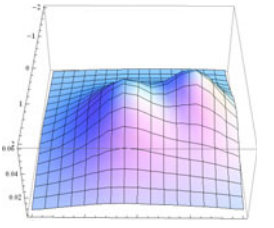


Fig. 8. When z=1,the 3d map that simulates the magnetic field changes of the nail

5 An Example of the Nail

5.1 The Experimental Program

In order to avoid the effect of iron ore, gravel, we choose a square sand area, four sides respectively paralleled to East, South, West and North. As shown in figure 9 shows, we fixed rope 2, 3 along the north and south direction. The rope1 could move, after measuring one column, moved rope 1 to the next column to continue to measure. Then we used proton magnetometer to measure every measurement sites. The distance between two points was 5cm. Date-treatment was carried out on the site scope 50 cm × 50 cm.

After measuring geomagnetic field without the influence of nails, we carried out anther measurement with a nail buried in the ground. In order to increase the credibility of the experiment, we had done experimental validations for different buried depth. The scheme for the experiments was shown as Fig.9: the nail was placed in the center of square area with the buried depth of 5 cm, 10 cm, the method of measurements was the same as above.

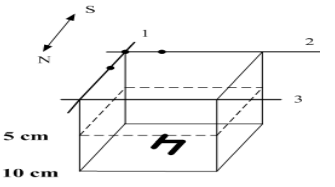


Fig. 9. The geomagnetic field measurement schematic diagram of the nail

In the experiment, the effect of space magnetic field on the nail was the difference between abnormal field (with nails) and normal field (no nails).We apply software "surfer" to obtain isograms, the solid lines represent positive, the dotted lines represent negative, and the thick lines represent zero, the isograms were as follows(interval 5 nT):

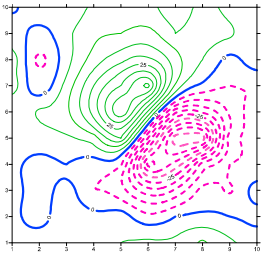


Fig. 10. The isograms underground 5 cm

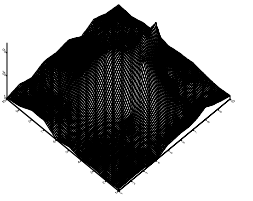


Fig. 11. The 3 d structure map

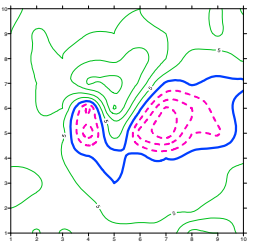


Fig. 12. The isograms underground 10 cm

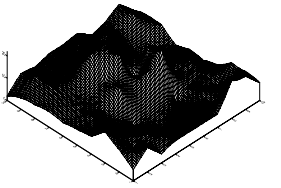


Fig. 13. The 3 d structure map

In the experiment, the geomagnetic field around the nail had significant change in two buried depths, which manifested as closed isograms. The dramatic changes of the geomagnetic field corresponded to the position of nail. With the buried depth increasing, closed isograms become thinner, which showed that the magnetic induction and detection strength become weaker.

In the experiment, the magnetic induction had been changed larger than before. We introduce absolute value into 3d structure map to clearly display the change range.

5.2 Results Analysis

Based on the calculation and experiment of the geomagnetic field, the error sources were shown as follows:

1) The error of magnetic susceptibility treatment that ferromagnetic substances were magnetized by external magnetic field. Theoretically when ferromagnetic substances are magnetized, even if additional magnetic field changes are small, susceptibility still has a small change. In order to solve the actual problem that susceptibility is difficult to determine, in the simulation of the model, we think susceptibility in weak magnetic field is almost invariable.
2) The error of measuring equipment used in the experiment (proton magnetometer) The error include the magnetic sensor itself, the placed attitude of the probe, the manual operation during the practical measurement, the precision of measurement grid and geomagnetic environment(The effect of sand particles, the rock, the iron ore in measuring field, and the effect of the ferromagnetic materials around the measuring field) , etc
3) The error of representing ferromagnetic substance by three magnetic dipoles superposition in the model. Magnetic dipole is the basic unit on magnetic theory research, the magnetic phenomenon in nature all can be equivalent to the superposition of the magnetic field of the certain magnetic dipole. In theory, the finite numbers of magnetic dipoles limit the accuracy of the model.

6 Conclusions

We obtain the simulation results by establishing the model of the magnetic body using three magnetic dipoles according to the shape of the ferromagnetic object, although three magnetic dipoles model is not consistent with the theoretical model, the analysis calculation of magnetic induction of the ferromagnetic material can guide the elimination of the permanent magnetic field of the ferromagnetic material. he distribution form and the intensity of the magnetic field anomaly is related to the location of field sources, magnetism content, the shape of own, the distance from the measuring point. With the distance from space point to the ferromagnetic object increasing, the effect of the local magnetization on the magnetic field of the ferromagnetic material becomes small; as a result, the magnetic field distribution is close to distribution of the magnetic body which can seem as uniformly magnetized body. The permanent magnetic field of the ship and the carrier Package used to load cables can be calculated by using this far-field magnetic dipoles model; it can guide the elimination of magnetic disturbances.

References

1. Sun, Z., Chen, T.: The application of magnetic gradient detecfion in engineering pile length testing. Shanghai Geology 25(3), 48–62 (2007)
2. Zhou, Y., Zhang, G.Y.: Ship magnetic fields analysis, pp. 224–225. National defense work Industry press, Beijing (2004)
3. Schlageter, V., Besse, P.A., Popv, I.C.R.S., et al.: Tracking system with five degrees of freedom using a 2D-array of hall sensors and a permanent magnetic. Sensors and Actuators A 92, 37–42 (2001)
4. Hou, W.S., Zheng, X.L., Peng, C.L.: Study of simplified magnetic-location model for micro medical device inside human body. Chinese Journal of Scientific Instrument 26, 895–897, 901 (2005)

A Cluster-Based QoS Information Approximation Algorithm in Multi-domain Networks

Yue Han[1,2] and YongJian Luo[2]

[1] ISN State Key Lab, Xidian University, Xi'an, China
[2] Xi'an Institute of Communications, Xi'an, China

Abstract. This paper proposes a method of improving the scalability of topology aggregation (TA) as network size grows. A staircase representation algorithm is employed to approximate the QoS information representation. The new staircase is represented by fewer points whose size will not grow with the network size by applying cluster analysis to the arrival requests and the area-minimization optimization is used to minimize the approximation distortion for determining these points. Simulation results show that the proposed method achieves better scalability compared with existing TA methods.

Keywords: topology aggregation, staircase, scalability, cluster analysis.

1 Introduction

Topology aggregation is generally used in QoS routing to deal with the scalability problem [1]. It is defined as a process of summarizing the internal topology information of a domain to reduce the routing overhead. Among the existing TA methods, geometry-based methods have turned out to be very effective for TA in the delay-bandwidth sensitive networks [2-6]. These methods use a geometric representation to approximate a staircase represented by some points that corresponds to the delay/bandwidth requirements. There are a few works on the approximation. In [2] a stretch-factor based method is introduced to contract the overestimated region. It is effective but still has higher crank-back ratio. A line segment method proposed in [3] is better than the stretch-factor method in average. Unfortunately, it still introduces obvious distortions as some points are far away from the line segment. The works in [4] define three approximation curves (polynomial curve, cubic spline, and polyline) which make the distortion less than the line-segment method makes, however, their space complexity is dramatically increased. So far, the staircase approximation has been the research focus of TA.

The main idea of the geometry-based methods is to minimize the area between the staircase and the approximating geometric representation. However, with the continuing growth of the network size, the area will increase dramatically, leading to what is called scalability problem. Due to the scalability problem, although the current staircase approximation methods may work well at a specified network size, these methods will fall into trouble as the network size grows. In addition, this scalability challenge is further complicated when networks are delineated into sizeable scale and hence the scalability problem of the staircase approximation has

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 207–211.
springerlink.com

been gaining importance. In order to improve the scalability, the work in [5] uses a new staircase represented by H predefined points to approximate the original staircase. However, with the rapid growth of network size, more points are necessary to represent the staircase for reducing the approximation distortion, so the time complexity becomes higher accordingly.

In light of the above analysis, we focus on the staircase representation method. In this paper, we take the user requests into account to determine the number of points that represent the staircase. The user requests are divided into G clusters by means of cluster analysis. As described in DiffServ researches[7] and [8], G is relatively stable and does not increase with the expansion of network scale and thus G is independent of network scale. According to the G clusters, a novel staircase representation algorithm is proposed by using the area-minimization optimization to minimize the approximation distortion. The simulation results show that our method outperforms existing TA methods in terms of the scalability.

2 Proposed Solution

Our proposed staircase representation method consists of two processes. The first process is to partition the user requests into G clusters by using cluster analysis technique. The second process is to form the new staircase according to the G clusters.

2.1 Cluster Analysis of the User Requests

As observed in existing networks and stated in DiffServ[7] and [8], although the number of user requests is large and vary greatly with the time variation, there are only a limited number of classes corresponding to the huge amount of user requests. Thus we can partition the user requests into different classes. Since the number of points that represent the staircase is determined with the granularity of resulted classes rather than the user requests. Therefore our proposed method is more scalable.

Similar to the network information expressed on the delay-bandwidth plane, the user requests can also be expressed by some points on the delay-bandwidth plane. Given that the distribution of these points corresponds to the characteristic of compact clusters, we use the distance between two points as the clustering metric. Next we concentrate on aggregating the points attached to each class into different clusters that conform to the underlying classes.

Any feasible and suitable clustering methods can be adopted to obtain the clusters. We further propose a rule to estimate the clustering quality. The rule include two aspects, first, the method should cluster as near points as possible; second, the points within a cluster should be closer than the points that does not belong to the same cluster. To implement the rule, we design it as follows. For different points p_i and p_j we define the distance between p_i and p_j as $dist(p_i, p_j)$. Given a cluster C_i, the mean and variance of the distances among the points within the cluster is denoted as $dist_{mean}(C_i)$ and $dist_{var}(C_i)$, respectively. For different clusters C_i and C_j, we define the distance between C_i and C_j as $dist(C_i, C_j)$

$$dist(C_i, C_j) = \min_{p_i \in C_i, p_j \in C_j} dist(p_i, p_j) \quad (1)$$

Then the mean of the distances among the clusters is denoted as $dist_{mean}$. Thus we give a new factor η, defined by

$$\eta = \frac{dist_{mean}}{\max_i(dist_{mean}(C_i))} \quad (2)$$

The optimal clustering method should be the one which achieves the maximum η while achieves the minimum $dist_{var}(C_i)$.

2.2 Staircase Representation Algorithm

Given a resulted cluster c_i, the points fallen within c_i are included in a rectangle region $[d_{i,min}, d_{i,max}] \times [b_{i,min}, b_{i,max}]$ which is denoted by s_i. We assume that there are G clusters. As shown in Fig.1, the G rectangles are mapped to the original staircase to divide the representative points into G groups. Assume that there are n_i representative points fallen within s_i, these n_i representative points are used to determine an optimal point of s_i as shown below.

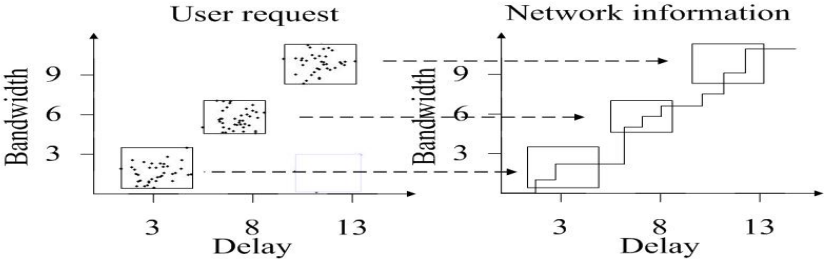


Fig. 1. Mapping user requests to the original staircase

Assume the representative points are ranked in order of delay from small to large. Given a rectangle $S_i = [d_{i,min}, d_{i,max}] \times [b_{i,min}, b_{i,max}]$, $\{p_i^1, \cdots, p_i^{n_i}\}$ are n_i repetitive points fallen within s_i. Then an optimal point can be determined in the following steps:

Step 1. Compute the area enclosed by the original staircase and s_i as illustrated in Fig 2. Denote this area as A. A defines the region of supported user requests.

Step 2. Make a curve f_1^i within s_i which satisfies:

$$f_1^i = \{(x, y) \mid (d_{i,max} - x)(y - b_{i,min}) = A, (x, y) \in S_i\} \quad (3)$$

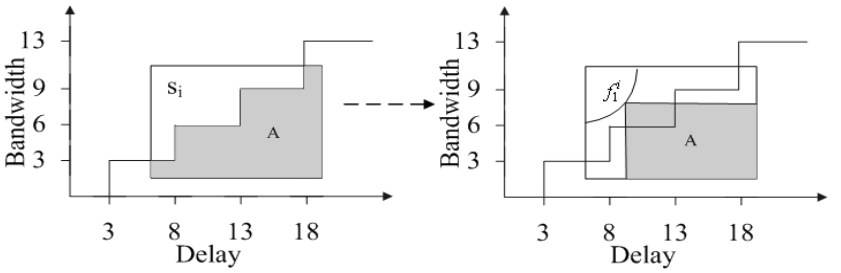


Fig. 2. Illustration for determining optimal point of s_i

Step 3. Find the last representative point before p_i^1 denoted as p_{i-1}^1 and the next representative point of $p_i^{n_i}$ denoted as $p_{i+1}^{n_i}$. Then the ordinate of p_{i-1}^1 is y_{i-1}^1 and the abscissa of $p_{i+1}^{n_i}$ is $x_{i+1}^{n_i}$. Given any point *(x,y)* on f_1^i, we can make a simple staircase f_2^i using p_{i-1}^1, *(x,y)* and $p_{i+1}^{n_i}$. As illustrated in Fig.3, f_2^i depends on *(x,y)* and consists of four segments, which is written as:

$$f_2^i(x', y') = \begin{cases} y' = y_{i-1}^1 & x' \in (d_{i,min}, x) \\ x' = x & y' \in (y_{i-1}^1, y) \\ y' = y & x' \in (x, x_{i+1}^{n_i}) \\ x' = x_{i+1}^{n_i} & y' \in (y, b_{i,max}) \end{cases} \tag{4}$$

Step 4. Let f_3^i denotes the original staircase within s_i. The area between f_2^i and f_3^i is

$$\Delta = \int_{d_{i,min}}^{d_{i,\max}} \left| f_2^i - f_3^i \right| dx \tag{5}$$

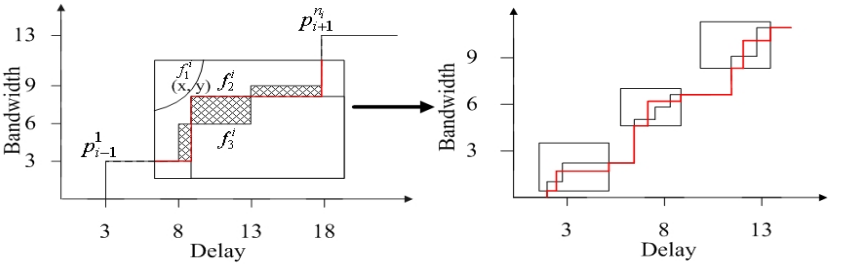


Fig. 3. Illustration for representing a new staircase

For all the points *(x,y)* on f_1^i, there is a point yielding the minimum Δ. This area minimization optimization can be formulated as a linear programming problem:

$$\begin{aligned} &\min \ \Delta \\ &s.t. \ (d_{i,max} - x)(y - b_{i,min}) = A, (x, y) \in S_i \end{aligned} \tag{6}$$

Since Δ is continuously differentiable, we can obtain the optimal point of s_i by calculating the derivative of Δ.

For G clusters of arrived user requests, we can find G optimal points. As illustrated in Fig.3, a new staircase with these G optimal points is formed which can also be used by the existing TA methods to achieve better information aggregation. Since we need to determine G optimal points, the time complexity of the staircase representation algorithm is $O(G)$.

3 Conclusion

This paper proposes a scalable staircase representation method for TA which maintains better scalability and lower approximation distortion. The method applies cluster analysis to the user requests to determine the number of points that represent the staircase so that it is scalable with the network size. Simulation results show that the proposed method always approximates the original staircase with much greater accuracy and hence achieves better scalability comparing with existing TA methods.

Acknowledgments. This work is supported by National Nature Science Foundation of China (NSFC) 61001129, 61179002.

References

1. Uludag, S., Lui, K.S.: Analysis of Topology Aggregation techniques for QoS routing. ACM Computing Surveys 39(3), Article 7 (August 2007)
2. Korkmaz, T., Krunz, M.: Source-oriented topology aggregation with multiple QoS parameters in hierarchical networks. ACM Trans. Modeling Computer Simulation 10(4), 295–325 (2000)
3. Lui, K.-S., Nahrstedt, K.: Routing with topology aggregation in delay-bandwidth sensitive networks. IEEE/ACM Trans. Networking 12(1), 17–29 (2004)
4. Tang, Y., Chen, S., Ling, Y.: State aggregation of large network domains. Computer Communications 30(4), 873–885 (2007)
5. Hou, R., Lui, K.-S., Leung, K.-C.: Routing with QoS information aggregation in hierarchical networks. In: 17th International Workshop on Quality of Service, pp. 1–9 (2009)
6. Zarifzadeh, S., Nayyeri, A., Yazdani, N.: Consumer oriented state aggregation using reinforcement learning approach. In: 3rd IEEE Consumer Communications and Networking Conference (CCNC), pp. 524–530 (2006)
7. Nichols, K., Blake, S., Baker, F., Black, D.: Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers, RFC 2474 (December 1998)
8. Blake, S., Black, D., Carlson, M.: An Architecture for Differentiated Services, RFC 2475 (December 1998)

Standby Power Control System Based on User's Location for Energy Saving in Smart Home

Kyoung-Mi Im[1], Chi-Su Kim[2], and Jae-Hyun Lim[2]

[1] Green Home Energy Technology Research Center, Kongju National University
[2] Dept. of Computer Science & Engineering, Kongju National University,
275, Budaedong, Cheonan, Chungnam, Korea
{omnibus,cskim,defacto}@kongju.ac.kr

Abstract. This paper designs a system to control the standby power of household appliances used at home. To achieve this, the system detects user's location by sensors in an area which is divided into several sections and controls standby power on or off. The current system controlling standby power is cut off the power automatically if electrical power below a certain amount of power is detected and leads to inconvenience caused by controlling the standby power manually when user is resupplied power from system. Therefore, this paper proposes a standby power control system based on location of user which controls standby power of home appliances in the section to meet both energy saving of electrical power and convenience. For this, the indoor space is separated into a number of zones and the system detects whether user is in the section or not by PIR sensor. To verify the system, we compare power between current system and proposed system through experimental scenario.

Keywords: Standby Power, Context-Awareness, Energy Saving System, Smart Home.

1 Introduction

Recently serious climate change has occurred across the world and countries are tightening the regulation on greenhouse gas emissions. In response, a lot of effort is being put in to reduce energy expenditure and research on reducing electrical energy expenditure, which accounts for approximately 20% of all energy expenditures, is being actively undertaken in Korea. As the electrical energy expenditure due to use of home or office appliances increases, interest is increasing not only electricity consumed when appliances are used, but also electricity of the standby power which is used while waiting for an external activation signal[1][2]. IEA(International Energy Agency) estimates that OECD member nations use standby power of 10% or 60W per family[3][4]. In addition, as various household appliances appear, a number of home appliance more than 1.5 per year is increasing and the rates of use of high-tech appliances through networking is also increasing. For this reason, they expect to increase electricity of standby power about 5.8% per year for 20 years from now. According to IEA in Swiss, if this trend continues, the electricity of standby power consumed in people's homes predicts that it will grow to 25% of total electrical consumption around 2020. But most people using home appliances have not realized

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 213–220.
springerlink.com

how saving electricity of standby power is very important. Therefore, we should develop a system that can reduce it.

The current methods of controlling standby power include the manual method of the user unplugging the appliance that is not in use or using standby power consent or multiple-tap to automatically stop the electricity and upon deactivation the resupplying of electricity according to the user's movements and more. But the two methods both have the disadvantage of requiring the user's voluntary participation and calls for a development of an intelligent standby power control system with the user's convenience in mind[5].

This thesis proposes a standby power control system that does not have the disadvantage of requiring the user's manual control and intelligently controls the standby power according to the location of the user to reduce the energy wasted in standby power. For this the indoor space is separated into a number of Zones and the user's location is found using sensors and the standby power of the corresponding zone is stopped or released. Here the system's context-awareness and control uses ZigBee technology based on IEEE 802.15.4[6-9].

2 Related Work

Standby power is electrical energy spent while waiting for internal and external activation signal or when the electronic equipment is not used when it is connected with the external power supply, and the standby power control systems currently used are using control consent. This consists of 4 different modes such as automatic, manual, timer movement and constant power supply, and in the case of automatic mode the power supply is automatically stopped when the amount of electrical power that is less than a certain value is maintained for a certain amount of time[10]. In manual mode On or Off is controlled by the user's forced control; in timer mode the power is cut or supplied at certain times depending on the user's time schedule setting and in constant power the standby power is not controlled.

Table 1. Standby power of appliances

Device	Standby Power	Device	Standby Power
DVD player	12.2W	Computer	3.26W
Audio system	8.61W	Printer	3.07W
Set-top box	7.85W	Microwave	2.77W
External modem	6.43W	Monitor	2.63W
Video	5.45W	Cordless phone	2.15W
TV	4.33W	Washing machine	1.90W
Bidet	3.39W	Charger of cellular phone	1.72W

Table 1 shows the standby power value of each household equipment and in the case of digital or office equipment that are connected to a network, it can be seen that there is a lot of standby power being wasted.

One of the currently used automatic standby power control systems in smart homes connect each appliance to PLC (Power Line Communication) as in Fig. 1 and stops or releases the standby power of the whole household depending on the user's presence in the house and is not able to control the standby power according to individual zones.

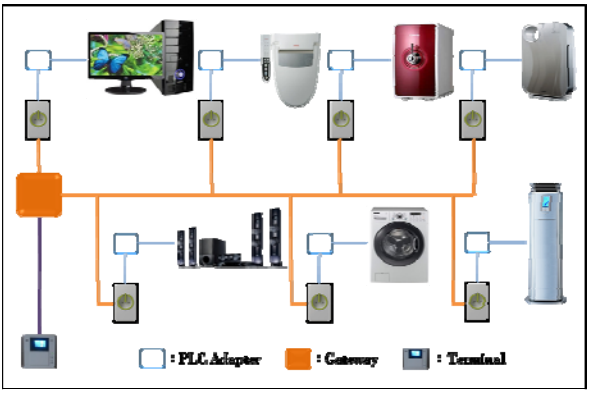


Fig. 1. Standby power control system in Smart Home

3 Standby Power Control System Based on User's Location

The Standby Power Control System based on User's Location suggested by this thesis provides a zone based service and has a system structure as shown in Fig. 2. The system is sorted into the user's location search step within the user's Zone and

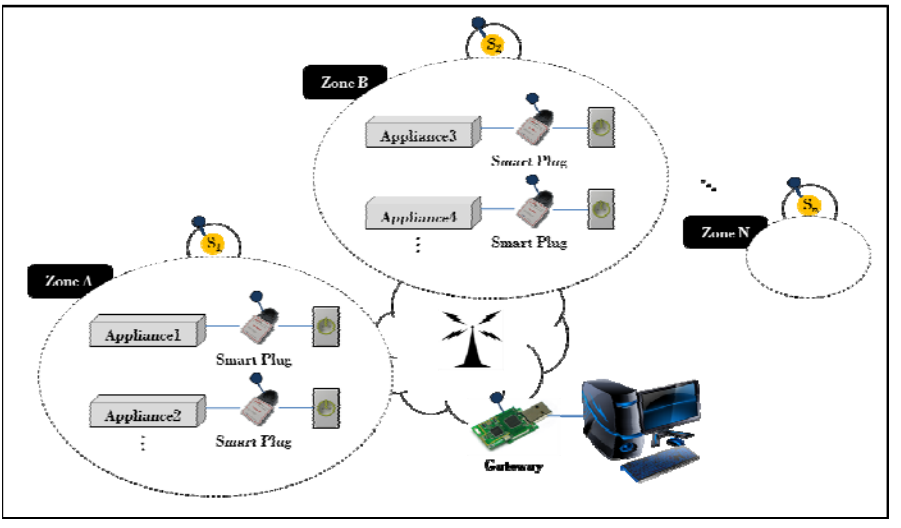


Fig. 2. System structure of proposed standby power control system

standby power control step which stops or releases the standby power whether the user is in the room or not, and the result of each step is sent to the system using ZigBee technology based on IEEE 802.15.4.

3.1 User Location Search

To detect the user is in the room or not, this thesis placed PIR sensors at the entrances according to each zone. To identify the entering and exit of the user, the PIR sensor used sticking on 2 things as seen in Fig. 3 and has sent out each distinguishable ID. If the user passes the first PIR sensor then the second PIR sensor the user is perceived to have entered the corresponding zone. Conversely, if the user passes through the second PIR sensor then the first PIR sensor, the user will be perceived to have exited the corresponding zone. The result on whether or not the user has reentered is sent to Gateway and transferred to the system for each zone.

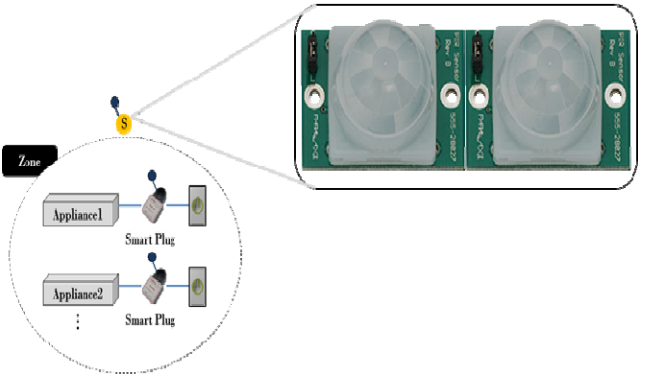


Fig. 3. System structure of proposed standby power control system

3.2 Control of Standby Power

Smart Plug was installed on the consent in order to measure the energy spent by each household appliance and to control the standby power. The Smart Plug is run on ZigBee communications of 2.4GHz and sends the value of energy spent in a certain period of time by each appliance to Gateway. Also, if there is a stop or release command from the system Actuator is generated and the standby power is controlled. However, for an appliance that power is switched will be supplied regardless of whether the user is in the room or not and power control will only take place for standby mode of less than a certain electrical power.

If the user's location is detected in a random zone A the system distinguishes household appliances located in the corresponding zone and re-supplies the stopped electricity and if the user's location is moved from zone A to zone B. Zone B's power is supplied and zone A's is cut.

4 Experiment and Evaluation

4.1 Control of Standby Power

For the experiment of the system the space in the hypothetical home was separated into 6 areas like in Fig. 4 and PIR sensors were installed in each entrance. The Living Room was designed to detect the reentering user's location from the vestibule, unlike other areas. If the reentrance of S4 is detected this is recognized as the return of the reentered user. Moreover the exit from other areas is deemed to be the movement to the Living Room and the exit from the Living Room is deemed to be a departure to the outside.

Fig. 4. Experimental environment

Table 2 is a chart showing the list of household appliances installed per zone and ID required to distinguish Smart Plug and each appliance's hourly standby power.

4.2 Energy Saving Evaluation through Scenarios

The scenario supposes that two users spend 12 hours a day from 9:00PM to 9:00AM at home and moves as shown in Fig. 5. The control of standby power starts with the detection User A's return to the Living Room and the system carries out the stopping and the supplying of electricity according to the user's entering and exiting of the zone.

If the system is a constant power system that does not have its standby power controlled and standby power is spent by all the appliances in Table 2 60.2W of energy is spent per hour and if thought of in terms of 24 hours approximately 1.44Kw of energy is wasted.

Table 2. Used appliances

Zone(Sensor)	Appliances	ID	Standby Power
Room1(S_1)	Desk lamp	1	3.0W
Room1(S_1)	TV	2	4.3W
Room2(S_2)	Desktop	3	3.3W
Room2(S_2)	Monitor	4	2.6W
Room2(S_2)	Printer	5	3.0W
Room3 (S_3)	Audio system	6	8.6W
Living Room(S_4)	TV	7	4.3W
Living Room(S_4)	Air cleaner	8	0.7W
Living Room(S_4)	Air conditioner	9	10.0W
Living Room(S_4)	DVD	10	12.2W
Kitchen(S_5)	Washing machine	11	2.0W
Kitchen(S_5)	Microwave	12	2.8W
Bathroom(S_6)	Bidet	13	3.4W

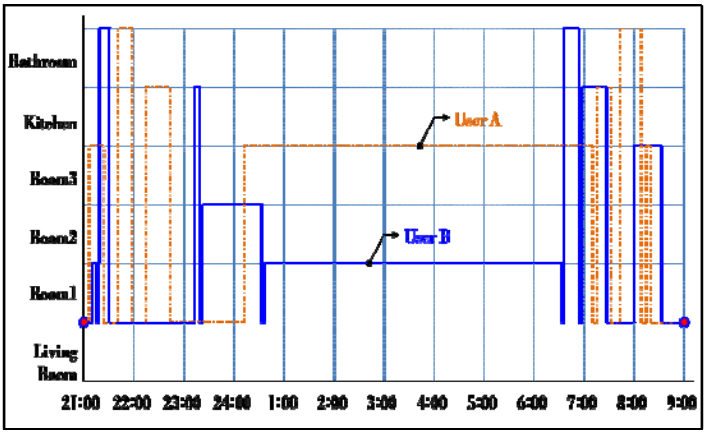


Fig. 5. Scenarios for locations

Fig. 6 shows the amount of time 2 users used each of the appliances with the x axis as time and y axis as the appliance's distinguishable ID. If the total standby power spent per zone according to the movement scenario of users A and B is shown it is as Table 3. Of the 2 standby power shown Standby Power 1 is excluding the power expenditure due to the scenario and Standby Power 2 is a comparison of the standby power spend due to the suggested system.

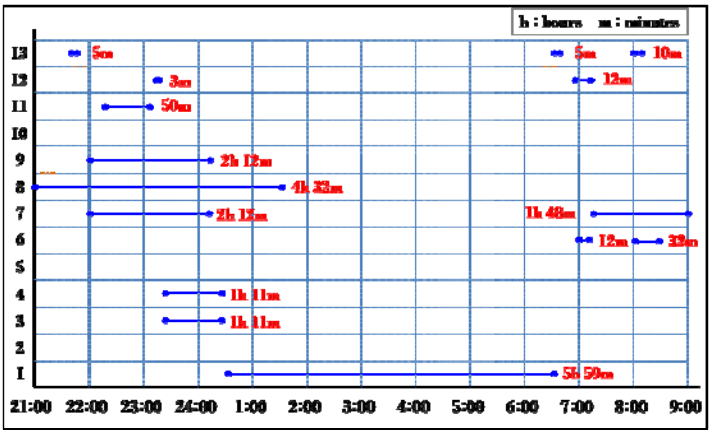


Fig. 6. Scenarios for appliances

Table 3. Standby power per zone

Zone	Standby Power 1	Standby Power 2
Room1(S_1)	69.65W	26.34W
Room2(S_2)	99.82W	3.55W
Room3(S_3)	96.75W	60.77W
Living Room(S_4)	284.73W	86.96W
Kitchen(S_5)	55.23W	4.16W
Bathroom(S_6)	39.67W	0.9W
Total	645.85W	182.68W

We were able to confirm that the suggested system's standby power expenditure was approximately 72% less than the Standby Power 1 previously used.

5 Conclusions

Currently operating standby power control systems consist of a structure that needs the user's manual control or control by a management system, or have a structure that uses occupancy sensor to detect movement to generally control the standby power. For manual control, it is a disadvantage of requiring the users to participate in the standby power that is allowing energy to be wasted, to be controlled. Also in the case of batch control using sensors there is a problem of the standby power of appliances in unused areas still being wasted.

This thesis proposes a Standby Power Control System based on User's Location for Energy Saving in Smart Home and separates the area within a home into a few zones and supplies electricity to the zone that the user is in and cutting off the standby power in other areas. The proposed system is a convenient system that does not

require the user's manual control and the reduction in energy spent was proved through experiment.

This thesis supplied to all the appliances in the current zone regardless of the appliances that were being used by the user. In the future, research into a standby power control system that enables control for each appliance within the zone using each user's individual pattern of appliance usage should be continuously developed.

Acknowledgments. This work was supported by Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0022977). This work was supported by the Human Resources Development of the Korea Institute of Energy Technology Evaluation and Planning(No. 20114010203040) grant funded by the Korea government Ministry of Knowledge Economy.

References

1. Ye, Z., Ji, Y., Yang, S.: Home Automation Network Supporting Plug-and Play. IEEE Transactions on Consumer Electronics 50(1), 173–179 (2004)
2. Calhoun, B.H., Chandrakasan, A.P.: Standby Power Reduction Using Dynamic Voltage Scaling and canary Flip-Flop Structures. IEEE Journal of Solid-State Circuits 39(9), 1504–1511 (2004)
3. Ross, J.P., Meier, A.: Whole-House Measurements of Standby Power Consumption. In: Proceedings of the Second International Conference on Energy Efficiency in Household Energy Economics, LBNL-45967, Rome (2000)
4. Heo, J., Hong, S.S., Kang, S.B., Jeon, S.S.: Design and Implementation of Control Mechanism for Standby Power Reduction. IEEE Transactions on Consumer Electronics 54(1), 179–185 (2008)
5. Kang, M.-S., Kim, S.-W., Kang, M.-S.: Design of Smart Home System through Isolation and Control of Standby Power. In: IEEK(The Institute of Electronics Engineering of Korea) Conference, vol. 33(1), pp. 1523–1526 (2010)
6. ZigBee Alliance, ZigBee Specification. version1.1(2006)
7. Zigbee Alliance, Smart Energy Profile Specification, version1.0(2009)
8. Park, W.-K., Choi, C.-S., Han, J., Han, I.: Design and Implementation of ZigBee based URC Applicable to Legacy Home Appliances. In: ISCE 2007 (IEEE International Symposium), pp. 1–6 (2007)
9. Han, I., Park, H.-S., Jeong, Y.-K., Park, K.-R.: An Integrated Home Server for Communication, Broadcast Reception, and Home Automation. IEEE Transactions on Consumer Electronics 52(1), 104–109 (2006)
10. Kim, S., Kim, D., Kwak, N., Oh, Y.: SPCS: Standby Power Control System to save energy. In: Proc. of KIISE, pp. 375–379 (2010)

Neural Network and Artificial Intelligence Study in Psychiatric Intelligent Diagnosis*

Bing Mei Chen[1,2,**], Xiao Ping Fan[1], Zhi Ming Zhou[3], and Xue Rong Li[2]

[1] College of Information Science and Engineering,
Central South University, Changsha, China
[2] The Second Xiangya Hospital, Central South University,
Changsha, Hunan, China, 410011
[3] Changsha Environmental Protection College,
Changsha, Hunan, China, 410004
cbm8882010@hotmail.com

Abstract. We have applied intelligent control theory into medical psychiatric diagnosis. During studying, we find selecting suitable neural network structure is very important to BP network. Suitable neural network structure will bring less error to diagnosis system. It is key to success or failure of psychiatric intelligent diagnosis system. We have found some rules that suit our diagnosis system. Such as : full connecting mode is better and at the same time adding suitable hidden node number can improve convergence effect and reduce error of network. But adding hidden layer number doesn't always improve network convergence effect under our studying. At the same time it makes network convergence speed to become slower and makes network training time to increase. We build an intelligent diagnosis system. Comparing the diagnosis by computer with the senior child psychiatrists, the consistent rate of intelligent diagnosis is 99%.

Keywords: Psychosis Diagnosis, Artificial Intelligence, Neural Network, Network Connecting Mode, Hidden Layer Node Number, Hidden Layer Number.

1 Introduction

Artificial neural network is a kind of method that can simulate human's brain visual thought and non-linearity parallel distributing treatment ability. It is usually used to simulate human's brain neural system structure and function.

Many large scope epidemiological investigations in China and abroad have found that child mental health disorders prevalent rate is about 10 percent. According to the prevalent rate, there are about 60 millions children in our country who need mental health care. But the child psychiatrists in China are not enough to serve so many kids.

* Foundation item: Project (39270262) supported by the National Natural Science Foundation of China, The research wins the third prize of Hunan province medical science and technology progress.
** Corresponding author.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 221–228.
springerlink.com

We hope that the abundant clinical experiences of senior child psychiatrists can help the younger doctors throughout the country by the advanced computer technology.

Our thought gain recognizing and granting of National Natural Science Foundation of China. So we get grant for development research. We have applied intelligence control theory into medical psychiatric diagnosis. It is also a very interesting practice that uses the computer simulating the human being brain.

2 Back Propagation Neural Networks

Man's neural system consists of 10^{10}~10^{11} neural cells. Brain neural cell is shown as Fig. 1.

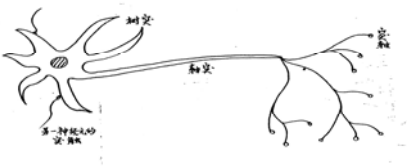


Fig. 1. Brain neural cell model

Artificial neural network processing element is a kind of simulation to man's brain neural cell. Artificial neural network processing element model is shown as Fig. 2.

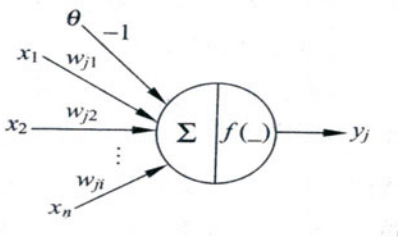


Fig. 2. Artificial neural network processing element model

Artificial neural network processing element consists of many inputs xi, i=1,2,…,n and an output yj. As be shown follows:

$$y_j(t) = f(\sum_{i=1}^{n} w_{ji} x_i - \theta_j) \tag{1}$$

θ_j is threshold of neural network cell, w_{ji} is connection weight of network cell j, n is input signal number, y_j is output of neural network cell, t is time, f (.) is output transform function.

Weight adjustment between PEs in back propagation is carried out according to the difference between the expectation target value dj and the practical output value yj of the neural network. In back propagation, the difference of the error is measured by the mean square error, as shown below:

$$E_k = \frac{1}{2}\sum_{j=1}^{q}(d_j - y_j)^2 \tag{2}$$

$$E = \sum_{k=1}^{m} E_k = \frac{1}{2}\sum_{k=1}^{m}\sum_{j=1}^{q}(d_j - y_j)^2 \tag{3}$$

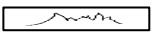
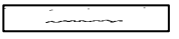
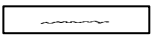
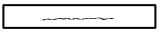
3 Effective Neural Network Structure Rules in Intelligent Diagnosis System

During studying neural network structure we has found some rules in psychiatric intelligent diagnosis system.

3.1 Rule 1: Adding Suitable Hidden Node Number Can Improve Convergence Effect and Reduce Error of Network

We can prove it from the second experiment. From table 1 we find that adding suitable hidden node number can improve convergence effect and reduce error of network. As below network1 is a 3 layer neural network with 21 input nodes, 1 hidden node and 1 output node. Network2 is a 3 layer neural network with 21 input nodes, 10 hidden nodes and 1 output node. Network2 only has different hidden layer node number from Network1. Each network learns the same samples .Their train count and recall error are shown in Table 1.

Table 1. The Comparison of Change Hidden Layer Node Number

	Network 1	Network 2
input node number	21	21
hidden1 node number	1	10
output node number	1	1
algorithm	BP	BP
transfer	S shape function	S shape function
network connecting mode	full connecting	full connecting
training count	2000	2000
Out-err shown in instrument		
training count again	10000	10000
Out-err shown in instrument		
recall result range	0.992769~0.993175	0.998003~0.998515
recall error range	0.007231~0.006825	0.001997~0.001485

It can be seen from Table 1 that when adding hidden layer node number, neural Network2 convergence effect have been improved at the learn the same number sample and the same training count. So adding suitable hidden node number can improve convergence effect and reduce error of network. The network recall error comparison is shown as Fig.3.

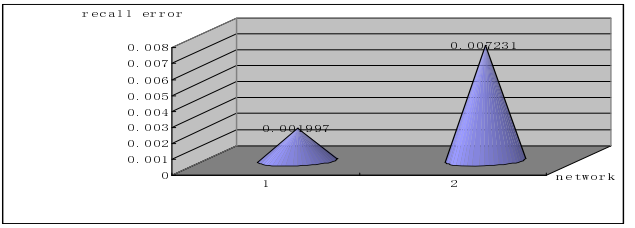
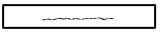
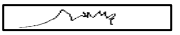
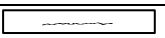
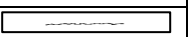


Fig. 3. The network recall error comparison

3.2 Rule 2: Adding Hidden Layer Number Doesn't Always Improve Network Convergence Effect Under Our Studying. At the Same Time It Makes Network Convergence Speed to Become Slower and Makes Network Training Time to Increase

We can prove it from the third experiment. From table 2 we find that adding hidden layer number doesn't always improve network convergence effect under our studying, At the same time it makes network convergence speed to become slower and makes network training time to increase. As below network1 is a 3 layer neural network with 21 input nodes, 10 hidden1 nodes and 1 output node. Network2 is a 4 layer neural network with 21 input nodes, 10 hidden1 nodes, 5 hidden2 nodes and 1 output node. Network2 only has different hidden layer number from Network1. Each network learns the same samples .Their train count and recall error results are shown in Table 2.

Table 2. The Comparison of Change Hidden Layer Number

	Network1	Network2
input node number	21	21
hidden1 node number	10	10
hidden2 node number	0	5
output node number	1	1
algorithm	BP	BP
transfer	S shape function	S shape function
network connecting mode	full connecting	full connecting
training count	2000	2000
error shown in instrument		
Training count again	10000	10000
error shown in instrument		
recall result range	0.998003~0.998515	0.996757~0.996882
recall error range	0.001997~0.001485	0.003243~0.003118

It can be seen from Table 2 that when adding hidden layer number. Network 2 recall error is bigger than Network 1. So adding hidden layer number convergence effect has not been improved under learning the same sample and the same training count experiment condition. The network recall error comparison is shown as Fig.4.

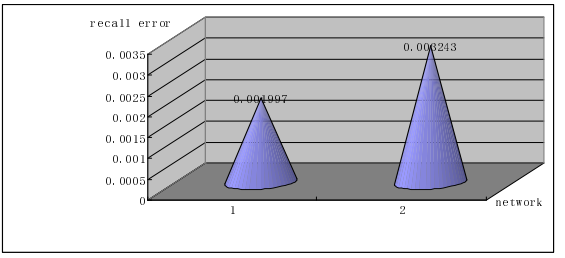


Fig. 4. The network recall error comparison

At the same time Network2 convergence speed become slower than Network1 under the same training count, i.e. training time of network2 increases after adding hidden layer number. The network training time comparison is shown as Fig.5.

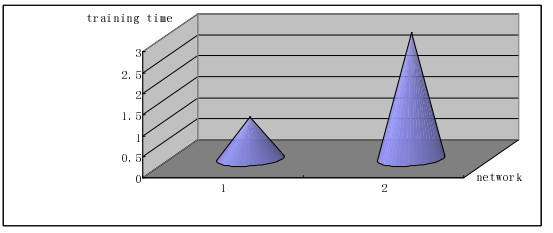


Fig. 5. The network training time comparison

3.3 Rule 3: Full Connecting Mode Suits our Medical System Better and Has Less Error

We can prove it from the first experiment. From table 3 we find that full connecting mode suits our medical system better and has less error. As below network1 is a 3 layer neural network with 21 input nodes, 10 hidden1 nodes, and 1 output node. Network2 only has different network connecting mode from Network1. Network1 is full connecting mode. Network2 is no full connecting mode. Each network learns the same samples .Their train count and recall error results are shown in Table 3.

It can be seen from Table 3 that Network 2 recall error is bigger than Network 1. So we think that non full connecting mode does not suit our medical system under our experiment condition. The network recall error comparison shown as Fig.6.

Table 3. The Comparison of Change Network Connecting Mode

	Network1	Network2
input node number	21	21
hidden1 node number	10	10
output node number	1	1
control strategy	BP	BP
transfer	S shape function	S shape function
Network connecting mode	full connecting	non full connecting
training count	12000	12000
recall result range	0.998003~0.998515	0.696756~0.696880
recall error range	0.001997~0.001485	0.303244~0.303120

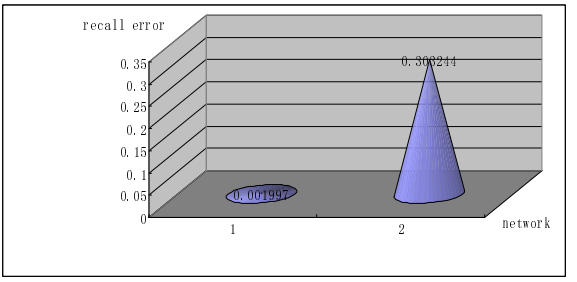


Fig. 6. The network recall error comparison

4 Artificial Intelligence and Neural Network Applied in Medical Psychiatric Intelligent Diagnosis

It is a beneficial practice combining senior psychiatrist knowledge with artificial intelligence and neural network . We build suitable neural network structure depending on above mentioned effective neural network structure rules. We use S shape function as neural network processing element output transform function. We select full connecting mode, etc. The research of medical psychiatric intelligent diagnosis and treatment system is granted by NSFC. The research is based on the diagnosis standard of ICD 10, DSM IV, more than 30 years clinical experiences of senior psychiatrists, the large scale data of epidemiological investigation and large amount of clinical research basis. The large amount sample data of epidemiological investigation comes from 14 cooperation hospitals all over the country. The diagnosis and therapy system can diagnose 61 kinds mental health disorders and give a treatment method suggesting. As below Fig.7 we give how picking up final diagnosis rules.

We build 17 kinds neural networks altogether. For example, one of network configuration is shown as Fig. 8.

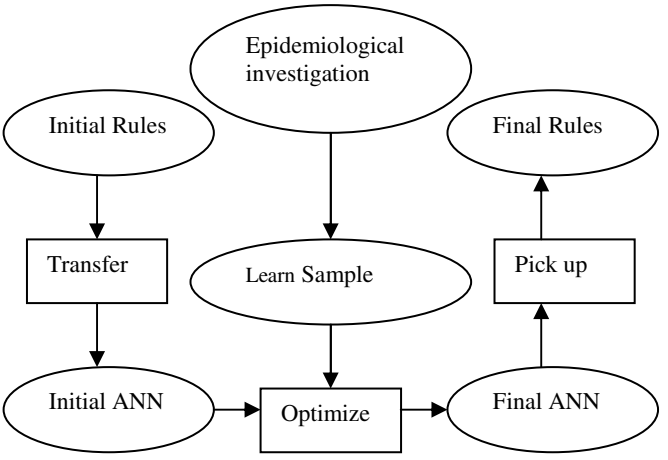


Fig. 7. Knowledge picking up

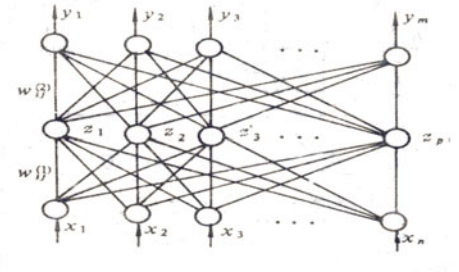


Fig. 8. Neural network connecting model

5 Results

We have applied intelligent control theory into medical psychiatric diagnosis. During studying, we find selecting suitable neural network structure is very important to BP network. Suitable neural network structure will bring less error to diagnosis system. It is key to success or failure of psychiatric intelligent diagnosis system. We have found some rules that suit our diagnosis system. Such as : full connecting mode is better and at the same time adding suitable hidden node number can improve convergence effect and reduce error of network. But adding hidden layer number doesn't always improve network convergence effect under our studying. At the same time it makes network convergence speed to become slower and makes network training time to increase. In view of the above, we build an intelligent diagnosis system. Comparing the diagnosis by computer with the senior child psychiatrists, the consistent rate of intelligent diagnosis is 99%.

6 Discussion

Computational intelligence and artificial intelligence has gotten fast development Nowadays. More mature application methodologies have appeared. For instance, SVM, ant algorithms, particle swarm optimization, rough set theory, etc. Some of them can let neural network convergence speed enhance and drop into entirely least point other than local least point. Some of them have less error and more intelligence.

Whether these new intelligent methods are better or not than artificial neural network in our system, we must study on them. We still have a lot of work to do for improving our intelligent diagnosis system. Because human brain is very complicated. We do not understand fully human brain oneself structure essence. We still have a lot of work to do for exploring better method to simulate human brain function and thought.

References

1. Hua, J.: Application of fuzzy neural network in multi-maneuvering target tracking. In: 2010 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR), pp. 92–95 (2010)
2. Chen, K., Zeniewski, J., Zhao, P., Chen, R.E.Y., et al.: Aggregate Volumetric Estimation Based on PCA and Momentum-enhanced BP Neural Network 643 Image based sieving for particle aggregate gradation. Journal of Electronics 25(2), 277–282 (2008)
3. Kapoor, R., Pal, D., Chakravartty, J.K.: Use of artificial neural networks to predict the deformation behavior of Zr-2.5Nb-0.5Cu. Journal of Materials Processing Technology 169, 99–205 (2005)
4. Ge, S.S., Hong, F., Lee, T.: Adaptive Neural Control of Nonlinear Time2 Delay Systems with Unknown Virtual Control Co2 efficients. IEEE Transaction on Systems, Man, and Cybernetics Part B: Cybernetics 34(1), 499–516 (2004)
5. Chen, M., Jiang, C., Wu, Q.: Back stepping Control for a Class of Uncertain Nonlinear Systems with Neural Network. International Journal of Nonlinear Science 3(2), 137–143 (2007)
6. Min, X.L., Liu, G.H.: BP network application development using MATLAB neural network toolbox. Computer Applications 21(8), 163–164 (2008)
7. Ni, Z.W., Ye, H.Y., Cao, H.H.: Efficient sequential pattern mining algorithm based on average value constraint satisfaction pruning strategy. Journal of University of Science and Technology of China (2), 217–220 (2007)
8. Su, J., Li, H., Dong, Q., Liu, P.: Modelling of Rolling and Aging Processes in Copper Alloy by Levenberg-Marquardt BP Algorithm. In: Wang, L., Chen, K., S. Ong, Y. (eds.) ICNC 2005. LNCS, vol. 3611, pp. 185–189. Springer, Heidelberg (2005)

A Kind of Motor-Function Evaluation Method for Upper-Limb Rehabilitation Robot*

Li Xing, Wang Jianhui, and Fang Xiaoke

College of Information Science and Engineering,
Northeastern University, Shenyang, China
lixing8245@163.com

Abstract. This paper gave motor-function evaluation content and established a synthetical evaluation model of motor function for 5-DOF upper-limb rehabilitation robot. Meanwhile, it proposed a kind of Fuzzy-AHP method with online self-correction function that are absent from the traditional methods. The result of the analysis and computation shows that Synthetical evaluation model is change rate, the faster model changes, the better effect of training is.This method can be used as an important complement to the evaluation method in hemiplegia after stroke and rehabilitation robot.

Keywords: 5-DOF upper-limb rehabilitation robot, Fuzzy-AHP, online self-correction function, synthetical evaluation model of motor function.

1 Introduction

During conventional training process of hemiplegic patient after stroke, training and evaluation mainly depend on the rehabilitation therapist's experience[1] according to these methods. But the mainkind subjective factors have bad effect on the motor-function evaluation. Therefore, it has practical significance to find effective ways and methods on motor-function rehabilitation evaluation[2] in stroke patients.

Analytic Hierarchy Process("AHP")adopts the method of combining the qualitative analysis with quantitative analysis to make decision-making process hierarchically, quantitatively [3]. But AHP could not consider fuzziness of weight, so it introduced fuzzy mathematics into AHP. Fuzzy-AHP and rehabilitation robot are magnificent prospects for rehabilitation evaluation in stroke patients. Fuzzy-AHP adopts properties combining the qualitative and quantitative supply chain, and can improve methods to solve the difficult problem of quantization[4]. So it is suit able to the uncertain evaluation of motor function in hemiplegia after stroke.

This paper adds online self-correction function based on Fuzzy-AHP, which can remove extreme factors. This method will be more precision, so as to substitute traditional evaluation method and provide complete medical dada for rehabilitation-training system.

* This work is supported by "Fundamental Research Funds for the Central Universities", "Foundation of Shenyang in 2010 (F10-205-1-57) "and "Science and Technology Planning Project of Liaoning Province in 2010 (2010020176-301)".

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 229–235.
springerlink.com

2 Fuzzy-AHP Method with Online Self-Correction Function

Evaluation factors

$C=\{c_1,c_2,c_3,\ldots,c_n\}$, in which $c_n=\{l_n,m_n,\ldots,u_n\}$, $m,n=1,2,3,\cdots$

Change expert's evaluation into fuzzy number a_{ij}

$$a_{ij}=(\frac{l_1+l_2+\ldots+l_n}{n},\frac{m_1+m_2+\ldots+m_n}{n},\ldots,\frac{u_1+u_2+\ldots+u_3}{n})$$

Calculate initial weight

$$\mathrm{D}_{\mathrm{i}}^{k}=\sum_{j=1}^{n}a_{ij}^{k}\div(\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}^{k}),i=1,2,...,n$$

Then certainty of outputs

$$v(D_1\geq D_2)=\sup_{x\geq y}[\min(u_{D1}(x),u_{D2}(y))]$$

$$v(D_1\geq D_2)=\mu(d)=\begin{cases}1 & D_1\geq D_2\\ \dfrac{l2-u1}{(D_1-u1)-(D_2-l2)} & D_1\leq D_2,\ u1\geq l2\\ 0 & otherwise\end{cases}$$

$$\mathrm{V}(D\geq D_1,D_2,\ldots\ldots D_{\mathrm{k}})=\min V(D\geq D_i),i=1,2,\ldots\mathrm{k}$$

Online detective

Average of weight for each expert $\overline{a}_i=\sum_{n=1}^{n}a_{ij}\Big/n$

Calculate D_k which is distance between vector of weight and center vector

$$D_k=\left[\sum_{i=1}^{n}(a_{ki}-a_i)^2\right]^{\frac{1}{2}}<0.5$$

It can delete extreme factors and recount average of weight.
Construct fuzzy judgment matrix

$$R_n=\begin{pmatrix}a_{11} & a_{12} & ,\cdots, & a_{1n}\\ a_{21} & a_{22} & ,\cdots, & a_{2n}\\ \cdots & \cdots & a_{ij} & \cdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{pmatrix}$$

With the same method, calculate other weights.
Synthetical evaluation model is:

$$B=D_i\circ R_n$$ [5]

The flow diagram of Fuzzy-AHP method with online self-correction function, which is shown in Fig.1:

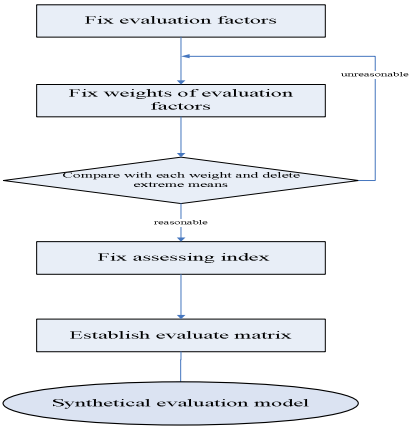


Fig. 1. The flow diagram of algorithm

3 Contents of Motor-Function Evaluation

Rehabilitative training has been practiced on hemiplegic patients by 5-DOF upper-limb rehabilitation robot. Contents of motor-function evaluation can be ensured by feedback data of rehabilitation robot.

3.1 Function Evaluation of Muscles and Joints B_1

Muscle strength level S_{muscle}

Muscle strength, which is the strength during muscle contraction, can be divided into five grades. The fifth grade is standard value.

$$S_{muscle} = \frac{\left| \frac{1}{n} \sum_{i=1}^{n} F_m - F_r \right|}{F_r} \times 100\% , \quad (n = 0,1,2,\ldots,n)$$

F_m —measurement value of muscle strength;

F_r —standard value of muscle strength;

n —the number of points sampled.

The nearer S_{muscle} approaches 100%, the nearer muscle strength approaches standard value, and the better muscle strength function.

Range of motion S_{range}

$$S_{range} = \left| \frac{1}{n} \sum_{i=1}^{n} S_m - S_r \right| , \quad (n = 0,1,2,\ldots,n)$$

S_m —the largest practical radian;

F_r —standard value of radian, which is the largest value of the active motion of each joint in rehabilitation robot;

$\frac{1}{n}\sum_{i=1}^{n} S_m$ —average radian which is produced by individual action of each joint in the same direction.

The smaller d-value of S_{range} is, the better independent movement of joint becomes.

3.2 Function Evaluation of Motion Control B_2

In the function evaluation of motion control, compulsory testing exercise is rectilinear motion of point to point, which is shown in Fig.2: Point A is beginning position of the end motion of patient's arm on fixed trunk condition and B is end position of motion, the target trajectory is the straight line of point A to point B. Red, blue are the actual trajectory of patient 1 and patient 2, respectively.

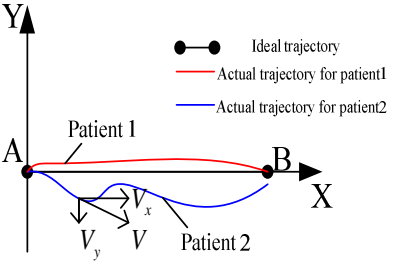


Fig. 2. Compulsory testing exercise

Degree of movement in the right direction $S_{direction}$

$$S_{direction} = \frac{1}{n}\sum_{i=1}^{n} \frac{\bar{V}_{xi}}{\bar{V}_{xi} + \bar{V}_{yi}} \times 100\%, \quad (n = 0,1,2,\ldots,n)$$

$\bar{V}_{xi}, \bar{V}_{yi}$ —average value of velocity components when patients do the translational motion laterally many times in x, y directions.

The nearer S_y approaches 100%, the more lateral motion, and the smaller longitudinal movement.

Motion time of testing exercise S_T

According to Reference [6], the time T can be shown as:

$$S_T = N / F(s)$$

N —the number of motion at particular paths;

$F(s)$ —recorded frequency.

3.3 Function Evaluation of Behavioral Coordination B_3

Degree of smoothness of the trajectory path S_{smooth}

$$S_{repeat} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(V_{yi} - \bar{V}_{yi}\right)^2} \times 100\% \ , (n = 0,1,2,\ldots,n)$$

V_{yi} —velocity components when patients do the translational motion laterally many times in y directions.

The smaller S_{smooth} and the aberrance extent of protrusive movement, the better the degree of smoothness.

Consistency of the trajectory path S_{repeat}

$$S_{repeat} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(V_{yi} - \bar{V}_{yi}\right)^2} \times 100\% \ , (n = 0,1,2,\ldots,n)$$

V_{yi} —velocity components when patients do the translational motion laterally many times in y directions;

$\bar{V}_{yi}$ —average value of velocity components when patients do the translational motion laterally many times in y directions.

The smaller S_{repeat}, the better repetitive movement. The trajectory path has well consistency.

4 Motor-Function Evaluation Model

The composition of synthetical evaluation system of motor function for upper-limb rehabilitation robot is shown in Fig.3:

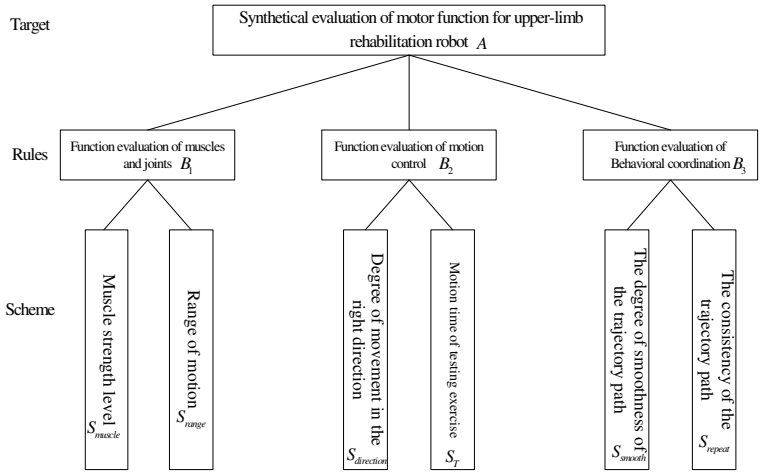


Fig. 3. Synthetical evaluation system of motor function for upper-limb rehabilitation robot

According to rehabilitation knowledge and three therapists' experience, fuzzy judgment matrix of B layer is shown in the Table1:

Table 1. Fuzzy judgment matrix of B layer

	B_1	B_2	B_3
B_1	(1,1,1)	(1/3,1/2,1/1) (1/3,1/2,1/1) (1/2,1/1,1/1)	(1/2,1/1,1/1) (1/2,1/1,1/1) (1/1,1/2,1/1)
B_2		(1,1,1)	(1,1,2) (1,2,3) (1,1,2)
B_3			(1,1,1)

So,

$$A=\begin{bmatrix}(1,1,1) & (0.39,0.67,1.00) & (0.44,0.83,1.00)\\ & (1,1,1) & (1.00,1.33,2.33)\\ & & (1,1,1)\end{bmatrix}$$

$$D_{c1}=\sum_{j=1}^{4}a_{ij}\div\sum_{i=1}^{4}\sum_{j=1}^{4}a_{ij}=(0.1509,0.2897,0.5083)$$

$$D_{c2}=(0.169,0.331,0.670)$$

$$D_{c3}=(0.1368,0.2731,0.5314)$$

Online detective each error of weight is small, so testing result is reliable.

$$V(D_{c1}\geq D_{c2})=\frac{(0.1690-0.5083)}{(0.2897-0.5083)-(0.3310-0.1690)}=0.8913,$$

$$V(D_{c1}\geq D_{c3})=1,$$

$$V(D_{c1}\geq D_{c4})=1,$$

$$d(C1)=\min V(D_{c1}\geq D_{c2},D_{c3},D_{c4})=\min(0.8913,1,1)=0.8913,$$

$$d(C2)=\min V(D_{c2}\geq D_{c1},D_{c3},D_{c4})=\min(1,1,1)=1,$$

$$d(C3)=\min V(D_{c3}\geq D_{c1},D_{c2},D_{c4})=\min(0.9583,0.8622,1)=0.8622,$$

Thus the resulting weights$(W_{c1},W_{c2},W_{c3})=(0.3086,0.3462,0.2985)$

Using the same method, the weights of S layer are:

$$(W_{s1},W_{s2},W_{s3},W_{s4},W_{s5},W_{s6})=(0.0714,0.0714,0.3214,0.1071,0.3214,0.1071)$$

Synthetical evaluation model of motor function for upper-limb rehabilitation robot is:

$$A=0.0714*S_{muscle}+0.0714*S_{range}+0.3214*S_{direction}+0.1071*S_T+0.3214*S_{smooth}$$
$$+0.1071*S_{repeat}$$

Synthetical evaluation of motor function for upper-limb rehabilitation robot, that is to say rehabilitation effect can be the shown by variational rate of A per unit time.The faster the changing per unit time, the better the rehabilitation effect is.So the function and technology of rehabilitation robot are well.

5 Summary

(1) The result of the analysis and computation shows that the weights value of $S_{direction}$ and S_{smooth} are larger, meanwhile, $S_{direction}$ and S_{smooth} are the determination for function evaluation of behavioral coordination and function evaluation of motion control.

(2) Based on fuzzy-AHP with online self-correction function, it makes up the shortcomings of traditional clinical treatment and the evaluation results are more accurate and reliable.

(3) This method has the important theory significance and the project practical value, to add an augment to evaluation method.

Acknowledgments. This work is supported by “Fundamental Research Funds for the Central Universities”, “Foundation of Shenyang in 2010 (F10-205-1-57)”and “Science and Technology Planning Project of Liaoning Province in 2010 (2010020176-301)”.

References

1. Krebs, H.I., Volpe, B.T., Aisen, M.L., et al.: Increasing productivity and quality of care: robot-aided neuro- rehabilitation. Journal of Rehabilitation Research and Development 37(6), 647–649 (2000)
2. Bin, M., Ying, G.: Evaluation for the outcome of stroke and the assumption to construct a clinical evaluation model by the combination of disease and syndrome. Chinese Journal of Clinical Rehabilitation 35(10), 155–158 (2006)
3. Sun, H.C., Tian, P., Wang, L.F.: Analytic Network Process and Decision Science. Defense Industry Press, Beijing (2011)
4. Gao, X.B.: Research on Experiment and Evaluation of Rehabilitation Training Robot. Chongqing University, Chongqing (2008)
5. Zhu, J.J.: Research on Some Problems of the Analytic Hierarchy Process and Its Application. Northeast University, Liaoning (2005)
6. Yang, X., Wang, Z., Ji, L., Bi, S.: Kinematics index for motor-function evaluation of shoulder and elbow joints. Beijing: J. Tsinghua Univ (Sci. & Tech.) 46(2), 172–175 (2006)

Dynamic Surface Control for Nonlinear Dynamic Positioning System of Ship

Yang Yang[1], Jialu Du[1], Guangqiang Li[1], Wenhua Li[2], and Chen Guo[1]

[1] School of Science and Technology, Dalian Maritime University,
Linhai Road 1, 116026 Dalian, China
[2] School of Marine Engineering, Dalian Maritime University,
Linhai Road 1,116026 Dalian, China
dujl66@163.com

Abstract. This paper addresses the problem of dynamic positioning of ships in the horizontal plane. A nonlinear controller including a first-order low-pass filter is proposed that yields convergence of the ship's position and orientation to a desired target point in the presence of unknown environmental disturbances. The proposed controller is derived by incorporating dynamic surface control technique into vectorial backstepping, which leads to no need for the differentiation of the model in the controller design and hence eliminates the problem of "explosion of terms" inherent in the backstepping approach, while being simpler to implement. In addition, it is proven that the control law can guarantee the uniformly ultimate boundedness of all signals of closed-loop dynamic positioning system of ship based on Lyapunov function under the assumption that the bound of uncertain disturbance is known. Finally, simulation results on the mathematical model of a supply ship are presented to validate the effectiveness of the proposed control scheme.

Keywords: dynamic positioning of ships, dynamic surface control, backstepping, uniformly ultimate boundedness, disturbances.

1 Introduction

Dynamic positioning (DP) is often utilized by many of the offshore applications. According to the International Maritime Organization (IMO) and the certifying class societies (DNV, ABS, LR, etc.), a dynamically positioned vessel is defined as a vessel that accurately maintains its position and heading at the fixed location or pre-determined track for marine operation purposes exclusively by means of active thrusters[1].

As ocean development moves to deeper and farther, dynamic positioning technology has more and more important practical significance and has received more and more attention. In the 1960s, the first dynamic positioning system was introduced using traditional PID control algorithms in cascade with low-pass and/or notch filter to suppress the wave-induced motion components [1]. In the 1970s, more advanced model-based control techniques based on multivariable optimal control theory and Kalman filter techniques were employed with the dynamic positioning problem by

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 237–244.
springerlink.com

[2]. Later this work was later improved and extended by numerous authors [3, 4]. However, in order to apply the linear optimal control theory and gain-scheduling techniques, these above dynamic positioning systems for ship are designed by linearizing the kinematic equations of motions about the predefined different yaw angles, from which there is no guarantee for global stability of the closed-loop system. In order to solve this problem, in the 1990s nonlinear dynamic positioning controller designs were proposed. Applying the backstepping design methodology and Lyapunov stability theory, the GUAS dynamic positioning system is designed by using 6 steps [5]. Furthermore, the design procedure is improved in [6], where the number of steps can be reduced to 2 steps in a vector setting by introducing the idea of vectorial backsteping and GES is proved. However, disturbances induced by wind, wave and current are neglected in the aforementioned designs, which is very restrictive in a practical system. In the work of [7] and [8], the important contribution of passive nonlinear observer is presented with the bias state estimation and wave filtering, where nonlinear passivity theory was used to reduce the number of tuning parameters in the control software getting rid of cumbersome linearizations. On the basis of [7], [9] presented a GAS controller for dynamic positioning of ship using only position measurements, which is validated by experimentation with a model ship scale 1:70. In [10], a global robust and adaptive output feedback controller for dynamic positioning of surface ships was designed under environmental disturbances based on a constructive method, where the ship's parameters are not required to be known and the control law forces the ship's position and orientation to globally asymptotically converge to the desired values.

Ships in operation are always disturbed by the external environment, which affects positioning accuracy of ships. Motivated by the result of dynamic surface control technique in [11], considering the external environment disturbances, we develop a new dynamic positioning control law of ship by incorporating dynamic surface control technique into the vector backstepping [6] in this paper. The designed controller eliminates the problem of "explosion of terms" inherent in the backstepping approach, while being simpler to implement. It is proven that the developed feedback controller can ensure not only the accuracy of dynamic positioning of ship, but also the stability of the ship dynamics. At last, the performances of the proposed control scheme on the dynamic positioning system are demonstrated by computer simulations of a thruster controlled supply vessel.

2 Problem Formulation

The nonlinear dynamic equations of dynamic positioning surface ships can be written in a vectorial setting in the presence of disturbances according to [1]

$$\dot{\eta} = R(\psi)\upsilon \tag{1.1}$$

$$M\dot{\upsilon} + D\upsilon = \tau + \Delta \tag{1.2}$$

where $\eta = [x, y, \psi]^T$ represents the earth-fixed positions (x , y) and the yaw angle ψ in vector. $\upsilon = [u, v, r]^T$ represents the ship's surge, sway and yaw velocities in the body-fixed frame. The rotation matrix R is defined as

$$R(\psi)=\begin{bmatrix}\cos\psi & -\sin\psi & 0\\ \sin\psi & \cos\psi & 0\\ 0 & 0 & 1\end{bmatrix} \tag{2}$$

which is orthogonal, that is $R^{-1}(\psi)=R^{T}(\psi)$, and has the property $\|R\|=1$. The matrix of inertia including hydrodynamic added inertia effects, M is invertible symmetric and positive definite under the assumption that the ship is port-starboard symmetric. D is linear damp matrix, which is standard for low speed application [1]. The vector $\tau\in R^3$ is the control vector provided by the ship propulsion system, which can produce surge and sway forces as well as a yaw moment. The disturbance vector $\Delta\in R^3$ is the uncertain equivalent forces and moment interacting with the vessel dynamics from the environment and unmodeled dynamics. There exists a known positive constant ρ such that $|\Delta|<\rho$, which is reasonable since the disturbance component can be largely attributed to the exogenous effects of the environment which have finite energy and, hence, are bounded.

The control objective of this paper is to design a feedback control law $\tau=\tau(\eta,\upsilon)$ for the system (1) such that the position (x, y) and the yaw angle ψ of the ship converge to a desired target position $\eta_d=[x_d \quad y_d \quad \psi_d]^T$.

3 Dynamic Surface Control Design

In this section, a dynamic surface control algorithm in [11] is extended to deal with the dynamic positioning control problem of ship (1) by incorporating with vectorial backstepping, and the stability analysis of the closed-loop system is also given. The design procedure consists of 2 steps.

Step 1. Defining the position error surface vector $S_1\in R^3$ of dynamic positioning ship (1)

$$S_1:=\eta-\eta_d \tag{3}$$

Then

$$\dot{S}_1=R\upsilon \tag{4}$$

Choose an intermediate control vector function $\phi_1\in R^3$ for the virtual control vector υ to stabilize (4) as follows

$$\phi_1=-R^{-1}K_1S_1 \tag{5}$$

where $K_1=K_1^T>0$ is the 3×3 design parameter matrix called as the surface gain matrix. If υ were to track ϕ_1 asymptotically, S_1 would asymptotically converge to the origin.

In order to avoid the problem the explosion of terms associated with the traditional backstepping, we introduce a new state vector $X_{2d}\in R^3$ and let ϕ_1 pass through a first-order low-pass filter with the time constant T_2 to obtain X_{2d} and its differentiation $\dot{X}_{2d}$:

$$T_2\dot{X}_{2d}+X_{2d}=-R^{-1}K_1S_1,\quad X_{2d}(0)=\phi_1(0) \tag{6}$$

where $T_2 > 0$ is a positive scalar. Then

Step 2. Defining the velocity error surface vector $S_2 \in R^3$ of dynamic positioning ship (1) as follows

$$S_2 := \upsilon - X_{2d} \tag{7}$$

Then we have

$$M\dot{S}_2 = -D\upsilon + \tau + \Delta - M\dot{X}_{2d} \tag{8}$$

Design

$$\tau = M\dot{X}_{2d} - K_2 S_2 + D\upsilon - \frac{S_2\rho^2}{2\varepsilon} \tag{9}$$

where $K_2 = K_2^T > 0$ is the 3×3 design parameter matrix called as the surface gain matrix. $\varepsilon > 0$ is an arbitrarily small positive design constant which is a measure of the accuracy of dynamic positioning control.

It is obvious that this control law (9) does not involve model differentiation and thus has prevented the explosion of terms, the problem inherent in the backstepping approach, which is that due to the presence of the first-order low-pass filter (6). The intermediate control vector function ϕ_1 isn't differentiated and therefore, the control law is simpler in the design and easier to implement.

From the definition of S_2 and Equation (4), we have

$$\dot{S}_1 = R(S_2 + X_{2d}) \tag{10}$$

According to (8) and (9)

$$M\dot{S}_2 = -K_2 S_2 - \frac{S_2\rho^2}{2\varepsilon} + \Delta \tag{11}$$

Theorems 1. Consider the nonlinear mathematical model of dynamic positioning ship (1) under disturbance vector with a constant ρ as its upper bound. Given any ε>0, there exist the surface gain matrices K_1 and K_2 and filter time constant T_2 such that the dynamic surface controller (9) with the first order filter (6) guarantees that the closed-loop error system (10) and (11) of dynamic positioning ship is globally uniformly ultimately bounded, and the actual position of ship eventually converges to the desired target position $\eta_d = [x_d \quad y_d \quad \psi_d]^T$ with arbitrary precision smaller than ε.

Proof. Define the boundary layer error function vector Y_2 as

$$Y_2 = X_{2d} - \phi_1 = X_{2d} + R^{-1}K_1 S_1 \tag{12}$$

In view of (6) and (12), then

$$\dot{Y}_2 = -\frac{Y_2}{T_2} + \dot{R}^{-1}K_1 S_1 + R^{-1}K_1\dot{S}_1 \tag{13}$$

Clearly

$$\left\| \dot{Y}_2 + \frac{Y_2}{T_2} \right\| \le \eta_2(S_1, S_2, K_1, \eta, \upsilon, X_{2d}, T_2) \tag{14}$$

for some continuous function, such as $\eta_2(S_1, S_2, K_1, \eta, \upsilon, X_{2d}, T_2) = \left|\dot{R}^{-1}K_1 S_1\right| + \left|R^{-1}K_1\dot{S}_1\right|$.

Then, the closed-loop dynamics in terms of the error surface vector $S_1$and S_2 and the boundary layer error vector Y_2 can be expressed as

$$\dot{S}_1 = R(S_2 + X_{2d}) \tag{15}$$

$$M\dot{S}_2 = -K_2 S_2 - \frac{S_2 \rho^2}{2\varepsilon} + \Delta \tag{16}$$

Consider the Lyapunov function candidate

$$V = \frac{1}{2} S_1^T S_1 + \frac{1}{2} S_2^T M S_2 + \frac{1}{2} Y_2^T Y_2 \tag{17}$$

Then

$$\dot{V} = S_1^T \dot{S}_1 + S_2^T M \dot{S}_2 + Y_2^T \dot{Y}_2 \tag{18}$$

In light of (15) and (12), considering $\|R\| = 1$ and Cauch inequality, we have

$$S_1^T \dot{S}_1 = S_1^T R(S_2 + X_{2d}) \le -S_1^T K_1 S_1 + \frac{\|S_1\|^2 + \|S_2\|^2}{2} + \frac{\|S_1\|^2 + \|Y_2\|^2}{2} \tag{19}$$

In light of (16), $|\Delta| < \rho$ and Cauch inequality, we have

$$S_2^T M \dot{S}_2 = -S_2^T K_2 S_2 + S_2^T \Delta - \frac{S_2^T S_2 \rho^2}{2\varepsilon} \le -S_2^T K_2 S_2 + \|S_2\| \cdot \rho - \frac{S_2^T S_2 \rho^2}{2\varepsilon} \tag{20}$$

Considering the following complete square inequality:

$$\|S_2\| \cdot \rho - \frac{\|S_2\|^2 \rho^2}{2\varepsilon} \le \frac{\varepsilon}{2} \tag{21}$$

Then

$$S_2^T M \dot{S}_2 \le -S_2^T K_2 S_2 + \frac{\varepsilon}{2} \tag{22}$$

From (13) and (14), we have

$$Y_2^T \dot{Y}_2 = Y_2^T \left(-\frac{Y_2}{T_2} + \dot{R}^{-1} K_1 S_1 + R^{-1} K_1 \dot{S} \right) \le -\frac{\|Y_2\|^2}{T_2} + \|Y_2\| \cdot \eta_2 \tag{23}$$

Then,

$$\dot{V} \le -S_1^T K_1 S_1 - S_2^T K_2 S_2 - \frac{\|Y_2\|^2}{T_2} + \varepsilon + \frac{\|S_1\|^2 + \|S_2\|^2}{2} + \frac{\|S_1\|^2 + \|Y_2\|^2}{2} + \frac{\|Y_2\|^2 \eta_2^2}{2\varepsilon} \tag{24}$$

Select the time constants of the filter inductively $1/T_2 = 1/2 + M_2^2/2\varepsilon + \alpha_0$, where $M_2 \ge \eta_2$. Therefore

$$\dot{V} \le -S_1^T K_1 S_1 - S_2^T K_2 S_2 - \alpha_0 Y_2^T Y_2 + \varepsilon + S_1^T S_1 + \frac{S_2^T S_2}{2} \tag{25}$$

Let $K_1 = I + \alpha_0 I$ and $K_2 = I/2 + \alpha_0 M$, where α_0 is a positive real number. Accordingly, Equation (25) can be rewritten as following

$$\dot{V} \le -S_1^T(I+\alpha_0 I)S_1 - S_2^T\left(\frac{M^{-1}}{2}+\alpha_0 I\right)MS_2 - \alpha_0 y_2^T y_2 + \varepsilon + S_1^T S_1 + \frac{1}{2}S_2^T M^{-1}MS_2$$

$$\le -2\alpha_0 V + \varepsilon \tag{26}$$

It is seen from Equation (26) that V is uniformly ultimately bounded. Hence, the error surface vectors S_1 and S_2 are uniformly ultimately bounded according to the definition of V. Furthermore, all signals of the closed-loop system of ship dynamic positioning (10) and (11) are uniformly ultimately bounded and the actual position of ship eventually converges to the desired target position $\eta_d = [x_d \quad y_d \quad \psi_d]^T$ with arbitrary precision smaller than ε. The Theorem 1 is proved.

4 Simulation

In our simulation study, we consider a supply vessel, whose length is $76.2m$ and mass is $4.591\times10^6\,kg$. The dynamic parameters of the ship model represented by (1) are given as follows [1]

$$M = \begin{bmatrix} 5.3122\times10^6 & 0 & 0 \\ 0 & 8.2831\times10^6 & 0 \\ 0 & 0 & 3.7454\times10^6 \end{bmatrix},\ D = \begin{bmatrix} 5.0242\times10^4 & 0 & 0 \\ 0 & 2.7229\times10^5 & -4.3933\times10^6 \\ 0 & -4.3933\times10^6 & 4.1894\times10^8 \end{bmatrix}$$

Let the desired target position of the ship be $\eta_d = [0 \quad 0 \quad 0]^T$. The disturbance is chosen as $\Delta = [1000N \quad 1000N \quad 1000Nm]^T$, then we can choose $\rho = 1000$ in the simulations. Here, the design parameters are taken as $K_1 = diag([1.1, 1.1, 1.1])$, $K_2 = diag([5.3\times10^5,\ 8.3\times10^5,\ 3.7454\times10^8])$, $\varepsilon = 0.09$ and $T_2 = 0.09$. The initial conditions of the ship are $\eta_0 = [10m \quad 20m \quad 5^\circ]^T$ and $\upsilon(0) = [0m/s \quad 0m/s \quad 0^\circ/s]^T$. Simulation results are depicted in Fig. 1-3.

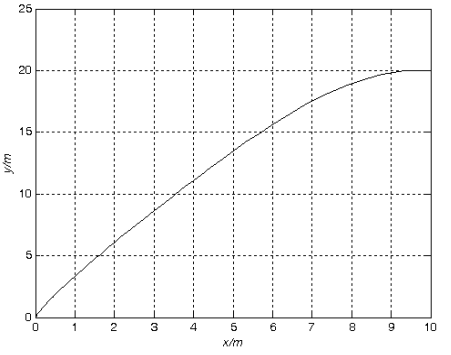


Fig. 1. The actual position plot of the supply ship in *xy*-plane. The desired position is (0m, 0m).

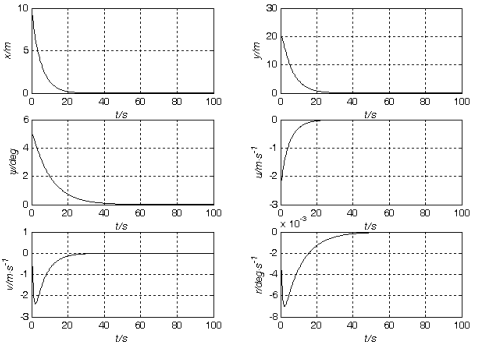


Fig. 2. The curves of the actual position, yaw angle, surge velocity, sway velocity and yaw rate versus time

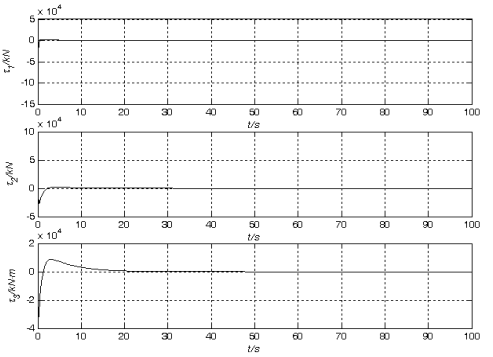


Fig. 3. The curves of the surge control force, sway control force and yaw control torque versus time.

Figure 1 shows that the proposed controller is able to force the ship to the desired position under disturbances. And the position, yaw angle, surge, sway and yaw rate of ship are presented in Fig.2 respectively, which indicate that the actual ship position (x , y) and the actual yawing angle ψ can approach to the desired position $\eta_d = [0 \quad 0 \quad 0]^T$ in around 45s with high precision. The corresponding control inputs are presented in Fig.3. It is seen that the control forces and torque are smooth and reasonable. Simulation results demonstrate the effectiveness of the proposed control scheme.

5 Conclusion

In this paper, a dynamic positioning control law of ship is proposed utilizing dynamic surface control approach and vectorial backstepping design tool under disturbances induced by wind waves, and ocean currents. The dynamic surface control algorithm, while providing the ease in designing and implementing nonlinear control laws,

guarantees the stability and dynamic performance of the ship dynamic positioning system, as well as arbitrarily bounded regulation of the actual position of ships to the desired target position under disturbances. The stability analysis and the simulation results show that the proposed design method is effective.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (No. 51074053 and 61074053), in part by the Higher Education Research Fund of Education Department of Liaoning (LT2010013), China, in part by Applied Basic Research Program sof Ministry of Transport of P. R. China (No. 2011-329-225-390), and in part by the Fundamental Research Funds for the Central Universities (No.DMU2011QN030).

References

1. Fossen, T.I.: Marine Control Systems: Guidance, Navigation and Control of Ships, Rigs and Underwater Vehicles. Marine Cybernetics AS. Trondheim, Norway (2002)
2. Balchen, J.G., Jenssen, N.A., Sælid, S.: Dynamic Positioning Using Kalman Filtering and Optimal Control Theory. In: Proceedings of IFAC/IFIP Symposium on Automation in Offshore Oil Field Operation, Bergen, Norway, pp. 183–186 (1976)
3. Balchen, J.G., Jenssen, N.A., Mathisen, E., Sælid, S.A.: Dynamic Positioning System Based on Kalman Filtering and Optimal Control. Modeling, Identification and Control 1(3), 135–163 (1980)
4. Fung, P.T.-K., Grimble, M.J.: Dynamic ship positioning using a selftuning Kalman filter. IEEE Trans. on Automatic Control 28(3), 339–349 (1983)
5. Grøvlen, A., Fossen, T.I.: Nonlinear control of dynamic positioned ships using only position feedback: An observer backstepping approach. In: Proceedings of the 35th IEEE Conf. Decision Contr., CDC 1996, December 11–13, pp. 3383–3393. Kobe, Japan (1996)
6. Fossen, T.I., Grøvlen, Å.: Nonlinear output feedback control of dynamically positioned ships using vectorial observer backstepping. IEEE Trans. On Control Systems Technology 6(1), 121–128 (1998)
7. Fossen, T.I., Strand, J.P.: Passive nonlinear observer design for ships using Lyapunov methods: Experimental results with a supply vessel. Automatica 35(1), 3–16 (1999)
8. Strand, J.P., Fossen, T.I.: Nonlinear passive observer for ships with adaptive wave filtering. In: Nijmeijer, H., Fossen, T.I. (eds.) New Directions in Nonlinear Observer Design, pp. 113–134. Springer-Verlag London Ltd., Heidelberg (1999)
9. Loria, A., Fossen, T.I., Panteley, E.: A Separation Principle for Dynamic Positioning of Ships: Theoretical and Experimental Results. IEEE Trans. on Control Systems Technology 8(2), 332–343 (2000)
10. Do., K.D.: Global Robust and Adaptive Output Feedback Dynamic Positioning of Surface Ships. In: The Proceedings of 2007 IEEE International Conference on Robotics and Automation, Roma, Italy, pp. 10–14 (April 2007)
11. Swaroop, D., Hedrick, J.K., Yip, P.P., Gerdes, J.C.: Dynamic Surface Control for a Class of Nonlinear Systems. IEEE Trans. on Automatic Control 45(10), 1893–1899 (2000)

Experimental Investigation of Performance Evaluation on Velocity and Acceleration of Motion System Based on ELT Linear Motion Unit

Shuaihe Zhao and Yaqing Zheng

College of Mechanical Engineering and Automation, Huaqiao University, 361021, Xiamen, China
feitianbao_@163.com, yq_zheng@hqu.edu.cn

Abstract. To evaluate the performance of velocity and acceleration of the motion system based on ELT linear motion unit, the experimental investigation of the motion of the linear motion unit driven by Panasonic servo motor controlled by PMAC control card in different conditions is made. This paper presents the establishment of the experimental prototype and introduces how to adjust the parameters of PMAC control card and how to control servo motor based on PID, which makes the system realize stable movement within a range of velocity and achieves the corresponding acceleration critical value. The experimental results have indicated that the velocity and the acceleration performance of the ELT linear motion unit can meet the experimental requirements.

Keywords: linear motion unit, control strategy, velocity, acceleration, performance evaluation.

1 Introduction

The motion system studied in this paper is a part of transmission system of the meter-sized prototype for 2 degree-of-freedom (DOF) gantry crane robot and 6-DOF parallel crane robot which includes 3 rigid-flexible hybrid chains. Mechanism theory and trajectory tracking control technique of the meter-sized prototype are studied in the project. This paper presents the performance evaluation of the velocity and acceleration of the experimental prototype.

The traditional transmission method of "rotary servo motor + ball screw" is widely applied, but there are several disadvantages, as follows [1, 2]: 1) It has weak stiffness, big inertia and is difficult to gain high acceleration and velocity; 2) It is seriously nonlinear and difficult to achieve closed-loop control; 3) It has a larger transmission error which can affect the accuracy of motion; 4) It has low efficiency of transmission.

The linear motion unit, which has improved the disadvantages of "rotary servo motor + ball screw" transmission method, is widely applied in robots and machine tools [3-6].

BLT60 linear motion unit, the spindle of which is driven with ball screw spindle produced by BAHR in Germany, is chosen as a part of transmission system of the

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 245–251.
springerlink.com

meter-sized prototype for 2 degree-of-freedom (DOF) gantry crane robot and 6 DOF parallel crane robot which includes 3 rigid-flexible hybrid chains [7-9]. This paper presents the establishment of the experimental prototype and introduces how to adjust the parameters of PMAC control card and how to control servo motor based on PID, which makes the system realize stable movement within a range of velocity and achieves the corresponding acceleration critical value.

2 The Establishment of the Experimental Platform

2.1 The Design Requirements and the Main Functions of the Platform

In the experiment, the stable movement within a range of velocity of the system composed by AC servo motor and linear motion unit should be realized, and the velocity of the linear motion unit should be limited in the range of 0 to 0.5 meters per second because the length of the experiment platform is 1.2 meters. Before the installation of the measuring equipment of linear motion, the velocity and acceleration, which can be converted by the measurable motor speed to be used in estimating the stability of the system, can not be measured directly.

2.2 The Configuration of the System

To ensure the stiffness of the system, high-strength aluminum alloy profiles are chosen as the material of the frame of the experimental platform. The ELT60 linear motion unit connected with Panasonic AC servo motor which is controlled by PMAC control card is fastened on the frame, shown in Fig. 1. The signal of the position detection is provided by the encoder of servo motor, and the real-time running situation of the motor can be gained by the half closed-loop system composed by PMAC and the servo system. In the control process of the motor, PID correction is used in the feedback channel and the feedforward control is used in the forward channel.

Fig. 1. The experimental platform of 2-DOF Gantry Crane Robot

2.3 The Method and the Principle of the Experiment

The AC servo motor should be controlled in velocity mode to get the required velocity of the linear motion unit. And some useful parameters can be gained by the Pmac TuningPro which is a part of the software suite of PMAC. It provides real-time plotting utilities, motor status screen, watch window, position window and a simple terminal to help monitor and perform different tasks on a system [10].

In the experiment, the data of motor speed can be gained at present, and the time of acceleration and deceleration can be read on Pmac TuningPro; so the velocity and acceleration of the linear motion unit can be converted by them. In the future, in order to measure and control the linear motion of the slider more accurately, the specialized detection device will be assembled on the linear motion unit to make up a closed-loop linear motion control system.

3 Investigation of the Experiment

3.1 Adjustment of the Motor Status

Step response is usually used in estimating the feedback filter for quadratic systems, for example the motor current controlled driving the inertia load. In the experiment, a step move is done and the plotted response which indicates the slow rise time of the system is shown in Fig. 2. In the case, Ix30 (Kp), Ix31 (Kd), Ix33 (Ki) should be adjusted to get the fastest rise time without a huge amount of overshoot. It is necessary for further tradeoffs between Kp and Kd to get the desired response, i. e. fast response with no overshoot as shown in Fig. 3. This figure shows the step response with a Kp of 68000, a Ki of 0 and a Kd of 2700.

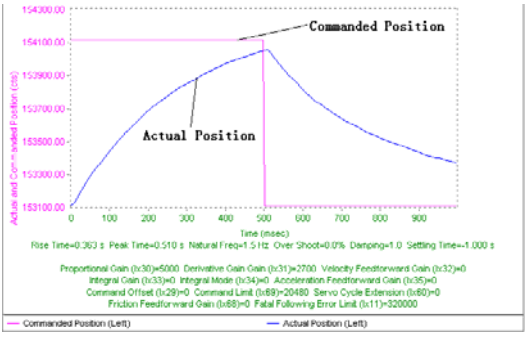


Fig. 2. Initial step move response

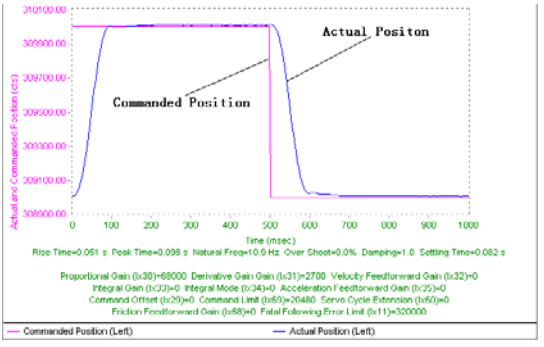


Fig. 3. Modified step move response

In a servo system without feedforward or dynamic error integration, there is a continual error between the commanded position and the actual position in a profiled move known as following error which is usually proportional to the velocity and/or the acceleration. In the experiment, a series of parabolic moves which are cubic in position, parabolic in velocity and linearly varying in acceleration are done to adjust the feedforward terms which can reduce the following error. In Fig. 4, the following error is between minus $150(cts)$ and $150(cts)$ when velocity feedforward gain (Kvff) is 5000 and acceleration feedforward gain (Kaff) is 0. By adjusting Kvff to be equal to 8255 and Kaff to be equal to 0, the maximum following error is reduced from 150 to 8 counts and the following resembles the velocity curve, shown in Fig. 5. Then an almost perfectly tuned motor is achieved.

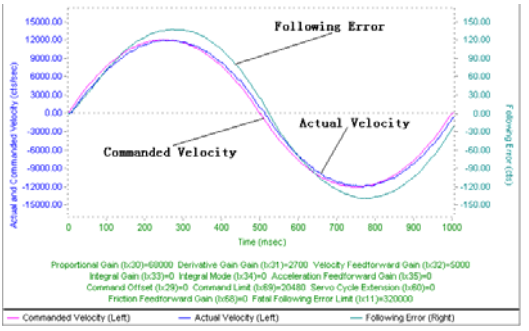


Fig. 4. Initial parabolic move response

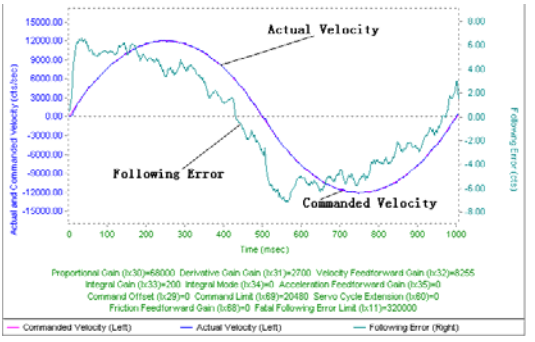


Fig. 5. Modified parabolic move response

3.2 The Performance of Velocity and Acceleration of the Motion System Based on Elt Linear Motion Unit

After repeating experiments, it is found that the values shown in analog monitor (RIN) of the servo motor coincide with the voltage output. For example when the analogue value in RIN is 1, the output voltage is $1(V)$. The rated speed of the motor

is $3000(r/\min)$, and the maximum speed is $5000(r/\min)$. In the condition of a Kd of 2000 to 3000 and a Kp of 10000, the maximum analogue value approximately equals to 1, the output voltage approximately equals to $1(V)$ and the output speed of the motor is $500(r/\min)$. The conversion relationship between the speed of linear motion unit and the motor speed is as follows:

$$v = \frac{A \cdot n \cdot ph}{60000} \tag{1}$$

Here, v denotes the velocity of linear motion unit (m/s); $A = 1$, denotes the reduction ratio, because the linear motion unit and the motor are coupled directly; n denotes the motor speed $(r/\min)$; $ph = 20(mm)$, denotes the lead of the linear motion unit which is double thread with the pitch of $10(mm)$.

Table 1. The operation situation of different acceleration of the system

Acceleration(g) (Motor speed is $500\,r/\min$)	Operation situation (No-load)	Operation situation (Load $10kg$)	Operation situation (Load $25kg$)
0.65	Smooth operation	Smooth operation	Smooth operation
0.69	Smooth operation	Smooth operation	Slight vibration
0.71	Smooth operation	Smooth operation	Strong vibration
0.78	Smooth operation	Slight vibration	No test
0.81	Slight vibration	Strong vibration	No test
0.85	Strong vibration	No test	No test

In the case of no load, in order to achieve different acceleration of the linear motion unit there are two steps to do. First, keep the velocity of the linear motion be $1/6(m/s)$ when the motor speed is $500(r/\min)$; second, adjust the parameters to get different response time. The operation situation of the system is shown in table 1. Smooth operation of the system can be got when the acceleration is less than $0.81g$ $(g = 9.8m/s^2)$; but the vibration occurs when the acceleration is $0.81g$; strong vibration comes when the acceleration is more than $0.85g$.

The experimental results with different loads are as follows: (1) With the load of $10(kg)$, the critical acceleration of vibration of the system is $0.78g$. At this time, the system has slight vibration; (2) With the load of $25(kg)$, the critical acceleration turns to be $0.69g$. So the stationarity of the system decreases with larger load.

Table 2. The operation situation of different motor speed of the system (the acceleration is less than 0.6g)

Motor speed ($r/\min$)	500	1000	1250	1500
The velocity of linear motion unit (m/s)	1/6	1/3	5/12	1/2
Operation situation (No-load)	Smooth operation	Smooth operation	Smooth operation	Smooth operation
Operation situation (Load $10kg$)	Smooth operation	Smooth operation	Smooth operation	Smooth operation
Operation situation (Load $25kg$)	Smooth operation	Smooth operation	Smooth operation	Smooth operation

When the motor speeds are $500(r/\min)$, $1000(r/\min)$, $1250(r/\min)$ and $1500(r/\min)$ respectively, the corresponding velocities of the linear motion unit are $1/6(m/s)$, $1/3(m/s)$, $5/12(m/s)$ and $1/2(m/s)$. Meanwhile adjust the parameters to keep the acceleration being less than $0.6g$, the smooth operation of the system can be got, as shown in table 2. Duo to the limit of the length of the linear motion unit, the higher velocity can not be tested. The system can operate smoothly within the required range of velocity from 0 to $1/2(m/s)$.

3.3 Analysis of the Experimental Results

The experimental results show that:

(1) The system can operate smoothly within the required range of velocity;
(2) The vibration of the system occurs within higher acceleration;
(3) The acceleration critical value of the system decreases with larger load.
(4) In the experiment, the minimum acceleration is $0.69g$ and it meets the requirement of the prototype.

4 Conclusions

The principle and experimental method of the performance evaluation on velocity and acceleration of motion system based on ELT linear motion unit are discussed, and the initial experimental investigations are done in this paper. We come to the conclusion that the motion system can meet the basic requirement of velocity and acceleration of prototype.

The experiment indicates that this evaluation method can verify the performance of the linear motion unit used in this situation. In further experiments, other parts of the prototype should be improved, for example to get more accurate data the detection device of the velocity and acceleration should be assembled on the linear motion unit.

References

1. Xiao, Z.: The Development Trend of Ball Screw (In Chinese). Manufacturing Technology and Machine Tool 4, 11–12 (2000)
2. Zhang, G., Zhou, J.: Analysis of Linear Motor and Ball Screw of High-Speed Linear Motion Unit (In Chinese). Machinist Metal Cutting 3, 25–27 (2007)
3. Huang, Z.: The Latest Development of the High Speed Precision Ball Screws and Its Application (In Chinese). Aeronautical Manufacturing Technology 4, 36–40 (2003)
4. Song, X., Liu, J., Wang, Z.: Research and Development of Test System of Conbination Property of High-Speed Ball Screw Unit 39(3), 34–36 (2005)
5. Xiao, Z.: The Development Trend of Structure and Properties of Ball Screw (In Chinese). Maschinen Markt 9, 100–102 (2001)
6. Li, F.: The Design of High-Speed Linear Motion Unit and Research of Experimental Technique of Performance Evaluation (In Chinese). Master Thesis, pp. 21–32. Chongqing University, Chongqing (2006)
7. Zheng, Y., Lin, Q., Wu, J., Mitrouchev, P.: Analysis of Inverse Kinematics and Dynamics of a 6-Degree-of–freedom Wire-Driven Parallel Gantry Crane Robot. In: Proceedings of 2009 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Singapore, pp. 1786–1791 (2009)
8. Zheng, Y.: Research on Key Theoretical Issues of Wire-Driven Parallel Kinematic Manipulators and the Application to Wind Tunnel Support Systems (In Chinese). PhD Dissertation, Quanzhou:Huaqiao University. pp. 4–104 (2004)
9. Zheng, Y., Wu, J.: Trajectory Control a 2-Degree-of-Freedom Gantry Crane Robot (In Chinese). Machine Design and Research 26(3), 23–28 (2010)
10. Software Reference Manual of Pmac Tuning Pro. Delta Tau Data Systems, Inc. (2003)

Batch to Batch Iterative Learning Control of a Fed-Batch Fermentation Process

Jie Zhang[1], Zhihua Xiong[2], Delautre Guillaume[1], and Alexandre Lamande[1]

[1] School of Chemical Engineering and Advanced Materials,
Newcastle University, Newcastle upon Tyne NE1 7RU, UK
[2] Department of Automation, Tsinghua University, Beijing 100084, China
jie.zhang@newcastle.ac.uk

Abstract. Batch to batch iterative learning control of a fed-batch fermentation process is studied in this paper. Taking the immediate previous batch as the reference batch, a linearised model relating the deviations in the control profiles with the deviations in the quality variable trajectories is obtained. The linearised model is used in calculating the control policy updating for the current batch through solving an optimisation problem. In order to cope with nonlinearities, the batch-wise linearised model is re-identified after each batch run. Simulation results show that the proposed method can overcome the effect of model-plant mismatches and unknown disturbance and improve the process operation from batch to batch.

Keywords: batch process, fermentation, data-driven model, iterative learning control.

1 Introduction

Fermentation is an important process in pharmaceutical and biochemical industry. Process optimisation is important in improving product quality and production efficiency. The off-line calculated control policies (recipes) for batch fermentation processes may not be optimal when implemented on the processes due to model-plant mismatches and/or the presence of unknown disturbances [1]. Utilising the repetitive nature of batch processes, fermentation operation recipes can be modified from batch to batch in order to overcome the detrimental effect of model-plant mismatches and unknown disturbances.

Recently, iterative learning control (ILC) has been used in the batch-to-batch control of batch processes to directly update input trajectory [2, 3]. The basic idea of ILC is to update the control policy for a new batch run using the information from previous batch runs so that the output trajectory converges asymptotically to the desired reference trajectory. Zhang [4] proposes a neural network based batch to batch control strategy where a linearised model is obtained from the neural network model and is used in ILC. Xiong and Zhang [5] present a recurrent neural network based ILC scheme for batch processes where filtered recurrent neural network prediction errors from previous batches are added to the model predictions for the current batch and optimisation is performed based on the updated predictions. Xiong and Zhang [6]

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 253–260.
springerlink.com

propose a batch to batch ILC strategy using linearised time variant perturbation models that are identified from process operational data.

This paper presents a study on using ILC to control a fed-batch fermentation process with batch-wise updated models. A linearised model, linearised upon the input and output trajectories of a reference batch, is identified from process operation data. After each batch run, the linearised model is re-identified to account for nonlinearities and process operating condition variations. In order to cope with process variations and disturbances, the reference batch can be taken as the immediate previous batch. In such a way, the model is a batch-wise linearised model and is updated after each batch. The newly obtained process operation data after each batch is added to the historical data base and an updated linearised model is re-identified. ILC strategy is derived based on the batch-wise linearised model through solving an optimisation problem.

2 A Fed-Batch Fermentation Process

The process considered in this paper is a fed-batch yeast fermentation process taken from [7], where a detailed kinetic and dynamic model is presented. The model is developed based on mass balances for glucose, ethanol, oxygen and biomass concentrations. In this study, a simulation programme is developed in MATLAB using the kinetic and dynamic model given in [7] and is verified with the results presented in [7]. The operation objective is to produce maximum amount of biomass by adjusting the glucose feed rate subject to operation constraints.

The batch time for this process is 16.5 hours. The batch duration is divided into 10 stages with equal length and the glucose feed rate is held constant during each stage. Thus the control policy consists of 10 glucose feed rates at these stages.

3 Batch to Batch Iterative Learning Control with Batch-Wise Updated Models

3.1 Linearised Models for Batch Processes

Consider batch processes where the batch run length (t_f) is fixed and consists of N sampling intervals or stages (i.e. $N=t_f/h$, with h being the sampling time or stage length). Product quality variables (outputs), $y\in R^n$ ($n\geq 1$), can be obtained off-line by analysing the samples taken during the batch run and the manipulated variable, $u\in R^m$ ($m=1$ in this work), can be measured at each sampling time on-line. The product quality and control trajectories are defined, respectively, as

$$\mathbf{Y}_k=[y_k^{\mathrm{T}}(1), y_k^{\mathrm{T}}(2),\ldots, y_k^{\mathrm{T}}(N)]^{\mathrm{T}} \tag{1}$$

$$\mathbf{U}_k=[u_k(0), u_k(1),\ldots,u_k(N\text{-}1)]^{\mathrm{T}} \tag{2}$$

where the subscript k denotes the batch index. The desired reference trajectories of product quality are defined as

$$\mathbf{Y}_d=[y_d^{\mathrm{T}}(1), y_d^{\mathrm{T}}(2),\ldots, y_d^{\mathrm{T}}(N)]^{\mathrm{T}} \tag{3}$$

A batch process is typically modelled with a dynamic model, but it would be convenient to consider a static function relating the control policy to the product quality sequences over the whole batch duration [6].

$$\mathbf{Y}_k = \mathbf{F}(\mathbf{U}_k) + \mathbf{v}_k \tag{4}$$

where $\mathbf{F}(\cdot)$ represents the non-linear static functions between $\mathbf{U}_k(t)$ and $y_k(t)$ at different sampling times and $\mathbf{v}_k=[v_k^T(0), v_k^T(1),\ldots, v_k^T(N\text{-}1)]^T$ is a vector of measurement noises.

Linearising the non-linear batch process model described by Eq(4) with respect to $\mathbf{U}_s$ around the nominal trajectories ($\mathbf{U}_s$, $\mathbf{Y}_s$), the following can be obtained.

$$\mathbf{Y}_k = \mathbf{Y}_s + \left.\frac{\partial \mathbf{F}(\mathbf{U}_k)}{\partial \mathbf{U}_k}\right|_{\mathbf{U}_s} (\mathbf{U}_k - \mathbf{U}_s) + \mathbf{w}_k + \mathbf{v}_k \tag{5}$$

where $\mathbf{w}_k=[w_k^T(1), w_k^T(2), \ldots, w_k^T(N)]^T$ is a sequence of model errors due to the linearisation (i.e., due to negaleting the higher oder terms) and $\mathbf{v}_k$ represents the effects of noise and unmeasured disturbances. Define the linearised model $\mathbf{G}_s$ as

$$\mathbf{G}_s = \left.\frac{\partial \mathbf{F}(\mathbf{U}_k)}{\partial \mathbf{U}_k}\right|_{\mathbf{U}_s} \tag{6}$$

The structure of $\mathbf{G}_s$ is restricted to the following lower-block-triangular form due to the causality.

$$\mathbf{G}_s = \begin{bmatrix} g_{10} & 0 & \cdots & 0 \\ g_{20} & g_{21} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ g_{N0} & g_{N1} & \cdots & g_{NN-1} \end{bmatrix} \tag{7}$$

The linearised model can be identified from historical process operation data using multiple linear regression (MLR) [6]. To cope with process drift, the linearised model can be re-identified after each batch with data from the most recent batch added to the historical process data. Furthermore, the control trajectory and quality variable trajectory from the most recent batch can be used as the reference trajectories.

In many batch processes, the control policies are typically determined to optimise the product quality at the end of a batch. Therefore, the control actions during different stages of a batch could be correlated. In such cases, appropriate linearised model may not be obtained from MLR. To overcome the colinearity in the regression variables, principal component regression (PCR) or partial least squares (PLS) regression [8] can be used to obtain the linearised models.

3.2 Batch to Batch Iterative Learning Control

The batch to batch iterative learning control strategy was developed in [6] and is briefly introduced here. As batch process dynamics are non-linear and the perturbation model is linearised around the nominal operation trajectories of a batch process, offsets always occur due to modelling errors and unmeasured disturbances. The perturbation model predictions of the current batch run can be corrected by adding model prediction residuals of previous batch runs.

The prediction of perturbation model is defined as

$$\hat{\overline{\mathbf{Y}}}_k = \hat{\mathbf{G}}_s \overline{\mathbf{U}}_k \tag{8}$$

and the absolute model prediction is defined as

$$\hat{\mathbf{Y}}_k = \mathbf{Y}_s + \hat{\overline{\mathbf{Y}}}_k = \mathbf{Y}_s + \hat{\mathbf{G}}_s \overline{\mathbf{U}}_k \tag{9}$$

After completion of the k^{th} batch run, prediction errors between off-line measured or analysed product qualities and their model predictions can be calculated as

$$\varepsilon_k = \mathbf{Y}_k - \hat{\mathbf{Y}}_k = \overline{\mathbf{Y}}_k - \hat{\overline{\mathbf{Y}}}_k \tag{10}$$

Based on the prediction errors of the k^{th} batch run, the modified prediction of perturbation model in the $(k+1)^{\text{th}}$ batch run is obtained as

$$\tilde{\overline{\mathbf{Y}}}_{k+1} = \hat{\overline{\mathbf{Y}}}_{k+1} + \varepsilon_k \tag{11}$$

The absolute modified model prediction is defined as

$$\tilde{\mathbf{Y}}_{k+1} = \hat{\mathbf{Y}}_{k+1} + \varepsilon_k = \mathbf{Y}_s + \hat{\overline{\mathbf{Y}}}_{k+1} + \varepsilon_k \tag{12}$$

The modified prediction error is defined as

$$\tilde{\varepsilon}_{k+1} = \mathbf{Y}_{k+1} - \tilde{\mathbf{Y}}_{k+1} = \overline{\mathbf{Y}}_{k+1} - \tilde{\overline{\mathbf{Y}}}_{k+1} \tag{13}$$

From the definitions in Eq(10) and Eq(11), we have

$$\tilde{\varepsilon}_{k+1} = \varepsilon_{k+1} - \varepsilon_k \tag{14}$$

Here we assume that the prediction error of the perturbation model is bounded by a certain small positive constant B_m such that

$$|\varepsilon_k| < B_m \tag{15}$$

The prediction error bound B_m is a measure to represent the deviation of $\hat{\overline{\mathbf{Y}}}_k$ from $\overline{\mathbf{Y}}_k$ or $\hat{\mathbf{Y}}_k$ from $\mathbf{Y}_k$. The higher the value of B_m is, the poorer the identified model is. The modified prediction error is bounded by $2B_m$ as follows

$$|\tilde{\varepsilon}_k| < |\varepsilon_k| + |\varepsilon_{k-1}| < 2B_m \tag{16}$$

The tracking errors of process and perturbation model are respectively defined as

$$\mathbf{e}_k = \mathbf{Y}_d - \mathbf{Y}_k = \overline{\mathbf{Y}}_d - \overline{\mathbf{Y}}_k \tag{17}$$

$$\hat{\mathbf{e}}_k = \mathbf{Y}_d - \hat{\mathbf{Y}}_k = \overline{\mathbf{Y}}_d - \hat{\overline{\mathbf{Y}}}_k \tag{18}$$

where $\overline{\mathbf{Y}}_d$ is the deviated desired trajectory and defined as

$$\overline{\mathbf{Y}}_d = \mathbf{Y}_d - \mathbf{Y}_s \tag{19}$$

The tracking errors of modified prediction of perturbation model is defined as

$$\tilde{\mathbf{e}}_k = \mathbf{Y}_d - \tilde{\mathbf{Y}}_k = \overline{\mathbf{Y}}_d - \tilde{\overline{\mathbf{Y}}}_k \tag{20}$$

From the definitions in Eq(10), Eq(17) and Eq(20), the following relationship among these three tracking errors can be obtained

$$\varepsilon_k = \hat{\mathbf{e}}_k - \mathbf{e}_k \tag{21}$$

$$\tilde{\mathbf{e}}_k = \hat{\mathbf{e}}_k - \boldsymbol{\varepsilon}_{k-1} \tag{22}$$

From Eq(18) and Eq(21), an iterative relationship for $\hat{\mathbf{e}}_k$ along the batch index k can be obtained as

$$\hat{\mathbf{e}}_{k+1} = \hat{\mathbf{e}}_k - \hat{\mathbf{G}}_s \Delta\overline{\mathbf{U}}_{k+1} \tag{23}$$

where $\Delta\overline{\mathbf{U}}_{k+1}$ is defined as

$$\Delta\overline{\mathbf{U}}_{k+1} = \overline{\mathbf{U}}_{k+1} - \overline{\mathbf{U}}_k \tag{24}$$

From the definition of perturbation variables, we can have

$$\Delta\overline{\mathbf{U}}_{k+1} = \overline{\mathbf{U}}_{k+1} - \overline{\mathbf{U}}_k = \mathbf{U}_{k+1} - \mathbf{U}_k \tag{25}$$

Substitute Eq(21) and Eq(23) to Eq(22), we have

$$\tilde{\mathbf{e}}_{k+1} = \hat{\mathbf{e}}_{k+1} - (\hat{\mathbf{e}}_k - \mathbf{e}_k) = \mathbf{e}_k - \hat{\mathbf{G}}_s \Delta\overline{\mathbf{U}}_{k+1} \tag{26}$$

On the other hand, Eq(21) can be rewritten as

$$\mathbf{e}_k = \hat{\mathbf{e}}_k - \varepsilon_k \tag{27}$$

From Eq(27) and Eq(23), an iterative relationship for $\mathbf{e}_k$ along the batch index k can also be obtained as

$$\mathbf{e}_{k+1} = \mathbf{e}_k - \hat{\mathbf{G}}_s \Delta\overline{\mathbf{U}}_{k+1} - \tilde{\varepsilon}_{k+1} \tag{28}$$

Given the error transition model in the form of Eq(26) and Eq(28), the objective of ILC is to design a learning algorithm to manipulate the control policy so that the

product qualities follow the specific desired reference trajectories. It is required that the learning algorithm has the following property [9]

$$\lim_{k\to\infty}\|\mathbf{e}_k\|_{\mathbf{Q}}^2 \to \min_{\mathbf{U}}\|\mathbf{e}\|_{\mathbf{Q}}^2 \tag{29}$$

By the certainty-equivalence principle [10], we consider solving the following quadratic objective function based on the modified prediction errors upon the completion of the k^{th} batch run to update the input trajectory for the $(k+1)^{\text{th}}$ batch run

$$\lim_{k\to\infty}\|\mathbf{e}_k\|_{\mathbf{Q}}^2 \to \min_{\mathbf{U}}\|\mathbf{e}\|_{\mathbf{Q}}^2 \tag{30}$$

where $\mathbf{Q}$ and $\mathbf{R}$ are positive definitive matrices. The objective function, Eq(30), has a penalty term on the input change $\Delta\overline{\mathbf{U}}_{k+1}$ between two adjacent batch runs, the algorithm has an integral action with respect to the batch index k [6]. The weighting matrices $\mathbf{Q}$ and $\mathbf{R}$ should be selected carefully. A larger weight on the input change will lead to more conservative adjustments and slower convergence. There are also other variants of the objective function. For example, the weighting matrices $\mathbf{Q}$ and $\mathbf{R}$ may be set as $\mathbf{Q}$ = diag{$Q(1)$, $Q(2)$, …, $Q(N)$}, $\mathbf{R}$ = diag{$R(0)$, $R(1)$, …, $R(N\text{-}1)$}, where $Q(i)$ and $R(j)$ increase with respect to the time intervals t in proportion to its effect of the final product quality. For the sake of simplicity, $\mathbf{Q}$ and $\mathbf{R}$ are selected in this study as $\mathbf{Q}=\lambda_q\mathbf{I}_N$ and $\mathbf{R}=\lambda_r\mathbf{I}_N$.

By finding the partial derivative of the quadratic objective function Eq(30) with respect to the input change $\Delta\overline{\mathbf{U}}_{k+1}$ and through straightforward manipulation, the following ILC law can be obtained

$$\Delta\overline{\mathbf{U}}_{k+1} = \hat{\mathbf{K}}\mathbf{e}_k \tag{31}$$

where $\hat{\mathbf{K}}$ is defined as the learning rate

$$\hat{\mathbf{K}} = [\hat{\mathbf{G}}_s^T\mathbf{Q}\hat{\mathbf{G}}_s + \mathbf{R}]^{-1}\hat{\mathbf{G}}_s^T\mathbf{Q} \tag{32}$$

From Eq(25) and Eq(31), the ILC law for the control trajectory can be written as

$$\mathbf{U}_{k+1} = \mathbf{U}_k + \hat{\mathbf{K}}\mathbf{e}_k \tag{33}$$

4 Simulation Results

In this study, only the final biomass concentration is considered as the controlled variable. The desired biomass concentration is selected as a constant value of 70 g/l. The iterative learning control parameters are selected as: $\mathbf{R}$ = 0.95$\mathbf{I}$, $\mathbf{Q}$ = 600. Batch to batch ILC was run for 60 batches. From the 40^{th} batch onward, the initial biomass concentration was changed to 11 g/l from its nominal value of 15 g/l to represent the presence of a disturbance. This could be due to the change in raw materials. Fig. 1 and Fig. 2 show, respectively, the final biomass concentrations and the control policies at different batches. It can be seen from Fig. 1 and that the final biomass concentrations

approached the desired one from batch to batch. The final biomass concentration dropped at the 40th batch due to the presence of disturbance. However, with batch-to-batch ILC, the final biomass concentrations were improved from batch to batch and approached the desired value from the 41st batch onward. This demonstrates that ILC is effective in addressing model plant mismatches and the presence of unknown disturbances.

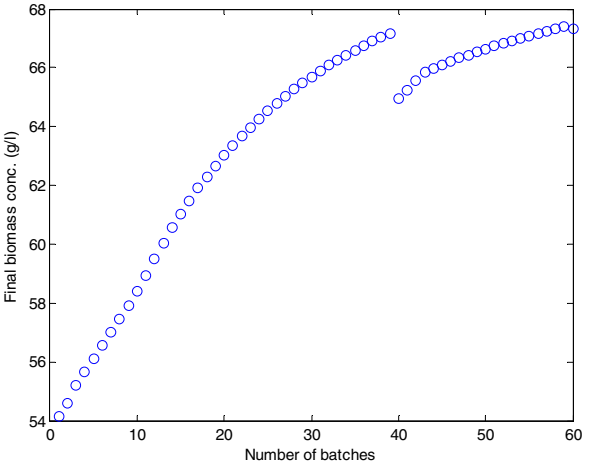


Fig. 1. RMSE at different batches

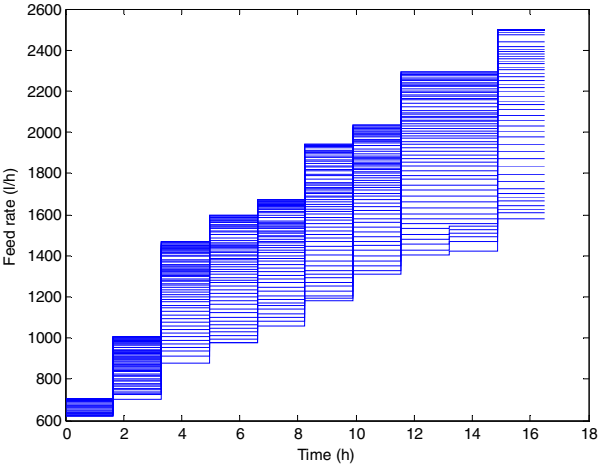


Fig. 2. Control Policies at different batches

5 Conclusions

A batch to batch iterative learning control method based on batch wise updated linearised models is developed. Through batch wise updating of the linearised model, the model can adapt to process variations and is more appropriate for the current operating condition. Model based ILC is derived through solving an optimisation problem. Application to a fed-batch fermentation process demonstrates the effectiveness of the proposed method.

Acknowledgments. The work is supported by the EU through the project iREMO – intelligent reactive polymer composite moulding (grant No. NMP2-SL-2009-228662).

References

1. Zhang, J.: A reliable neural network model based optimal control strategy for a batch polymerisation reactor. Ind. Eng. Chem. Res. 43, 1030–1038 (2004)
2. Gao, F., Yang, Y., Shao, C.: Robust iterative learning control with applications to injection molding process. Chemical Engineering Science 56, 7025–7034 (2001)
3. Lee, J.H., Lee, K.S., Kim, W.C.: Model-based iterative learning control with a quadratic criterion for time-varying linear systems. Automatica 36, 641–657 (2000)
4. Zhang, J.: A neural network based strategy for the integrated batch-to-batch control and within batch control of batch processes. Transactions of the Institute of Measurement and Control 27, 391–410 (2005)
5. Xiong, Z., Zhang, J.: Batch-to-batch iterative optimisation control based on recurrent neural network models. Journal of Process Control 15, 11–21 (2005)
6. Xiong, Z., Zhang, J.: Product quality trajectory tracking in batch processes using iterative learning control based on time-varying perturbation models. Ind. Eng. Chem. Res. 42, 6802–6814 (2003)
7. Yüzgeç, U., Türker, M., Hocalar, A.: On-line evolutionary optimization of an industrial fed-batch yeast fermentation process. ISA Transactions 48, 79–92 (2009)
8. Geladi, P., Kowalski, B.R.: Partial least squares regression: a tutorial. Analytica Chimica Acta 185, 1–17 (1986)
9. Lee, J.H., Lee, K.S., Kim, W.C.: Model-based iterative learning control with a quadratic criterion for time-varying linear systems. Automatica 36, 641–657 (2000)
10. Lee, K.S., Lee, J.H.: Model predictive control for nonlinear batch processes with asymptotically perfect tracking. Computers Chem. Engng. 21, 873–879 (1997)

An Innovative Time Synchronization Algorithm for Wireless HART Networks

Li Dongdong, Zhang Sheng, Wen Houming, and Lin Xiaokang

Department of Electronic Engineering, Tsinghua University,
100084, Beijing, China
lidongdonghaoyun@163.com

Abstract. WirelessHART is widely used in the field of industrial process control. The communication between the devices is performed using Time Division Multiple Access (TDMA) with time slots of 10ms. In addition, a series of actions must be completed in the slot unit of 10 ms, including the data reception, transmission and acknowledge, etc. Time synchronization is the most important premise to ensure the communication secure, reliable and conflict-free.

Keywords: WirelessHART, time synchronization, Kalman filtering, TDMA.

1 Introduction

In September 2007, HART Foundation released WirelessHART protocol, making it the first open international standard to focus on the field of process control and famous for its features of security, simplicity, reliability.

WirelessHART is a complete wireless mesh networking protocol mostly based on the IEEE 802.15.4-2006 DSSS physical layer standard, operating in the license-free ISM 2.4GHz band with a data rate of up to 250 kbps. The data link layer defines a strict 10ms timeslot and utilizes TDMA and FHSS technology to provide collision free and deterministic communications [2]. The network layer employs mesh topology structure, and support graph routing mechanism to provide redundant paths which allows messages to be routed around physical obstacles, broken links, and interference. Its transport layer is responsible for establishing secure bi-direction transport pipe which can improve the end-to-end data transmission reliability. Both the MAC layer and network layer provide security services, using the CCM* mode together with AES-128 as the underlying block cipher to generate and compare the MIC. All these features make WirelessHART a simple, secure and reliable standard for wireless sensor network.

WirelessHART data link layer supports multiple superframes, which are constructed by a fixed number of slots and form a network cycle with a fixed repetition rate. A series of actions must be completed in the slot unit of 10 ms, including the data reception, transmission and acknowledge, etc. Fig.1shows the details of a WirelessHART slot. For successful and efficient TDMA communications, synchronization of clocks between devices in the network is critical.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 261–267.
springerlink.com

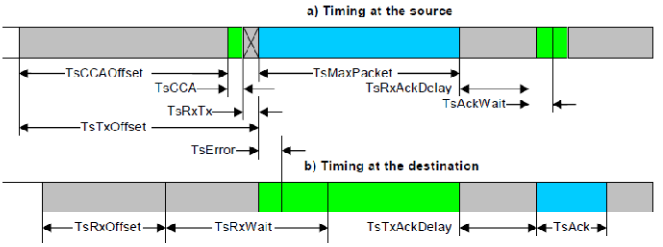


Fig. 1. Slot timing details [1]

A WirelessHART PPDU contains a preamble of 4 bytes, a SFD of 1 byte, and a PHR of 1 byte. According to the data rate of 250 kbps, the transmission of preamble, SFD, PHR needs totally 192 microseconds. When the source sent the PHR, the receiving node must be in the receiving window (TsRxWait). So, it must satisfy the formulas of 2-1 and 2-2:

$$\Delta t + \text{TsRxOffset} \le \text{TsTxOffset} \tag{2-1}$$

$$\Delta t + \text{TsTxOffset} + 192 \le \text{TsRxOffset} + \text{TsRxWait} \tag{2-2}$$

where Δt is the time deviation between two neighbor nodes. According to 2-1 and 2-2, we calculate that the smallest time deviation is less than1000 microseconds.

The rest of this paper is organized as follows: A WirelessHART time synchronization algorithm will be present in section 2. In section 3, an improved synchronization algorithm will be described, introducing the Kalman filtering technique to increase the synchronization accuracy. Section 4 will list the results of a software simulation, and give the performance evaluation. Also, the results of a hardware simulation will be described in section 5. At last, we will draw a conclusion for this paper in section 6.

2 WirelessHART Time Synchronization Algorithm

We divide WirelessHART time algorithm into active time synchronization before joining the network and passive time synchronization after joining the network.

2.1 WirelessHART Passive Time Synchronization

Before joining the network, a WirelessHART node should listen continuously until receive some DLPDUs. It executes the passive time synchronization algorithm, using the timestamp of reception, as follows:

a) Let n=0, $T_{current\_slot} = 0$, Δt = TsTxOffset, where n denotes the total number of receiving the DLPDU, and $T_{current\_slot}$ denotes the start time of current slot, and Δt is the value of time between DLPDU reception time and slot starting time.

b) When receives the fist DLPDU, the source node will record the reception time as t_1. As a result, the next slot start time is $T_{next\_slot} = t_1 + (10\text{ms} - \text{TsTxOffset})$, and the start time of current slot is $T_{current\_slot} = t_1 - \text{TsTxOffset}$. T_{next_slot} and $T_{current\_slot}$ need

updating during every slot whether it receives a DLPDU or not. But, Δt is not updated until a DLPDU was received during current slot.

c) When the receiving node receives the fist DLPDU, it will record the reception time as t_1. So, n = 2, $T_{current\_slot} = t_1 + 10ms - TsTxOffset$, $\triangle t = [(t_2 - T_{current\_slot}) + \Delta t]/2$, $T_{next\_slot} = T_{current\_slot} + \Delta t + (10ms - TsTxOffset)$.

d) If the DLPDU received is a Advertise-DLPDU, and the Network-ID matches the true value, we will get ASN from the payload of the DLPDU, and synchronize the slot of the receiving node. So, $ASN_{node} = ASN$.

e) Let t_n denote the reception time of the n-st DLPDU, then $T_{current\_slot} = t_{n-1} + 10ms - TsTxOffset$, $\triangle t = [(t_n - T_{current\_slot}) + \Delta t*(n-1)]/n$, $T_{next\_slot} = T_{current\_slot} + \Delta t + (10ms - TsTxOffset)$.

2.2 WirelessHART Active Time Synchronization

A WirelessHART device will synchronize its clocks through periodic transmission of Keep-Alive DLPDU to its neighbors as soon as joins the network. The sink node loads the time deviation into the payload of ACK DLPDU and sends it to the source node. After receiving the ACK DLPDU, the source node will synchronize its clock if the sink node is a clock source. The process of active synchronization is as follows:

a) When the synchronizing period expires, node B sends a Keep-Alive DLPDU to node A at t_1. The slot start time of node B is t_{0B}, as that of node A is t_{0A}.

b) When receives the DLPDU, node A records the reception time as t_2. Then, the time deviation of node B is $TsError = t_2 - t_{expect}$, where $t_{expect} = t_{0A} + TsTxOffset$, is the expect time of receiving the DLPDU. At last, node A loads TsError into the payload of ACK DLPDU and sends it to node B.

c) When receives the ACK DLPDU from node A at t_4, node B gets TsError and synchronizes its clock if node A is a source clock, and the start time of next slot is $t_{next\_slot} = t_4 + (10ms - t_{ACK}) + TsError + t_4 - (t_{0B} + t_{ACK})$.

2.2 WirelessHART Network-Wide Time Synchronization

Either the active synchronization or the passive synchronization is time synchronization of neighbors. However, network-wide time synchronization consists of three phases: level discovery, selecting source clock, time synchronization [3].

2.2.1 Level Discovery

The network manager creates a time topology in corroding to the relationship among nodes, so the location and the level of every node can be known. The process of level discovery is as follows:

a) Initialize the network manager and the gateway, and create a network. Let level_0 denote the level of the gateway. Similarly, let location_0_1 denote the location of the gateway, where "0" is the level and "1" is the number in this level.

b) The fist node joins the network through the gateway with its level being level_1 and location being location_1_1.

c) If the parent node of the second node is located in the level_0, the level of the second node is level_1, and the location is location_1_2. If the parent node of the

second node is located in the level_1, the level of the second node is level_2, and the location is location_2_1.

d) Similarly, if the parent node of the n-st node is located in the location_p_q, we record the level of it as level_p+1. If there have been k nodes located in the level_p+1, we recognize the location of the n-st node as location_p+1_k+1.

2.2.2 Selecting Clock Source

Because the hierarchical time topology has been established, we can select clock source for every node from nodes of upper level, making the least hops to the gateway and synchronization error smaller. One node can select not only its parent but also its uncles as clock sources.

a) Let the gateway act as the clock root, which can be equipped with a GPS module.

b) A node can be selected as a clock source of its baby and nephew. However, the nodes cannot act as a clock source of its parent, uncles or brothers.

c) If a node finds a new neighbor on its discovery link, it will report the health condition of its new neighbor to the network manager. If the new neighbor is its uncle, the network manager will appoint the new neighbor as its clock source.

d) If someone leaves the network, the network manager will delete everything in connection with it and update the clock topology.

2.2.3 Time Synchronization

After selecting their clock sources, WirelessHART nodes can execute the efficient time synchronization. The gateway acts as the clock root of the whole network. The nodes of lower level actively synchronize towards the nodes of upper level periodically until the nodes of lebel_1 send a synchronization request to the gateway.

3 Improved Time Synchronization Algorithm Based on Kalman Filtering

Kalman filter is an optimal information processor described by a series of recursive mathematical formulas [4]. An efficient method is provided by Kalman filtering algorithm to estimate the process state, making the mean square error the smallest.

We minimize the transmission delay and make proper time compensation to decrease the clock error. When node B sends a Keep-Alive DLPDU to node A, we assume that the reception time expected is that of SFD, that is, $t_{expect} = t_{0A} + TsTxOffset + 160us$, where t_{0A} is the start time of current slot, and 160us is the time compensated for PHR. Node A records the reception time in the physical layer, so we can ignore the sending delay, accessing delay, transmitting delay, receiving delay and processing delay. The transporting delay can also be ignored because the transporting speed of electromagnetic wave is equal to that of the light.

The improved time synchronization algorithm doesn't adjust the local clock according to the time deviation measured immediately, but to the most optimal value estimated of the time deviation. According to the principle of Kalman filter, we create a time synchronization system model as follows:

$$s(k) = s(k-1) + w(k-1) \tag{3-1}$$

$$y(k) = s(k) + v(k) \tag{3-2}$$

Where, formula 3-10 is the signal model, and formula 3-11 is the state model. With k denoting the number of synchronize, s(k) represents the time deviation of the k-st synchronization, and y(k) is the measurement of time deviation. The process noise w(k-1) drives the dynamic system, whose standard deviation is d_w, and the observation noise is given by v(k), whose standard deviation is d_v. According to the principle of the smallest MSE, we get the most optimal estimation from formula 3-12:

$$x(k)= x(k-1) + b(k)*[y(k) – x(k-1)] \tag{3-3}$$

where Kalman filter gain is:

$$b(k) = p_1(k)*[p_1(k) + d_v^2]^{-1} \tag{3-4}$$

$$p_1(k) = p(k-1) + d_w^2, \tag{3-5}$$

and the covariance is:

$$p(k) = p_1(k-1) – b(k)*p_1(k) \tag{3-6}$$

The measurement of time synchronization, TsError, in the payload of ACK DLPDU is the equivalent of y(k) in formula 3-3. We can get the most optimal estimation of time deviation through formula 3-4, and we use the estimation x(k) to synchronize the clock.

4 Simulation and Evaluation

In this section, we give a simulation for our algorithm using MATLAB software. In our simulation, we get 200 time deviation randomly whose mean is 30. We assume that the standard deviation of process noise is $d_v=10^{-6}$, and that of measurement noise is $d_w=10^{-1}$. The initial value is set as x(1)=50, p(1)=2. We should be careful that the initial covariance cannot be zero. Considering the Kalman filtering technique is a recursive scheme, the estimation x(k) converges to the actual value after a number of messages. Fig.2 shows that changes of the most optimal estimation x(k) are consistent with that of actual time deviation s(k), but it is different for the measurement y(k). We also can see that the estimation can quickly converge to the actual time deviation after 6 times of synchronization. So, the improved time synchronization algorithm based on the Kalman filtering technique can increase the synchronization accuracy and decrease the synchronization error for a single hop, making the network- wide error accumulated the least. It tells us the most optimal estimation is more creditable.

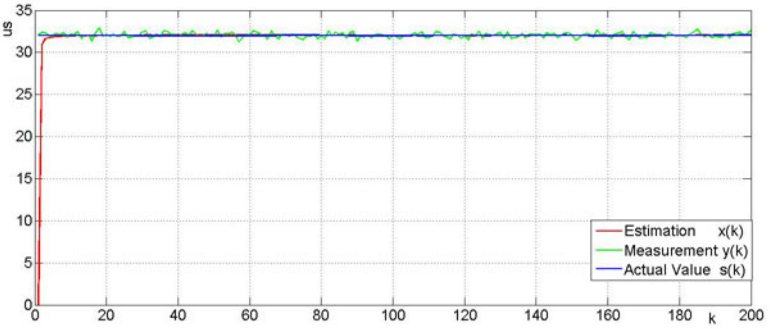


Fig. 2. The trend of measurement, estimation and actual value

In the network-wide synchronization, the local clock error of the node which is the farthest from the gateway may be accumulated continuously, leading to disconnect from the network at last. Its clock error accumulated is

$$\mathrm{TSERROR} = \sum_{n=1}^{N} \mathrm{TsError_n}, \tag{4-1}$$

where N denotes the maximum of the hop, and $\mathrm{TsError_n}$ is the clock error of node n. On one hand, this algorithm creates a hierarchical clock topology and selects the clock sources in the principle of the least hop to decrease the time error accumulated. On the other hand, it uses the Kalman filtering technique to improve the synchronization accuracy, making the time error accumulated the least. At the same time, this algorithm expands the network topology and increases the capacity of the network.

5 Implementation and Analysis

We have implemented the time synchronization algorithm with the processer STM32F103 and the transceiver CC2520. Five nodes were arranged into a queue. A 16-bit timer was used as a counter, which reserved every 10ms, and the measurement was calculated in the interrupt program. We found that the synchronization accuracy was only 163us and was accumulated continuously without this time synchronization algorithm. Conversely, the synchronization accuracy was stable at about 32us between two neighbors, and the error accumulated was very small, only 41us. This experiment proves that the time synchronization algorithm based on the Kalman filtering technique can improve the synchronization accuracy and decrease the error accumulated. The experiment had not reached the expectant effect, because the delay of SPI operation was not compensated, etc.

6 Conclusion and Future Work

In this paper, we propose a time synchronization algorithm for WirelessHART network which can decrease the error accumulated and improve the synchronization accuracy. Nevertheless, the accuracy can be improved in some degree. In further work, we will focus on how to maximize the energy efficiency and to improve the robustness of network.

References

1. Industrial communication networks – Wireless communication network and communication profiles –WirelessHART™ IEC 62591
2. Lennvall, T., Svensson, S., Hekland, F.: A comparission of WirelessHART and ZigBee for industrial applications. In: The Intl. Workshop on Factory Communication Systems (2008)
3. Ganeriwal, S., Kumar, R., Adlakha, S., Srivastava, M.: Network-wide Time Synchronization in Sensor Networks. In: UCLA Technical Report, NESL01-01-2003
4. Jian, T., Haoshan, S., Ronghui, H., XueSong, L.: Time synchronization algorithm in sensor networks. In: Communication Technology, ICCT 2006, pp.1–4 (November 2006)

The Design and Application of Boost DC Converter Power Sliding Mode Controller

Cao Long-han, Tian Li, Liu Lu, Wu Zheng-yi, and Zhang Ying-chao

Key Laboratory of Control Engineering,
Chongqing Communication Institute, Chongqing, 400035, China
ta_fengle@126.com

Abstract. A strategy which connects the pulse width modulation with sliding mode control is put forward and a boost DC converter sliding mode controller in continuous current mode is designed, using DC-DC conversion technology that changes the single low-voltage lithium battery of 3.2V into the working power station of 12V. Building the controller simulation model in Matlab / Simulink environment, simulating controller performance in the case of changing the load resistance and input voltage; Simulation results validate its performance in the large-scale disturbance, and show its good dynamic performance, robustness, and smaller steady-state error.

Keywords: DC-DC converter, Sliding mode control, Pulse Width Modulation, Continuous Current Mode, Robustness.

1 Introduction

Portable radio is the key to achieve mobile communication. Lithium battery single working voltage is up to 3.2V or 3.6V. Traditional wireless radio is generally used to adopt power supply of multiple class single cells in series into battery groups of 12Vor24V. This approach is easy to realize; But in long-term use, because of the material of each single cell aging at different rates, this difference will lead to smaller overall capacity of the battery and reduce battery life. More that when charging the battery pack, Cause each single cell over charge or under-charging phenomenon in the battery pack, which will lead the capacity drop, cause the battery early damage.

In view of this, a boost DC converter sliding mode controller in continuous current mode is designed, using DC-DC conversion technology that changes the single low-voltage lithium battery into the working power station, and overcoming the shortcoming of life and capacity reducing[1-3].

2 DC Converter Topology Selection

The power supply of a portable radio is 12Vand peak current is 2A, belongs to A small power transformation range. This paper uses boost DC converter to transform single low-power lithium battery output voltage of 3.2V into the radio working voltage of 12V.According to performance analysis, choosing Boost-topology, shown in figure1;

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 269–276.
springerlink.com

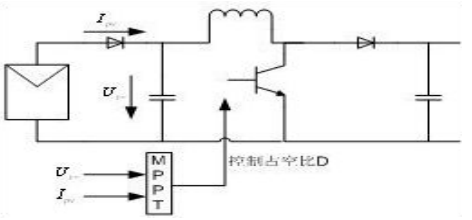


Fig. 1. Boost topology

It is specially used for the boost transform, consistent with the DC transformation unit demand[4-6].

3 The Basic Principles of Sliding Mode Control

According to the definition of sliding mode controller, design sliding mode controller mainly following the two design process below:

(1)Select the appropriate switching function $S(x)[0,1]$;

(2) Select the appropriate control law $u^{\pm}(x)$;

Also meet the following three objectives:

1) Switching surface meets the reaching condition. The state of the system tracks must be tend to switching surface movement, and in limited time can reach. The condition can be simplified as $S\dot{S}=0$;

2) The switching surface should exist, to ensure state points can turn into the sliding mode;

3) Sliding movement should be able to asymptotic stability to achieve a dynamic quality control system requirements.

4 PWM Sliding Model Variable Structure Control Strategy Based on Boost Converter

Based on the idea of PWM modulation technique, sliding mode variable structure control based on PWM can get a fixed switching frequency. The advantage of this strategy is no additional auxiliary circuit, also can't weaken its dynamic characteristics and robustness.

4.1 The Design of PWM Sliding Mode Variable Structure Controller

1) selection of the switching surface function in continuous mode

Only when the system state little exercise in sliding mode stage, the system can manifest the advantages of sliding mode variable structure control.

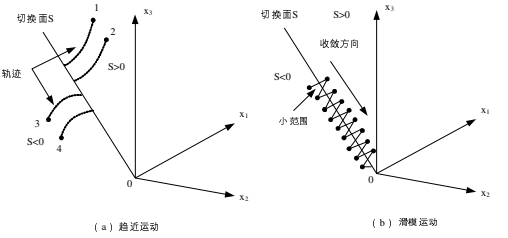


Fig. 2. State point trajectory's behavior in continuous current mode under sliding mode control

When boost convertor work in inductive current converter mode, all state points trajectory must follow as shown in figure 2 showing trajectory. As the sliding mode control of the control signal the end to act on the power switch; Under the PWM power switch, only has two states of off and turn on, so based on continuous mode variable structure model, taking the switching surface function:

$$S = a_1 x_1 + a_2 x_2 + a_3 x_3 = J^T x \tag{1}$$

Where: $x = \begin{bmatrix} x_1 = v_{ref} - \beta v_o \\ x_2 = \dfrac{\beta v_o}{R_L C} + \int \dfrac{\beta (v_o - v_i)\overline{u}}{LC} dt \\ x_3 = \int (v_{ref} - \beta v_0) dt \end{bmatrix}$; $\overline{u} = 1 - u$ is logic reverse for the control law; L, R_L, C represent the inductor, the equivalent resistance and capacitance values; β is feedback network divider ratio; $J^T = [a_1, a_2, a_3]$, a_1, a_2, a_3 is switching surface coefficients.

2) conditions of reaching to surface under continuous mode

To ensure the switching surface up, switch up to the surface must meet the conditions:

$$\lim_{s \to 0} S\dot{S} < 0 \tag{2}$$

Function (2) of the switching surface on both sides of the derivation:

$$\dot{S} = J^T \dot{x} \tag{3}$$

after the derivative of X to the standard state space form into the following formula to get:

$$\dot{x} = Ax + B\overline{u} + D \tag{4}$$

Where: $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -\dfrac{1}{R_L C} & 0 \\ 1 & 0 & 0 \end{bmatrix}$; $B = \begin{bmatrix} 0 \\ \dfrac{\beta v_o}{LC} - \dfrac{\beta v_i}{LC} \\ 0 \end{bmatrix}$; D=0.

Because $v_i - v_o < 0$,standard form of switching surface for reaching as the relationship:

$$0 < \beta L\left(\frac{a_1}{a_2} - \frac{1}{R_L C}\right) i_c - \frac{a_3}{a_2} LC(v_{ref} - \beta v_o) < \beta(v_0 - v_i) \tag{5}$$

3) stability analysis of continuous mode switching surface

Ackermann formula used to re-define the switching surface coefficient selection method, which can automatically switch to meet the existence of a stable surface conditions.

Sliding mode controller, the final point of the system state can expect in the switching surface for sliding movement, that's $S=0$. Switching surface variable x_1、x_2、x_3 into a function, written in typical second-order system, the standard form:

$$\ddot{x} + 2\zeta\omega_n \dot{x} + \omega_n^2 x = 0 \tag{6}$$

Where: $\omega_n = \sqrt{a_3 / a_2}$ is the angular frequency of undamped oscillation;

$\zeta = a_1 / (2\sqrt{a_2 a_3})$ System damping;

The controller operating bandwidth is $f_{BW} = \omega_n / 2\pi$, so deriving to get:

$$\frac{a_3}{a_2} = \frac{1}{4\zeta^2}(\frac{a_1}{a_2})^2 \tag{7}$$

Classical control theory that the second-order system response characteristics depending on the system damping ζ , according to the size of damping, system can be divided into three states: underdamped state, critical damping condition and overdamped state. But this article defaults the designed variable structure controller in the underdamped condition, so only analyse the conditions of switching surface stability in this case.

When system works in underdamped state($0 < \zeta < 1$), in order to ensure the state points access to Boost converter sliding mode, the output error should be:

$$x_1(t) = A_1 \frac{e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}} \cos(\omega_n \sqrt{1-\zeta^2}\, t - \psi), t \geq 0 \tag{8}$$

Where: A_1 determined by the system initial state; phase angle $\psi = \arctan\frac{\zeta}{\sqrt{1-\zeta^2}}$; transient component time constant decay rate depends on the countdown: $\frac{1}{\tau} = \zeta\omega_n = \frac{a_1}{2a_2}$.

Assuming the system response time $T_s = 5\tau$,to ensure stable switching surface must satisfy the following relationship:

$$\frac{a_3}{a_2} = \frac{25}{\zeta^2 T_s^2} \tag{9}$$

Where: $\zeta = \sqrt{\frac{[In(\frac{\sigma_P}{20n})]^2}{[\pi^2 + In(\frac{\sigma_P}{20n})]^2}}$; σ_p is the maximum overshoot of system response.

Based on the selection process of parameters of Ackermann formula, switching surface coefficients of sliding mode control depends only on the operating bandwidth of the controller and the expected response time. If there will be a zero problems of right-half plane(Right Half Plane Zero, RHPZ) in boost converter, commonly used method is to significantly reduce the operating bandwidth value of the controller to solve the added phase lag generated by RHPZ.

4) the equivalent equation under continuous mode

In order to introduce sliding mode control to PWM modulation techniques, need to converts control signal in the sliding mode control into duty cycle modulation signal in PWM modulation. Because of the invariance $\dot{S} = J^T Ax + J^T B\bar{u}_{eq} = 0$,can derive the equivalent control:

$$\bar{u}_{eq} = -[J^T B]^{-1} J^T Ax \tag{10}$$

Substituted into the coefficient matrix A ,Control matrix B, combin $0 < \bar{u}_{eq} < 1$, can get:

$$0 < u^*_{eq} = -\beta L(\frac{a_1}{a_2} - \frac{1}{R_L C})i_c + \frac{a_3}{a_2} LC(v_{ref} - \beta v_o) + \beta(v_o - v_i) < \beta(v_o - v_i) \tag{11}$$

Where: u^*_{eq} is duty cycle in PWM; D_{ccm} is equivalent control signal;

Control signal in the PWM modulation is compared with a fixed frequency ramp signal, in order to modulate duty cycle,so get:

$$0 < D_{ccm} = \frac{v_c}{\hat{v}_{ramp}} < 1 \tag{12}$$

Obtain the equivalent control law as follows:

$$v_c = -K_{p1} i_c + K_{p2}(v_{ref} - \beta v_o) + \beta(v_o - v_i) \tag{13}$$

$$\hat{v}_{ramp} = \beta(v_o - v_i) \tag{14}$$

Where: $K_{p1} = \beta L(\frac{a_1}{a_2} - \frac{1}{R_L C})$; $K_{p2} = LC\frac{a_3}{a_2}$;

v_c is the control signal; $\hat{v}_{ramp}$ is the peak ramp signal.

Formula (13) and Formula (14) are continuous mode of the equivalent sliding mode control equations. According to the equivalent equation, then build a continuous mode Boost converter PWM modulation sliding mode controller, the controller structure shown in Fig3.

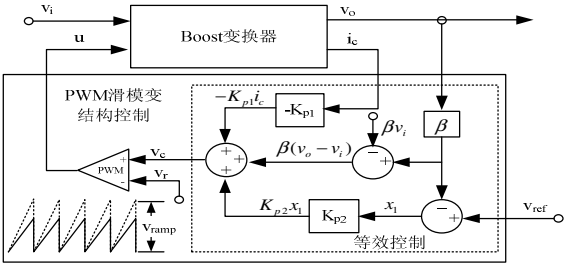


Fig. 3. Schematic diagrams of PWM-based sliding mode controlled Boost converter in CCM

Where: v_i is the input voltage; v_o is the output voltage; i_r is the load current; i_c is the capacitor current; v_r is the ramp signal; v_{ramp} is the peak ramp signal; u is the control pulse signal.

4.2 Simulate and Results Analysis

In order to verify the continuous mode PWM control sliding mode controller performance, based on sliding mode control in continuous mode the equivalent control formula (13), formula (14), in Matlab / Simulink environment to build the controller simulation model, Simulation time is set to $40ms$,step is set to $10ns$, the algorithm uses the default settings; Seting the damping coefficient $\zeta = 0.2$,bandwidth $\omega_n = 1.25krad/s$, the time constant $\tau = 0.8ms$, expected adjustment time $T_s = 2ms$; Output voltage reference value $v_{ref} = 2V$, then the ratio of the feedback divider network $\beta = v_{ref} / v_o = 1/6$. According to the conditions of stable switching surface, in less resistance state calculate the switching surface coefficients $a_1 / a_2 = 2500$, $a_3 / a_2 = 3906200$; equivalent controled coefficient $K_{p1} = 0.0486$, $K_{p2} = 1.769$.

Continuous mode Boost converter parameters: Input voltage (max) v_{imax}=3.2V, Input voltage (min) v_{imin}=2.7V; Capacitance C =300 μF ; inductance L =150 μH ; Load resistance (min) R_{min} =6 Ω ;Working frequency f =200 kHz ; Output voltage v_o =12 V. In order to verify the continuous mode PWM control sliding mode controller demonstrated the dynamic characteristics and robustness, simulating the load resistance and input voltage in the case of changes. Simulation experiments carried out as follows:when $t = 0.02s$,mutation load resistance from 6Ω to 12Ω .When $t = 0.03s$, mutation load resistance from 12Ω to 6Ω . Input voltage test using the same way,When $t = 0.02s$, The input voltage down from $3.2V$ to $3.0V$;When $t = 0.03s$, further down the input voltage from $3.0V$ to $2.8V$.

Fig4 shows the load disturbance corresponds output voltage waveform of Boost converter. the output voltage is not sensitive to disturbance, without significant fluctuations. The system adjusts fast. From the point of disturbance to the system enters a new steady state, adjust the time is 2ms, consistent with the expected response time, steady-state error control in less than 10mv, showing strong robustness.

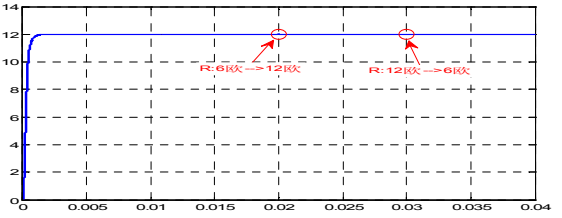


Fig. 4. Output voltage waveforms of Boost converter operating at step changes load resistances between 6Ω and 12Ω under PWM-based SMC in CCM

Fig5 shows the input disturbances corresponds Boost converter output voltage waveform. A wide range of input perturbation disturbance, the output voltage change is not obvious, steady-state error is small, showing good invariance of control system.

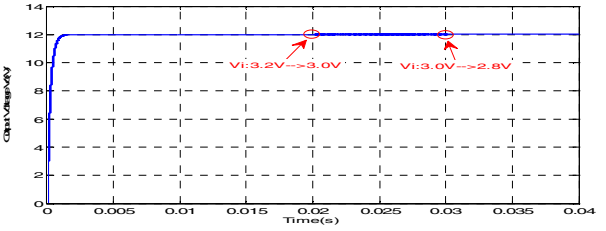


Fig. 5. Output voltage waveforms of Boost converter operating At step changes input voltage beween 3.2Vand 2.8Vunder PWM-based SMC in CCM

The above a simulation analysis for continuous mode PWM control sliding mode controller, when changing in input voltage and load resistance, it can maintain a fixed switching frequency. The output voltage are shown insensitivity to changes in the input voltage and load resistance, the controller showed excellent dynamic performance and strong robustness.

5 Conclusion

Boost converter is core unit of portable radio power supply based on the single lithium battery;The introduction of control technology to control is striving to obtain a more excellent dynamic performance and robustness. Control technology in the

sliding mode control has invariable and completely adaptive advantages to the parameter perturbation and external disturbance, in recent years having been greatly developed in various fields. Boost converter has the characteristics of variable structure, particularly suitable for the use of sliding mode control strategy. Article is committed to research the problems appearing in application-specific of the sliding mode control strategy in Boost converter. Based on theoretical analysis research, design a sliding mode controller and its control performance are verified through simulation experiments, showing its good dynamic performance, robustness, and strong practical value.

Acknowledgement. The results of this paper has obtained the supports of colleges and universities outstanding achievements conversion project in Chongqing (code: Kjzh10219) and the supports of scientific and technological project in Chongqing (code: CSTC-2011AC4046), to express my heartfelt thanks.

References

1. Line, J.L., Chen, S.J.: μ-based controller design for a DC-DC switching power converter with line and load variations. IEEE Industrial Electronic, Control and Instrumentation (2), 1029–1034 (1996)
2. Liu, J.-K.: Variable Structure Control Matlab Simulation. Tsinghua University Press (2005)
3. Kazimierczuk, M.K.: Transfer function of current modulator in PWM conveners with current mode control. IEEE Transactions on Fundamental Theory and Applications, Circuits and Systems 47(9), 1407–1412 (2000)
4. Ackermann, J., Utkin, V.: Sliding mode control design based on Ackermann's formula. IEEE Transactions on Power Electronics 43(2), 234–237 (1998)
5. Kwon, I.S., Cho, W.: Sliding mode control of a zero-current switching resonant converter. IEEE Electronics Letters 5(30), 381–382 (1994)
6. Krein, P.T., Papenfuss, S.: Equalization requirements for series VRLA batteries. In: The Sixteenth Annual Battery Conference on Applications and Advances, vol. 16(1), pp. 125–130 (2001)

Multi-Classification LSSVM Application in Fault Diagnosis of Wind Power Gearbox

Bin Jiao[1] and Zhixiang Xu[1,2]

[1] Electric Engineering School, Shanghai DianJi University,
No.690, Jiang Chuan Rd., Min Hang District, Shanghai 200240, China
jiaob@sdju.edu.cn
[2] College of Information Science and Engineering,
East China University of Science and Technology,
No.130, Mei Long Rd., Shanghai 200237, China
xzxkbjy@163.com

Abstract. For wind turbine gearbox fault diagnosis problem, we propose a multi-classification least squares support vector machines (MCLSSVM) model. According to failure mechanism and vibration characteristics of gearbox, it investigates some formulas of fault diagnosis. Through the combination of voting method and decision tree, it constructs the MCLSSVM decision-making structure, and then it is applied on the fault diagnosis of wind turbine gearbox. Tests show that MCLSSVM can be effectively used in the fault diagnosis of wind turbine gearbox. It solves the studying problem of small sample, and overcomes the shortcoming of artificial neural network (ANN) when it is used in fault diagnosis.

Keywords: wind turbines gearbox, fault diagnosis, MCLSSVM, artificial neural network.

1 Introduction

As the installation location and operating conditions of wind turbine gearbox, the relationship between faults and symptoms is not very clear. Then the fault diagnosis of wind turbine gearbox is a very complex issue, involving dynamic information processing, computers, artificial intelligence and many other fields of knowledge. At the area of vibration signal processing and extraction, Yang JieMing and Xiong Shibo analysis the early fault feature extraction of gearbox based on wavelet packet, which achieves the separating of frequency of a gear vibration signal [1]; At the area of fault pattern recognition, Wang uses averaging method of signal establish the AR model about analysis of gearbox vibration signal [2].

With the development of computer technology, the artificial neural network (ANN) is widely used in gearbox fault diagnosis. But it has the problems of training slow, require large samples to learn and easy to fall into local minimum. The performance is difficult to be guaranteed.

Support Vector Machine (SVM) is an artificial intelligence method with superior performance. In 1999, SuyKensJ.A.K proposed Least Squares Support Vector Machine

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 277–283.
springerlink.com

(LSSVM) [3], which is simplified and variant of SVM. LSSVM has a lot of advantages include of simple training, better performance and so on. Because there are many faults in the gearbox, the paper proposes the MCLSSVM and does some test to prove its availability compare with ANN.

The remainder of the paper is designed as follows. In section 2, the basic principles of LSSVM depicted briefly. The following section 3 introduces the gearbox of wind power and the basic faults. Next, the MCLSSVM model is builded and it is used in fault diagnosis of wind power gearbox. Finally, we draw a conclusion in section 4.

2 Principle of LSSVM

LSSVM is the improvement of SVM, it turn inequality constraints into equality constraint [4]. It develops the speed of computation and lets the optimization problem become easy.

The basic principle is listed below:

The training sample is $(x_i, y_i), x_i \in R^n$, $y_i \in \{-1, +1\}$, $i = 1, 2, \cdots N$ .N is the total sample space, n is the dimension of sample space, y is the symbol of sample category. In the feature space, SVM model is:

$$y(x) = \omega^T \phi(x) + b \tag{1}$$

According to the structural risk principle, LSSVM classification problem can be described as following:

$$\min R(\omega, \xi) = \frac{1}{2}\omega^T \omega + \frac{1}{2}\gamma \sum_{i=1}^{N} \xi_i \tag{2}$$

$$y_i[\omega^T \phi(x) + b] = 1 - \xi_i, i = 1, 2, \cdots, N .$$

Where, $\phi(x)$ is mapping function in the kernel space, ω is the weight, b is the deviation, ξ_i is error variables, $\gamma > 0$ is punish coefficient, it can adjust error variables. The LSSVM corresponds to the Lagrangian function as following:

$$L(\omega, b, \xi; a) = \frac{1}{2}\omega^T \omega + \frac{1}{2}\gamma \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} a_i \{ y_i \left[\omega^T \phi(x_i) + b \right] - 1 + \xi_i \} . \tag{3}$$

Where $a_i \geq 0$ is Lagrangian multiplier ($i = 1, 2, \cdots, N$).

Finally, we can get the optimal classification decision-making for:

$$f(x) = sign\left[\sum_{i=1}^{N} a_i y_i K\left(x_i, x_j\right) + b \right]. \tag{4}$$

Where, sign is symbol function, $K\left(x_i, x_j\right)$ is kernel function. Frequently-used kernel functions are following: Polynomial kernel function, RBF kernel function, and Sigmoid kernel function [5].

Detailed derivation can be seen in literature [3].

3 Fault Diagnosis of Wind Turbine Gearbox

3.1 Introduction of Gearbox

Gearbox is the key components of wind turbine, which is between the impeller and generator. It transfers the power created by the impeller to generator, and at the same time, it turns the lower speed into the higher speed that the generator needed. Wind power generators are in general installed the place where around wind power and without sunscreen, such as field, seaside, the mountain pass. Gearbox is installed in the narrow cabin above tower, where is apart from the ground dozens of meters or more than one hundred meters. Gearboxes undergo extremely hot or cold weather perennially and bad natural environments. It is very difficult to repair because of inconvenient traffic, and fault period is generally appearing at the peak of the power generation. That will seriously affect the economic benefits of wind field once gearbox appears fault, Therefore, the reliability and working life of the gear box are needed the high requirements.

The typical faults of gearbox are listed below: wear of gear, broken teeth, tooth face scoring, tooth contact fatigue and bearing damage. In this paper, we use the MCLSSVM model to solve the classification problem of fault diagnosis with four kinds of fault for the example, such as fault of inner ring gear, spalling of outer ring gear, attrition of tooth surfaces and broken teeth.

3.2 The Model of MCLSSVM

Although LSSVM can effectively solve the problem of small samples learning, it can be seen from the classification principle that LSSVM is designed to solve the two classification problem. But in this paper we need to build five basic models of the gearbox fault, which is a multi-classification problem, therefore, it must be realized by using multi-classification LSSVM technology. At present, it is by the means of combining several two types of classifier to build multi-class classifier, the basic idea is resolving multi-classification problem into two kinds of problem to solve problem. There are three branch algorithms: one against rest (OAR), one against one (OAO) and decision directed acyclic graph (DDAG), among which OAO algorithm has the superior performance.

OAO classification method is building $m(m-1)/2$ decision-making functions between m types of fault, that is to say we build a classifier with any two fault samples and train with the corresponding fault. Through calculating the posterior probability of fault samples x between i and j, and combining with $m(m-1)/2$ two types of classifier to make sure the final posterior probability of sample belongs to each type of fault. Detailed classification principle can be seen in literature [6].

Ballot method [6] is the most commonly method to calculate the posterior probability of sample x in the $m(m-1)/2$ two types of classifier. We diagnose the fault sample x with each classifier, if the certain classifier estimates that it is belong to the ith kind fault, the ballot of ith kind adds one, or the ballot of jth kind adds one. At last, the fault type of most votes means that final posteriori probability of sample x belong to this fault sample is largest, so the vote strategy diagnoses the sample x as this type fault.

This paper combines MCLSSVM with ballot method and basic idea of decision tree to realize fault diagnosis of gearbox, and then it designs the MCLSSVM model of gearbox fault diagnosis in figure 1. A is the normal sample, B, C, D and E are for the fault samples.

In figure 1, the first floor is used to diagnose gearbox normal or not, and A is seen as the positive sample (+1), other faults are seen as the negative sample (-1) to train. There are two reasons why to do so, the first is that gearbox is on the normal state in the most of the time, its probability is larger than other faults. Second, it can improve the rate of diagnosis and reduce the scale of classification. The second floor of MCLSSVM model is used to recognize the 4 common fault of gearbox, which is realized by OAO strategy.

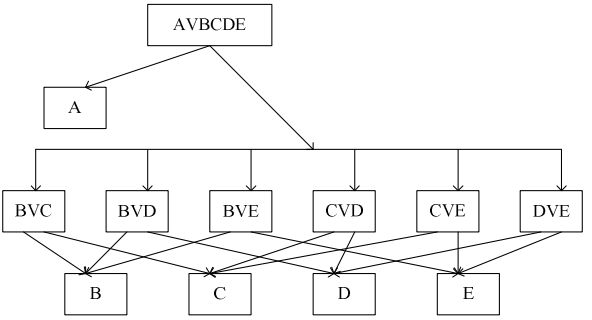


Fig. 1. MCLSSVM fault decision model of gearbox

Firstly, we build 6 LSSVM classifiers between any two of the 4 typical faults, each classifier is trained by the two corresponding fault samples. When we build the decision structure as figure 1, the decision rules for fault diagnosis of gearbox as following:

Step1: Use the first floor of LSSVM model to diagnose sample x, if it is normal, x is belong to A, gearbox is normal operating, or turn to step 2.

Step2: Use the 6 classifiers in the second floor of LSSVM model to diagnose sample x and vote.

Step3: According to the principle of "Max-wins" decide the fault mode of sample x. If some fault is the most votes, then we think the sample x belongs to this fault, diagnosis is finished. Or, two candidate fault modes have the same votes, it is needed the further diagnosis, turn to step4.

Step4: If the votes of i and j are same, according the posterior probability of sample x belongs to i and j to make sure the fault mode of it.

3.3 Experiment of Gearbox Fault Diagnosis

In this paper, we use the characteristic parameters in literature [7] which collects the vibration signal under different fault state, as shown in table1(have normalized).

Table 1. Sample characteristic parameters of gearbox for diagnosis

number	practical fault type	peak index	skewness index	frequency variance	harmonic factor
1	normal	0.0865	0.1184	0.9987	0.4798
2	normal	0.1076	0.2243	1.0001	0.1765
3	fault of inner ring gear	1.0001	0.5987	0.9756	0.13788
4	fault of outer ring gear	0.6523	0.6523	0.4826	0.0004
5	attrition of tooth surfaces	0.7623	0.4223	0.9872	0.1827
6	broken teeth	0.9712	1.0001	0.6634	0.0723

According to the samples that offered by literature [7], we train the 7 two types of LSSVM classifiers in figure 1 and build the corresponding decision functions. Where, set $\gamma = 15$, take RBF function $K(x_i, x_j) = \exp(-\left\|x_i - x_j\right\|^2 / 2\sigma^2)$ as the kernel function, set $\sigma = 1$.

To do the experiment by MCLSSVM with the samples of gearbox listed in table 1, the output diagnosis results of different LSSVM classifier are shown in table 2. According the output signs of each decision-making function decide the attribution of sample x and add 1 vote to the corresponding fault type. The final votes of each sample for diagnosis are listed in table3.

It can be seen from table 3 that the results of diagnosis is the same as the fault types set in literature [10], its effect is better than ANN in literature [8].

Table 2. The output of decision-making function for diagnosis samples

number	AVBCDE	BVC	BVD	BVE	CVD	CVE	DVE
1	0.8765	/	/	/	/	/	/
2	0.7612	/	/	/	/	/	/
3	-1.0000	1.0000	1.0000	1.0000	1.0000	-1.0000	-1.0000
4	-0.9987	-0.8762	0.0876	0.2218	1.0000	0.9962	-0.9987
5	-0.2345	-0.9976	-1.0000	-0.9987	-0.8765	1.0000	0.7643
6	-1.0000	0.4532	0.8754	-0.9875	1.0000	-1.0000	-1.0000

Table 3. Votes of diagnosis samples and result of diagnosis

number	type of fault					result of diagnosis
	A	B	C	D	E	
1	1	0	0	0	0	normal
2	1	0	0	0	0	normal
3	0	3	1	0	2	fault of inner ring gear
4	0	2	3	0	1	fault of outer ring gear
5	0	0	2	3	1	attrition of tooth surfaces
6	0	2	1	0	3	broken teeth

4 Conclusion

ANN model has been used widely in the fault diagnosis, but in the training process, ANN will fall into local minimum easily and convergence speed is slow, it needs large samples to diagnose. For this weakness, LSSVM is used to the fault diagnosis of gearbox, and then according the distribution characteristics of practical fault to build MCLSSVM model and design corresponding classification rule. Compare with ANN, MCLSSVM model is concise and high accuracy. On the other hand, its parameter is less and it just needs small sample to fault diagnosis. In a word, MCLSSVM is very effective and it has great potential.

Acknowledgement. This work was supported in part by Shanghai Municipal Science and Technology Commission. (Grant No.10JC1405800), Project of Science and Technology Commission of Shanghai Municipality(08DZ1200505), Project of Shanghai Municipal Economic and Information Commission(09A118), Project of Min Hang District Science and Technology Commission(2010MH181), Science and Technology Commission Key Discipline of Shanghai Municipal Education Commission(J51901), and Shanghai Dianji University(09C401).

References

1. Yang, J., Xiong, S.: Wavelet packet analysis method application in gearbox fault feature extraction. Vibration, Test and Diagnosis (2000)
2. Wang, W., Wong, A.K.: Autoregressive model based gear fault diagnosis. Journal of Vibration and Acoustics (2004)
3. Suykens, J.A.K., Van Gestel, T.: Least squares support vector machines. World Scientific, Singapore (2002)
4. Cristianini, N., Taylor, J.: An introduction to support vector machine. Electronic Industry Press (2004)
5. Chen, S.: The parameters optimization and improvement of LSSVM. East China University of Science and Technology Journal (natural science edition) (2008)

6. Moreira, M., Mayoraz, E.: Improved pairwise coupling classification with correcting classifiers. In: Processing of the 10th European Conference on Machine Learning, Germany Chemnitz: [S. n.] (1998)
7. Wu, D.: A fault diagnosis of gearbox based on SVM. Vibration, Test and Diagnosis (2008)
8. Meng, H.: Fault diagnosis of gearbox based on ANN and grey theory. North University of China (2005)

MUNOLD: Landslide Monitoring Using a Spatial Sensor Network

Ping Lu[1], Hangbin Wu[1,2], Gang Qiao[1,2], Weiyue Li[1,2], Xiaohua Tong[1,2], and Rongxing Li[1,3]

[1] Center for Spatial Information Science and Sustainable Development, Tongji University, China
[2] Department of Surveying and Geoinformatics, Tongji University, China
[3] Mapping and GIS Lab, the Ohio State University, USA

Abstract. MUNOLD is a novel landslide monitoring system composed of a set of spatial sensors. It is designed to be an early warning system which enables real-time monitoring of remote landslide areas. As its first version of implementation, 12 sensors were employed to collect geomorphologic, geotechnical and geophysical features in a down-scaled landslide simulation test site on the campus of Tongji University. All sensed data are collected and transferred to a database server through wireless communication. The whole system is expected to be further developed and applied in landslide monitoring in western China.

Keywords: spatial sensors, sensor network, landslide monitoring.

1 Introduction

Landslide is one of the major natural hazards. It is a threat not only to infrastructure facilities, but also to human lives [1], [2]. To reduce the risk level and damages caused, research on how to collect real-time precise spatial information before, during and after a landslide is critical for in-depth study of landslide mechanism and development of early warning systems.

Traditional landslide monitoring approaches focus on intensive field surveys after the installation of geotechnical and geophysical instruments. Despite its feasibility, this method is inefficient and time-consuming, thus making the approach impractical to carry out. This is especially true when a landslide site is located in remote areas, thereby making it difficulty in receiving real-time monitoring data and distributing early-warning messages. Moreover, conventional methods concentrate on sensors that work individually instead as system. As a result, they cannot provide the complete information in the global, regional and local areas, and further cannot provide the accurate information to validate the landslide prediction model.

A spatial sensor network is a network of sensor nodes that sense and control the environment, enabling interactions between persons/computers and the surroundings [3], [4]. It potentially provides an alternative approach for a real-time monitoring of landslide hazard. By integrating diverse sensors, the network can efficiently characterize geomorphologic, geotechnical and geophysical features of a landslide

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 285–289.
springerlink.com

with continuously-collected data, and meanwhile gains benefits from the combination of both remote sensing and in-situ sensors. The data then can be transferred instantly through wireless network to the port of a database server, thereby making subsequent monitoring, analyzing and visualization processes possible.

In this paper, MUNOLD (**MU**lti-Sensor **N**etwork for **O**bserving **L**andslide **D**isaster), a novel landslide monitoring system from a spatial sensor network, is presented. The sensor distribution, communication, and data storage of this system are demonstrated on a down-scaled landslide test site which is a simulation of Hongkou County in Sichuan Province, China. This paper briefly discusses the first three parts of the system: the down-scaled landslide model, sensor network and data transmission.

2 Construction of the Down-Scaled Landslide Body

In order to test the functionality of MUNOLD in landslide monitoring applications, a landslide simulation body was constructed on the campus of Tongji University, down-scaled from the in-situ landslide in Hongkou County of Sichuan Province. The length, width and height of the landslide body are 6m, 1.5m and 3m, respectively (Fig. 1). The landslide body was divided into 3 parts, with corresponding slope angle of 30^{o}, 15^{o} and 5^{o}, respectively, and 2m in length for each part. To simulate the dense gravel and loose soil characters of Hongkou area, the landslide mass was composed of two layers: the 0.5m height gravel and sand layer at the bottom, and the 0.3m height sand and clay layer on the top. The gravel and sand layer consists of clay, sand, and gravel with a proportion of 1:1:5. To monitor the internal slide of the landslide mass, a 0.1m diameter color sand bar was arranged at the outer side of the tempered glass every 1m. The detailed design and layout of the landslide body is shown in Fig. 1. A precipitation system was then installed above the landslide body, using the pumped water from two small reservoirs nearby. Also, three water pipes were put on the ground of the landslide body so as to simulate the underground water.

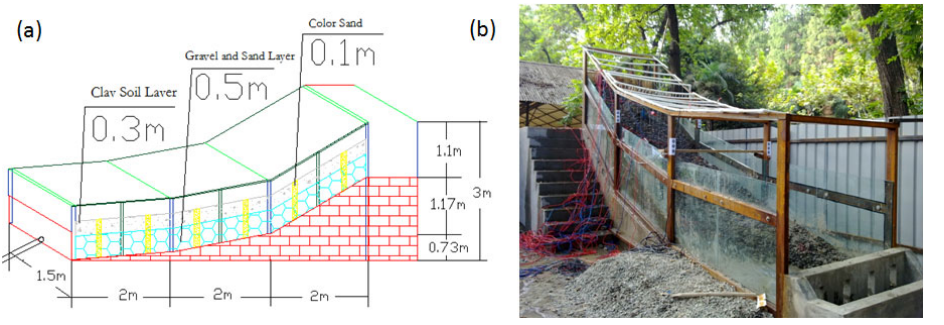


Fig. 1. The down-scaled landslide body for the MUNOLD test: (a) the design sketch of the landslide body and (b) the constructed landslide body.

3 Spatial Distribution of Multiple Sensors

To capture geomorphologic, geotechnical and geophysical features during the landslide process, 12 spatial sensors (see details in Table 1) were installed on the landslide body. A summary of available sensors and their monitoring targets is listed in Table 1. These sensors target on different landslide characteristics such as sliding surface displacement, slope inclination, pore water pressure of groundwater flows, landslide body acceleration, rainfall intensity *etc*. Among them, six accelerometers were installed in order to form a vector network of moving accelerations. In addition, eight pore water pressure gauges were aligned along the sliding direction, so as to interpret the pore water pressure changes before and after surface failure. Moreover, a weather station was set to collect meteorological features surrounding the landslide occurrence, for the purpose of weather forecast. The spatial distribution of these sensors is illustrated in Fig. 2. These sensors were synchronized using the time of the Data Taker (dataTaker®, DT80).

Table 1. A summary of deployed sensors and their monitoring targets for MUNOLD

Sensor	Precision	Freq.(*Hz*)	Monitoring Targets
Displacement meter	0.01mm	2	Surface displacement
Pore water pressure gauge	0.1Pa	1	Pore water pressure
Big tiltmeter	0.01degree	0.5	Slope inclination
Small tiltmeter	0.01degree	2	Slope inclination
Accelerometer	0.01g	80	Surface acceleration
Crackmeter	0.01mm	2	Crack displacement
Soil pressure gauge	0.1N	2	Pressure inside landslide body
Osmometer	0.1Pa	2	Osmotic pressure
Raingauge	0. 1mm	2	Rainfall intensity
GPS	mm - cm	60	Surface displacement
Weather station	N/A	1	Temperature, humidity, rainfall, *etc.*
Camera	N/A	25	DSM, color sand bar changes

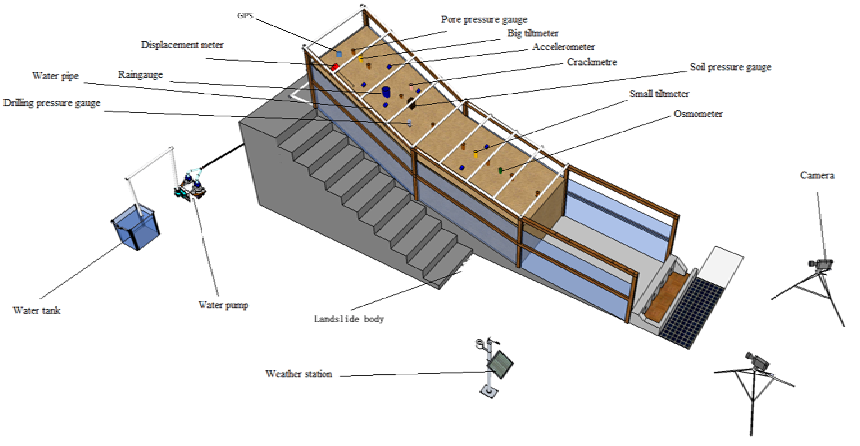


Fig. 2. The spatial distribution of applied sensors in MUNOLD

4 Wireless Transferring and Storage of Collected Data

The transferring methods are divided according to three types of sensors: narrowband high frequency sensors (*e.g.* six accelerometers, up to 80Hz), narrowband low frequency sensors (*e.g.* raingauge) and broadband sensors (*e.g.* images and videos). To obtain an instant data analyzing and decision making for early warning, all the narrow-band data are collected in a Data Taker and serial-port server and then transferred to the port of the remote database server through wireless and 3G/GPRS communication.

3G and GPRS are primarily used to transfer the narrowband data. In particular, GPRS was used to transfer the data collected by the Data Taker whereas 3G is employed to transfer the data captured by serial-port server, Ethernet and other methods. Through public internet, narrowband sensor data were then transferred to the database server. The communication speed of 3G and GPRS can reach to 350k/s and 30k/s maximally, enough for narrowband data acquisition and storage. Port 2002 and 2003 of the database server are used for data communication using TCP protocol. Port 2002 was used to maintain narrowband high-frequency data while port 2003 was used to manage the narrowband low frequency data. Once seized by client software, original sensor data transmitted as a data stream is divided and then stored in SQL Server database. Due to the data volume, high-resolution images and videos are separately transferred through a network of four iMesh instruments which enable an establishment of broadband channel with maximum transferring rate up to 8M/s. All the data stored in the database can be simultaneously analyzed using geotechnical model and the corresponding result is visualized in the decision making center in real-time (Fig. 3).

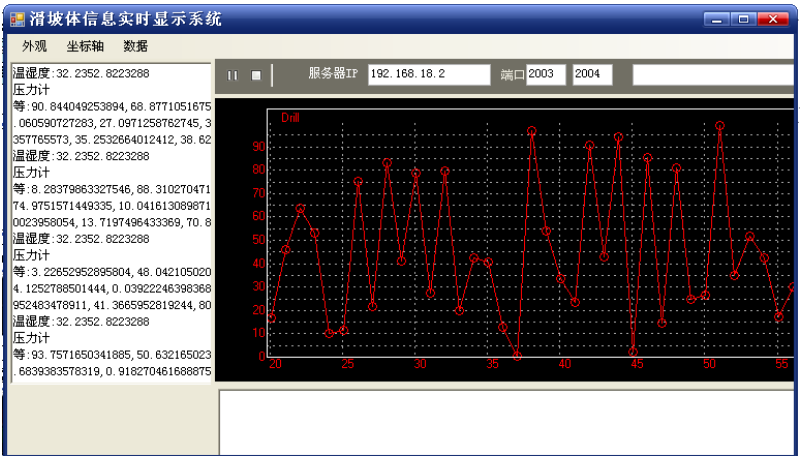


Fig. 3. One of the visualization windows of MUNOLD at the decision making center

5 Discussions and the Future Work

This paper introduces the initial research result of a novel landslide monitoring system of MUNOLD using a spatial sensor network. The system was tested in a down-scaled landslide simulation experiment on the campus of Tongji University, Shanghai, China. In total 12 spatial sensors were used to collect the geotechnical and geophysical features of the sliding body. The data was transferred wirelessly in real time to the remote port of the database server. After a few more tests, the system will be deployed in a real-world landslide site in western China. Further improvements should focus on building a stable system adaptable to field settings, an establishment of solar panel system for power supply and a remote control system for sensor commanding.

Acknowledgement. The authors appreciate the help of staff and students in the Center for Spatial Information Science and Sustainable Development, Tongji University, for their continuous support and contribution in this project. The project is supported by Guanghua Foundation and 985 Educational Fund.

References

1. Lu, P., Stumpf, A., Kerle, N., Casagli, N.: Object-oriented change detection for landslide rapid mapping. IEEE Geoscience and Remote Sensing Letters 8, 701–705 (2011)
2. Lu, P., Casagli, N., Catani, F., Tofani, V.: Persistent Scatterers Interferometry Hotspot and Cluster Analysis (PSI-HCA) for detection of extremely slow moving landslides. International Journal of Remote Sensing (in press), doi:10.1080/01431161.2010.536185
3. Verdone, R., Dardari, D., Mazzini, G., Conti, A.: Wireless and Actuator Networks. Elsevier, London (2008)
4. Li, R., The CSISSD research team: Advanced Spatial Sensor Network Systems: Review, Status, and Applications. In: Proceedings of the 2011 International Symposium on Image and Data Fusion, Tengchong, Yunnan, China, August 9-11 (2011)

KF vs. PF Performance Quality Observed from Stochastic Noises Statistics and Online Covariance Self-adaptation

Ken Chen[1], Meng Zhang[1], and Celal Batur[2]

[1] College of Information Science and Engineering,
Ningbo University, Ningbo 315211, China
[2] College of Engineering, University of Akron,
Akron, Ohio, 44325, U.S.A.
chenken@nbu.edu.cn

Abstract. The workability of Kalman filter is explored in the perspective of various stochastic noise statistics. The Gaussage is defined and applied at different levels in both Kalman filtering and particle filtering to evaluate the performance quality. Two parameters, DOM and DOD, are introduced and used for checking the consistence between the assumed stochastic noise covariance in Kalman filter and the truly existing covariance. Based on the proposed parameters, the method of online covariance adaptation is engineered, which is aimed to ultimately achieve the optimal Kalman performance quality.

Keywords: Kalman filter, particle filter, Gaussage, self-adaptation, performance quality.

1 Introduction

The Kalman filter (KF) and its extended form, the extended KF (EKF) [1], as well as the particle filter (PF) [2], have been widely applied in radar and video target tracking due mainly to their merits in optimal estimation for partially or entirely unobserved dynamic states by suppressing the stochastic noises existing in both process and measurement. Both KF and PF possess their genetic strengths and weaknesses. In the case that the given system is linear, and the process and measurement stochastic noises are white and Gaussian-distributed with zero mean, KF is proved the optimal approach for estimating the unattainable states by its regressive minimization of mean squared errors. KF is characterized by its relatively low computational complexity, memory space saving and easy implementation, so it has been selected as one of the principal algorithms in target tracking.

In real world, due to the time-variant changes in the states and in external environment, and in particular, the random disturbances represented by stochastic noises, the existing tracking algorithms bear such an inevitable shortcoming, that is, none of them is perfectly competent throughout the whole tracking process in terms of performance quality. In this work, the authors try to explore how the nature of stochastic noise affects the tracking performance. Two parameters, DOM and DOD,

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 291–298.
springerlink.com

are introduced and applied for on-line checking the KF performance quality. The parameters can virtually serve as the KF performance indicators, and can be used to adapt the assumed stochastic noise statistics so as to match with the true noise statistics. In so doing, the KF performance is improved.

2 Gaussage of Stochastic Noise Sequence

Due to the fact that a target of interest may move with high mobility, the dynamic model of the target motion may not sufficiently fit the true kinetics, plus other unknown or hard-to-quantify factors, all these concurrent elements exert, directly or indirectly, some adverse impact on the tracking performance. These unwanted but unavoidable components appear randomly as time elapses, and in terms of statistics, each individual magnitude may be uncertain, cross-uncorrelated, and unpredictable. These non-deterministic elements can all be taken in the form of stochastic noises of certain or uncertain distribution [1].

Because the KF (and the EKF) requires the stochastic noise to be white-and-Gaussian-distributed [3], in this paper, a new concept, Gaussage, is introduced to express "how gaussian" the stochastic noise is, which literally means "the grade of being Gaussian". The Gaussage is classified into the following categories: standard Gaussian (SG), over-Gaussian (OG), under-Gaussian (UG) and non-Gaussian (NG). An additional term, fuzzy Gaussian (FG), incorporates both OG and UG.

The Gaussage is related with four factors: kurtosis, *k*, skewness, *s* [4], μ for the mean of the stochastic noise sequence, and σ for its variance. *k* is used to determine how "symmetrically close" the probability density function (p.d.f) curve is to the Gaussian distribution geometrical center line; *s* is utilized to describe how skew the p.d.f curve is symmetrically to the same center line; μ is to express the location of the distribution center; and σ represents how deviating the distribution is from the μ.

Let *X* be a sequence of stochastic sequence, k, s, μ, σ are expressed as follows:

$$\mu = E(X), \tag{1}$$

$$\sigma = E[X - E(X)]^2, \tag{2}$$

$$s = \frac{E[X - E(X)]^3}{\{E[X - E(X)]^2\}^{3/2}}, \tag{3}$$

$$k = \frac{E[X - E(X)]^4}{\{E[X - E(X)]^2\}^2} - 3. \tag{4}$$

If the given stochastic sequence observes SG, then *k* and *s* tend to zero, and Gaussage, *G*, is defined as

$$G = \frac{E[X - \mu]^3}{\sigma^{3/2}} + \frac{E[X - \mu]^4}{\sigma^2} - 3, \tag{5}$$

When *G* is close to zero, the stochastic sequence satisfies FG distribution, otherwise is NG distribution. Further for FG, if $\sigma \ll 1$, the stochastic sequence is considered OG; if $\sigma \gg 1$, the sequence is deemed UG; when $\sigma \approx 1$, then SG.

3 Impact of Stochastic Noises on Performance Quality with KF and PF

To investigate the impact caused by stochastic noise of various statistics on the performance quality used with KF and PF, in this section, by simulation, the stochastic noise with different Gaussage, as is defined in Section 2, is to be applied in the given tracking algorithms, and the performance quality are to be compared with each other.

3.1 Performance Quality of KF and PF for Linear System

To simplify the problem, assume the state and measurement equations to be given as follows:

$$x(k+1) = -x(k) + w(k), \tag{6}$$

$$z(k+1) = x(k+1) + v(k), \tag{7}$$

In above equations, assume the covariance of process and measurement noise to be set as $Q=1, R=1$, respectively. The number of sampled particles for PF is 100, and the simulated epoch length takes 500. The process and measurement noises, $w(k)$ and $v(k)$, are regulated to separately observe UG, SG, OG and NG distribution.

To quantitatively describe the performance quality, Tab.1 gives the performance quality affected by noises using the root of mean squared errors (RMSE), defined as

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(x(k) - \hat{x}(k))^2}, \tag{8}$$

Table 1. Comparison of RMSE by stochastic noise for KF and PF

Algo. type / Noise type	Tracking by KF	Tracking by PF
UG distribution Q=R=100, G=0.0079	*RMSE* =7.9836	*RMSE* =0.6288
SG distribution Q=R =1, G =-0.1783	*RMSE* =0.8331	*RMSE* =0.4975
OG distribution Q=R =0.01, G =-0.0153	*RMSE* =0.0745	*RMSE* =0.1783
NG distribution Q=R =100, G =10.0273	*RMSE* =93.8539	*RMSE* =0.6971

3.2 Performance Quality of EKF and PF for Non-linear System

Assume the non-linear state and measurement equations being as

$$x(k+1) = 0.5x(k) + \frac{25x(k)}{1+x(k)^2} + 8\cos(1.2k) + w(k), \tag{9}$$

$$z(k+1) = \frac{x(k+1)^2}{20} + v(k), \tag{10}$$

As tabulated in Tab. 1, Tab. 2 shows the performance quality affected by noise of different *G*.

Table 2. Comparison of RMSE by stochastic noise for EKF and PF

Algo. type / Noise type	Tracking by EKF	Tracking by PF
UG distribution Q=R =100, G =0.0079	*RMSE* =28.5359	*RMSE* =0.4932
SG distribution Q=R =1, G =-0.1783	*RMSE* =13.0941	*RMSE* =0.3859
OG distribution Q=R =0.01, G =-0.0153	*RMSE* =7.7681	*RMSE* =0.3087
NG distribution Q=R =100, G =10.0273	*RMSE* =23.6463	*RMSE* =0.5415

From both Tab.1 and Tab.2, it can be noted that, for a linear system, if the stochastic noises satisfy the requirement for being Gaussian, KF performance is good in terms of tracking accuracy; however, if being Gaussian is not met, PF shows the higher tracking quality. In comparison, for a non-linear system, the stochastic noise variance Q and R need being preset, which may very likely not be in agreement with the true Q and R. This discrepancy may result in huge errors arising from performance divergence, causing significant tracking errors [5].

4 Check of Consistence between Assumed and True Q and R for KF

The KF filtering equation takes form of

$$\hat{x}_k = \hat{x}_k^- + K(z_k - H\hat{x}_k^-) \tag{11}$$

If the closer to the true states the priori estimation (prediction) of states, $\hat{x}_x^-$, is, the smaller the Kalman gain, K, is, indicating that the result of KF filter (correction) better approaches to the true states; contrarily, the further away from the true states the predicted $\hat{x}_x^-$ goes, K needs to weigh heavier, so that the filtering result achieves better accuracy. Because K is related to Q and R, so whether the assumed noise covariance in KF, denoted as Q_{kf} and R_{kf}, is consistent with the true covariance, expressed for Q_{truth} and R_{truth}, respectively, will have impact on the appropriateness of K, which ultimately plays a significant role in KF performance quality.

4.1 First Parameter for Consistence Check: DOM

Define the innovation sequence in KF algorithm as [5]

$$\sigma_k = H(x_k - \hat{x}_k^-) + v_k, \tag{12}$$

The "theoretic" covariance matrix of innovation sequence can be obtained by KF algorithmic computation, given as

$$C_{\sigma_k} = E[\sigma_k \sigma_k^T] = HP_k^- H^T + R_k \tag{13}$$

On the other hand, the "true" innovation covariance can be acquired by sampling with a sliding window of size *N*, denoted as $\hat{C}_{\sigma_k}$, computed from

$$\hat{C}_{\sigma_k} = \frac{1}{N} \sum_{j=j0}^{j0+N-1} \sigma_j \sigma_j^T, \tag{14}$$

where σ_j is determined using (12). The discrepancy between C_{σ_k} and $\hat{C}_{\sigma_k}$ is described by the degree of mismatch (DOM) [5], defined as

$$DOM = C_{\sigma_k} - \hat{C}_{\sigma_k} \tag{15}$$

If the discrepancy is significantly small, meaning DOM is around zero, it indicates that the assumed variances, Q_{kf} and R_{kf}, are in agreement with true ones, Q_{truth} and R_{truth}, respectively. Otherwise, if DOM is apparently non-zero, it indicates the gap between the two, and assumed variance needs being adjusted accordingly.

In simulation, the epoch length used in Eq. (6) and (7) is 1000, $Q_{kf} = 0.01$ and $R_{kf} = 1$ is assumed. For the first 300 points, the assumed variance agrees with the true one observing the SG, that is, $R_{truth1} = R_{kf} = 1$, $Q_{truth} = Q_{kf} = 0.01$; for the rest 700 points, it follows $R_{truth2} = 100$, suggesting UG, with $Q_{truth} = Q_{kf}$.

The conclusion may be reached based on the simulating results: when the stochastic noises are UG distributed, the error between the KF estimated states and true states is greater than that when SG is observed. When Q_{kf} and R_{kf} are pre-set to be smaller than true ones, DOM is negative.

Again assume $Q_{kf} = 0.01, R_{kf} = 1$, with $Q_{truth} = Q_{kf} = 0.01$. The first 300 points are mixed with measurement noise observing OG, with $R_{truth1} = 0.01$; for another 700 points, the noise is SG distributed, with $R_{truth2} = R_{kf} = 1$.

Based on the simulating results again, it comes to the conclusion that, when the stochastic noises observe OG, the error between the KF estimated states and true ones is less than that when SG is observed. When Q_{kf} and R_{kf} are pre-set to be greater than true ones, DOM is positive.

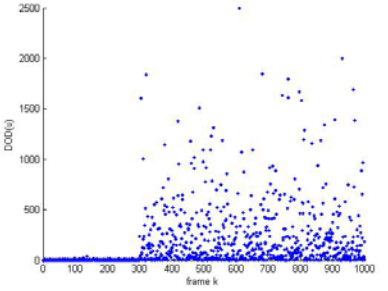
4.2 Second Parameter for Consistence Check: DOD

The degree of divergence is computed as [5]

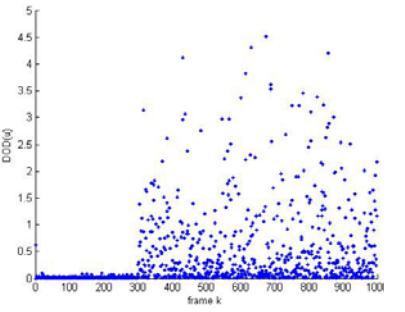
$$DOD = \frac{\sigma_k^T \sigma_k}{tr(C_{\sigma_k})} \tag{16}$$

where $tr(C_{\sigma_k})$ is the trace of covariance matrix of innovation sequence. The consistence check result conducted by DOD between the assumed noise covariance

and the true noise covariance is given in Fig.1. In Fig.1 (a), the Q and R are set identically with case described in Section 4.1, in which DOM is negative; and in Fig.1 (b), the Q and R are set identically also with that in Section 4.1, but in which DOM is positive.



(a) DOD with negative DOM



(b) DOD with positive DOM

Fig. 1. DOD using previously given Q and R in Section 4.1

The noise covariance consistence check by DOD is done through comparing the computed DOD and the threshold μ_T .

Let $R_{kf} = R_{truth}$, and $Q_{kf} = Q_{truth}$, and the process noise observes OG distribution with $Q_{kf} = Q_{truth} = 0.01$, the measurement noise is UG distributed with $R_{kf} = R_{truth} = 100$. The DOD is found to be 8 on these pre-set conditions, therefore the threshold μ_T is set for 8. Referring back to Fig.1 (a), $R_{kf} = 1$, recall that for the second set of 700 points, $R_{truth} = 100$, the DOD obtained in this case is found to be as high as 2500, which is far above the threshold $\mu_T = 8$.

The conclusion may be drawn upon the above experiment: when DOD is far greater than the threshold μ_T , the R_{kf} (and Q_{kf}) needs being increased.

Again let $R_{kf} = R_{truth}$, and $Q_{kf} = Q_{truth}$, both process noise and measurement noise observe OG distribution with $Q_{kf} = Q_{truth} = 0.01$ and $R_{kf} = R_{truth} = 0.01$, the DOD acquired in this case is less than 10, so the threshold is set to be 10. Referring back to Fig.1 (b), in which $R_{kf} = 1$, and for the first 300 points it holds $R_{truth} = 0.01$, the DOD is found to be almost zero, far less than $\mu_T = 10$.

Based upon the above experiment, it may conclude as follows: when DOD is far less than the threshold μ_T , the R_{kf} (and Q_{kf}) needs being decreased.

To sum up, when the stochastic noise is UG distributed, the error is significant, and in this case, when the assumed covariance is not in agreement with the true one, the error can grow even greater. Both DOM and DOD can be used as the parameters to check whether the assumed covariance is consistent with the true one.

5 Self-adaptation of R_{kf} and Q_{kf}

When DOM stays around zero, it indicates that the assumed R_{kf} and Q_{kf} are sufficiently consistent with R_{truth} and Q_{truth}, while DOM is significantly greater than zero, R_{kf} and Q_{kf} need to be re-set for smaller value, while DOM is significantly less than zero, both are required to be adjusted for greater value. Based on this identified correlation, the following focuses on using functional relation to automatically adjust R_{kf} and Q_{kf} as well as DOM, achieving the agreement between the assumed covariance and true one so that KF optimal performance is attained as a result.

When UG is observed, and the assumed covariance is much smaller than the true one, for example, notably $R_{kf} < R_{truth}$, it will cause the error to go greater. Now suppose the relation between the "current" R_{kf} and the "previous" R_{kf} as

$$R_{kf}(k) = \beta R_{kf}(k-1) \tag{17}$$

in which β is the function of DOM. Further, when $DOM > 0$ significantly, $0 < \beta < 1$; while $DOM < 0$ remarkably, $\beta > 1$. The relation between β and DOM is found to be

$$\beta = \begin{cases} e^{-DOM/100} & |DOM| > 2 \\ 1 & \text{others} \end{cases} \tag{18}$$

Fig.2 shows the error when for the first 300 points $R_{truth} = 1$, and for the rest 700 points $R_{truth} = 10$, in this case $R_{kf} = R_{truth}$. In contrast, in Fig.3, R_{truth} is the same as given in Fig. 2, but starting with $R_{kf} = 1$, indicating the significant inconsistence for the rest 700 points because $R_{truth} = 10$. Fig.3 shows the error after the self-adaptation of R in this case, and it can be noted that the error is not worsened due to the application of self-adaptation of R_{kf}, which would have been so if the self-adaptation was not applied.

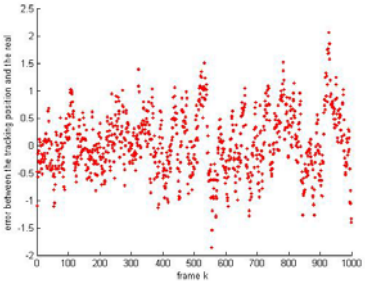


Fig. 2. Error from consistent noise covariance $R_{kf} = R_{truth}$

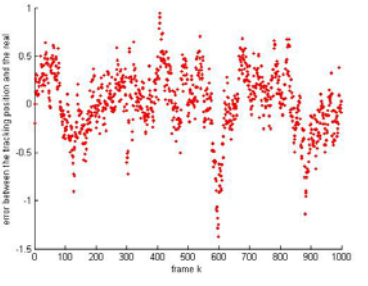


Fig. 3. Error from inconsistent noise covariance with self-adaptation of R_{kf} adaptation of R

6 Conclusion

In this paper, the concept of Gaussage is proposed for describing the stochastic noises statistics, and categorized into over-, standard-, under-, and non-Gaussian cases. These four different cases are applied in the Kalman filtering and particle filtering respectively, and the performance quality of each filter is analyzed accordingly.

Two parameters, DOM and DOD, are introduced and employed to check the consistence between the assumed stochastic noises covariance and the true one. Based on the variation of these two parameters, the assumed covariance can be self-adapted to make it adjust to the value of the true covariance which exists but is unknown in the Kalman filtering process. All the technical efforts devoted in this work are to observe how the stochastic noise statistics play a role in the performance of KF and PF, and ultimately improve the performance quality of KF.

Acknowledgments. Thanks are particularly given to Ningbo Natural Science Foundation of Zhejiang Province (2010A610109) for its sponsor to this research project.

References

1. Welch, G., Bishop, G.: An Introduction to the Kalman Filter, http://www.cs.unc.edu/~welch
2. Khan, Z., Balch, T., Dellaert, F.: An MCMC-Based Particle Filter for Tracking Multiple Interacting Targets. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3024, pp. 279–290. Springer, Heidelberg (2004)
3. Maybeck, P.S.: Stochastic models, estimation, and control, vol. 1, ch.1. Academic Press, New York (1979)
4. Jiang, Y., Chen, X., Tao, J.Y., Zhang, C.H.: Numerically Simulating Non-gaussian Random Processes with Specified PSD,Skewness and Kurtosis. Journal of System Simulation 18(5), 1127–1130 (2006) (in Chinese)
5. Jwo, D.J., Cho, T.S.: A practical note on evaluating Kalman filter performance optimality and degradation. Applied Mathematics and Computation 193(2), 482–505 (2007)

Approximate Estimates of Semiparametric Reproductive Dispersion Nonlinear Mixed Effect Models

Hongxing Shi

School of Primary Education, Chuxiong Normal University,
675000, Chuxiong, P.R. China
shx@cxtc.edu.cn

Abstract. Semiparametric reproductive dispersion nonlinear mixed effect model(SRDNMM) is an extension of semiparametric nonlinear mixed effect models and nonlinear reproductive dispersion models. It appears to be very flexible and useful in the analysis of complex longitudinal data not only because the covariates are often included into explain the inter-individual and intra-individual variations as well as the time effect by an unknown way, but also because it allows for a variety of distributions which is much wider than exponential family for response. Although for semiparametric reproductive dispersion nonlinear mixed effect model, the likelihood method is a standard approach for inference, it can be computationally challenging. So it is necessary for us to develop some computationally efficient approximate methods. In this article, we propose a novel approximate estimate method for semiparametric reproductive dispersion nonlinear mixed effect models and a simulation study is used to illustrate the proposed methodologies.

Keywords: Longitudinal data, Semiparametric model, Reproductive dispersion model, Mixed effect.

1 Introduction

Longitudinal data such as repeated measurements arise frequently from many biomedical and clinical studies as well as from other scientific areas. Semiparametric nonlinear mixed effect models are powerful tools for modeling the relationship between a response variable and the covariates in longitudinal studies, e.g., Ke and Wang[1]; Wu and Zhang[2]; Liu and Wu [3].

Moreover, in practical, there exist some complex data which can not be described well either by normal distribution or by exponential family distribution. Jorgense[4] proposed a reproductive dispersion model which includes a wide range of distributions such as normal, binomial, exponential, Gamma, Vonmises, simplex and Gumbel distributions as its special cases and these families of distributions have received a lot of attention in past decades. For example, Tang et al[5-6] and Chen et al[7] discussed some important issues such as statistical inference, model diagnosis and local influence analysis on the basis of the assumption of responses from a reproductive dispersion model for nonlinear models, semiparametric models as well as nonlinear mixed effect models etc.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 299–306.
springerlink.com

In this paper, we mainly focus on the issue of modeling the longitudinal data in which the response is from a reproductive dispersion family while semiparametric effect exists between the covariates, and propose a new class of semiparametric nonlinear mixed effect models. Because of the complexity of the model, the observed data log-likelihood function involves intractable high-dimensional integrals and in the usual condition, a Monte Carlo EM (MCEM) algorithm is an optimal choice. However, for our model considered, the "curse of dimensionality" which results from the mixed effect may appear and the Monte Carlo approximation may become unconvergence in this case. Then, it is necessary for us to develop some new methods to overcome this difficult.

In general, we establish a first-order linear approximate algorithm which is rather effective in parameter estimation for the semiparametric nonlinear mixed effect models and a simulation study is used to illustrate the proposed methodologies.

2 Notation and Model

Suppose that the response variable $y_{ij}(i=1,2,\cdots,m, j=1,2,\cdots,n_i)$ is independent observation of the i-th individual at the time point t_{ij}, and we assume that $y_{i1}|b_i, y_{i2}|b_i, \cdots, y_{in_i}|b_i$ are conditional independent, each has the following probability density function:

$$P\left(y_{ij}|b_i\right) = a\left(y_{ij};\sigma^2\right)\exp\left\{-\frac{1}{2\sigma^2}d\left(y_{ij};\mu_{ij}\right)\right\} \tag{1}$$

where μ_{ij} is the location parameter which does not denote the mean in general, $\sigma^2 \in \Lambda(\Lambda \subset R^+)$ is referred to as the dispersion parameter which is known or can be estimated separately. $a(\cdot,\cdot)$ is a suitable known function, $d(y,\mu)$ is a unit deviance defined on $C\times\Omega$ ($\Omega \subseteq C \subseteq R$ is an open interval) and satisfies $d(y,\mu)=0$ for $\forall y \in C$ and $d(y,\mu)>0$ for $\forall y \neq \mu$, and is twice continuously differentiable with respect to (y,μ) on $C\times\Omega$.

Following Jorgensen [4], model (1) is called the reproductive dispersion model (RDM), it includes normal distribution, extreme value distribution, Gamma distribution, Vonmises distribution, simplex distribution, Gumbel distribution and exponential family distribution as its special cases. Now, we extend the reproductive dispersion model to the following semiparametric nonlinear mixed effect model as:

$$\mu_{ij} = f\left[h\left(\alpha, b_i, x_{ij}, z_{ij}\right), g\left(\eta(t_{ij}), v_i(t_{ij})\right)\right] \tag{2}$$

where f, h, g are known link functions which may be nonlinear and reflect the associations between the responses and the mixed effects, α denotes fixed effect while b_i is random effect with their corresponding covariates x_{ij} and z_{ij}, and we assume that $b_i \sim N(0, D_b), (i=1,...,m)$. Moreover, h is the parametric mixed effect components and g denotes the nonparametric mixed effect components,

where $\eta(t)$ and $v_i(t)$ are unknown smooth fixed effect and random effect functions respectively. In addition, $v_i(t)$ is assumed to be i.i.d. copy of a zero-mean Gaussian stochastic process $V(t)$ with covariance function $\gamma(s,t)$, that is $v_i(t) \sim GP(0,\gamma)$ and $E[b_i v_i(t)] = \gamma_b(t)$. Thus, we call the model given in (1) and (2) as semiparametric reproductive dispersion nonlinear mixed effect model(SRDNMM), which can be regarded as a unified mixed effect model in the sense that it includes almost all mixed effect models as it special cases.

3 Nonparametric Smooth Technique

Let χ be the support of t, and assume that $\eta(t)$ is an element of a smooth function space $S_\eta(\chi) \subseteq L^2(\chi)$ such as the Sobolev space. Denote a completely orthonormal basis of $S_\eta(\chi)$ by $\Psi(t) = [\psi_0(t), \psi_1(t), \cdots]^T$, where $\psi_0(t) = 1$.

Thus, $\eta(t) = \sum_{k=0}^{\infty} \beta_k \psi_k(t)$,

in which $\beta_k = \int_\chi \eta(t) \cdot \psi_k(t) dt$.

If we write $\Psi_r(t) = [\psi_0(t), \cdots \psi_{r-1}(t)]^T$ and $\beta = (\beta_0, \cdots, \beta_{r-1})^T$, then we have $\eta_r(t) = \sum_{k=0}^{r-1} \beta_k \psi_k(t) = \Psi_r^T(t) \cdot \beta$ and $\eta_r(t)$ will converge to $\eta(t)$ in L^2-norm as r tends to infinity. In general, the value of r can be chosen based on AIC or BIC criterion. Furthermore, we can obtain an approximation to $v_i(t)$ in a similar way. Let $\Phi_q(t) = [\phi_0(t), \cdots, \phi_{q-1}(t)]^T$, and $C_i = (c_{i0}, \cdots, c_{i,q-1})^T$, then it is hold that $v_{iq}(t) = \sum_{k=0}^{q-1} c_{ik} \phi_k(t) = \Phi_q^T(t) \cdot C_i$.

Clearly, $\eta(t) \approx \eta_r(t) = \Psi_r(t)^T \cdot \beta$ and $v_i(t) \approx v_{iq}(t) = \Phi_q^T(t) \cdot C_i$, where β and C_i are unknown vectors of fixed and random coefficients, respectively. Based on the assumption of $v_i(t)$, we can treat C_i as i.i.d. zero-mean random vector. In practice, we may consider natural cubic spline bases with percentile-based knots, and determine the number of knots by the AIC or BIC.

According to the argument given above, the SRDNMM can be represented as:

$$\begin{cases} P(y_{ij} | b_i) = a(y_{ij}; \sigma^2) \exp\left\{ -\dfrac{1}{2\sigma^2} d(y_{ij}; \mu_{ij}) \right\} \\ \mu_{ij} = f\left[h(\alpha, b_i, x_{ij}, z_{ij}), g(\beta, C_i, \Psi_r(t_{ij}), \Phi_q(t_{ij})) \right] \end{cases} \tag{3}$$

Moreover, it is reasonable to assume that the coefficient $C_i, (i = 1, 2, \cdots, m)$ is i.i.d. copy of a normal zero-mean random vector C with covariance matrix D_c. Let $\mathrm{cov}(b_i, c_i) = D_{bc}$, then, we can obtain the following approximation:

$$\gamma(s,t) \approx \Phi_q(s)^T D_C \Phi_q(t), \quad \gamma_b(t) = D_{bc} \cdot \Phi_q(t) \tag{4}$$

For convenience, we write random effect coefficients as $r_i(t) = \left(b_i^T, C_i^T\right)^T$, then, we have $r_i \sim N(0, D)$, where $D = \begin{bmatrix} D_b & D_{bc} \\ D_{bc}^T & D_c \end{bmatrix}$ is unknown covariance matrix.

4 Approximate Estimate Method Based on Iteration

For simplicity, let $\theta = (\alpha, \beta, D)$ denotes all unknown parameters except σ^2. Based on the observed data set $\left\{\left(y_{ij}, x_{ij}, z_{ij}, t_{ij}\right), i = 1, \cdots, m, j = 1, \cdots, n_i\right\}$, the observed log-likelihood function takes the form:

$$l\left(\theta, \sigma^2 \middle| y, x, z, t\right)$$

$$= \int \left\{ \sum_{i=1}^{m} \sum_{j=1}^{\eta_i} \left[\log a\left(y_{ij}; \sigma^2\right) - \frac{1}{2\sigma^2} d\left(y_{ij}; \mu_{ij}\right) \right] - \sum_{i=1}^{m} \left[\frac{1}{2} r_i^T D^{-1} r_i + \frac{r}{2} \log(2x) + \frac{1}{2} \log|D| \right] \right\} dr \tag{5}$$

Obviously, the above log-likelihood function involves high-dimensional integrals that are hard to handle, then, the complete data log-likelihood function is considered:

$$Q\left(\theta \middle| \theta^{t-1}\right) = E_{r|y}\left[\sum_{i=1}^{m} \log p\left(y_i, r_i \middle| \theta, \middle| \theta^{(t-1)}\right) \right] \approx \frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{m} \log p\left(y_{ij}, r_i^{(n)} \middle| \theta^{(t-1)}\right) \tag{6}$$

The Q-function plays an important role in EM algorithm, which requires drawing random samples $\left\{r_i^n, i = 1, \cdots, N\right\}$ by Gibbs sampler to approximating (6). However, it should be noted that as the mixed effect involved in our model, the Gibbs algorithm may exhibit very slow convergence or non-convergence. Meanwhile, when the dimension of the random effect is small, the numerical integration methods such as Gaussian quadrature may not be feasible. Consequently, we propose a computationally efficient approximate method which completely avoids integration and thus offer a big computational advantage.

Inspired by Lindston and Bates [8], Wu [9], Zhang and Wu [10], we approximate this model using a Taylor expansion in the random effect vector r_i about the mean value $E(r_i) = 0$, which yields:

$$\mu_i \approx \bar{\mu}_i = \left(\bar{\mu}_{i1}, \cdots, \bar{\mu}_{in_i}\right) = f_i\left\{w\left(\bar{X}_i, \theta, 0\right)\right\} + F_i(\theta, 0) \cdot \Delta_{r_i}(\theta, 0) r_i \tag{7}$$

where $f_i\left\{w\left(\bar{X}_i, \theta, r_i\right)\right\} = f\left[h\left(d, b_i, x_i, z_i\right), g\left(\beta, C_i, \Psi_r(t_i), \Phi_q(t_i)\right)\right]$,

$$\bar{X}_i = \left(x_i, z_i, \Psi_r(t_i), \Phi_q(t_i)\right), \quad w(\cdot) = \left(h(\cdot), g(\cdot)\right)$$

and $F_i(\theta, 0) = \dfrac{d}{dw_i}\left[f_i(w_i)\right]\Big|_{w_i = w(\bar{x}_i, \theta, 0)}$, $\Delta_{r_i}(\theta, 0) = \dfrac{d}{dr_i} w\left(\bar{X}_i, \theta, r_i\right)\Big|_{r_i = 0}$.

Furthermore, suppose $H_i = F_i(\theta,0)\Delta_{r_i}(\theta,0)$, and U_i is an orthonormal matrix which meets $U_i \cdot U_i^T = H_i^T \cdot D^{-1} H_i$. Let $\overline{r}_i = U_i r_i$, $m(\overline{r}_i | y_i, \theta) = d(y_i; \overline{\mu}_i) + \overline{r}_i^T \cdot r_i$, then, we can establish the iteration algorithm for estimating the interested parameter θ and the random effect r_i as follows:

Step1: given the initial values of $\theta^{(0)}$ and $\overline{r}_i^{(0)}$, $\theta^{(0)}$ can be obtained by classical estimation based on Fisher-score algorithm without random effect and $\overline{r}_i^{(0)}$ can be determined by $\overline{r}_i^{(0)} = -U_i \cdot e_i^{(0)} / 2$, where $e_i^{(0)} = \dfrac{\partial d(y_i;\mu_i)}{\partial \mu_i}\Big|_{\theta=\theta^{(0)}}$.

Step2: with the current $\theta^{(k)}$ and $\overline{r}_i^{(k)}$ for the k-th iteration and each fixed i, update $\overline{r}_i$ according to the following equation:

$$\overline{r}_i^{(k+1)} = \overline{r}_i^{(k)} - \left[\ddot{m}\left(\overline{r}_i^{(k)} \middle| y_i, \theta^{(k)}\right)\right]^{-1} \dot{m}\left(\overline{r}_i^{(k)} \middle| y_i, \theta^{(k)}\right) \tag{8}$$

where $\ddot{m}(r_i | y_i, \theta) = U_i^T H_i^T V_i H_i U_i + 2I$ and $V_i = \dfrac{\partial^2 d(y_i;\mu_i)}{\partial \mu_i \cdot \partial \mu_i^T}$. Given small value δ_1 (e.g. 0.001), if $\left\|\overline{r}_i^{(t)} - \overline{r}_i^{(t-1)}\right\|_\infty < \delta_1$, thus, the convergence is claimed, where $\|\cdot\|_\infty$ denotes the infinite norm of vector and we write $\overline{r}_i^{(k+1)} = \overline{r}_i^{(t)}$.

Step3: once $\theta^{(k)}$ and $\overline{r}_i^{(k+1)}$ are known, we can obtain $\theta^{(k+1)}$ according to the iteration equation: $\theta^{(k+1)} = \theta^{(k)} + \left[J\left(\theta^{(k)}\right) + \lambda I_p\right]^{-1} S\left(\theta^{(k)}\right)$, where λ is a pre-designated positive constant which is used to adjusting the Fisher information matrix, $J(\cdot)$ and $S(\cdot)$ are Fisher information matrix and score vector corresponding to θ for model given in (3) and (4), and I_p is a identity matrix in which p denotes the dimension of θ.

Step4: repeat step 2 to step 3 until there exists some s rendering $\left\|\theta_i^{(s)} - \theta_i^{(s-1)}\right\|_\infty < \delta_2$, where δ_2 is a pre-designated positive constant and thus, we write $\hat{\theta}(\lambda) = \theta^{(s)}$.

Step5: for each fixed λ, repeat step 1 to step 4, and we can obtain $\hat{\theta}(\lambda)$. Choose λ so that each component of $\hat{\theta}(\lambda)$ appears to be even and then, we write $\theta = \hat{\theta}(\lambda)$ to be the maximum likelihood estimation of θ.

5 Simulation Study

In this section, a simulation study is presented to illustrate the application of our proposed method in estimation of the semiparametric reproductive dispersion nonlinear mixed model. Firstly, we generate the response $y_{ij} (i = 1,...,m; j = 1,...,n_i)$ from the following simplex distribution with the probability density function:

$$p(y_{ij};\mu_{ij}\mid b_i)=\begin{cases}\left[2\pi\sigma^2\{y_{ij}(1-y_{ij})\}^3\right]^{-1/2}\exp\left\{-\dfrac{(y_{ij}-\mu_{ij})^2}{2\sigma^2 y_{ij}(1-y_{ij})\mu_{ij}^2(1-\mu_{ij})}\right\}, & \text{if } 0<y_{ij}<1\\ 0, & \text{otherwise}\end{cases}$$

where parameter σ^2 is assumed known and the systematic part of the model is given by:

$$\mu_{ij}=f[h(\alpha,b_i,x_{ij},z_{ij}),g(\eta(t_{ij}),v_i(t_{ij}))],$$

where $f(h,g)=\log(e^{-h}+e^{-g})$ and $h(\alpha,b_i,x_{ij},z_{ij})=\alpha_0+x_{1ij}\alpha_1+x_{2i}\alpha_2+b_{i0}+z_{1i}b_{i1}$, $g(\eta(t_{ij}),v_i(t_{ij}))=\eta(t_{ij})+v_i(t_{ij})$, and $\eta(t)=\sin(\pi t/2)/\{1+2t^2[sign(t)+1]\}$ and $v_i(t)=v_i\sin(t), v_i\sim N(0,1)$.

Moreover, we assume $x_{1ij}\sim U[-1,1]$, both x_{2i} and z_{2i} follow a discrete uniform distribution of set {-1,1}, $t_{ij}\sim U[-2,2]$ and we assume random effects $(b_{i0},b_{i1})\sim N(0,D_b)$,where $D_b=\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{pmatrix}$, and the true values of the unknown parameters are set to be $\alpha_0=1.5,\alpha_1=-1,\alpha_2=2,\sigma_{11}=\sigma_{22}=1,\sigma_{12}=0.5$ and $m=50,n_i=15$.

For the data set generated above, we adopt the proposed first-order linear approximate algorithm given in section 4 to obtain the estimation of unknown parameters. Table 1 shows the corresponding estimation results for 100 iterations, where "Bias" denotes the absolute value between the true value and the mean of estimation and "ASE" denotes the average standard error from each estimates. Examination Table 1 we can found that the performance of our algorithm is quite well and the Bias of interested parameters are rather small while those corresponding to covariance matrix for random effect are slightly bigger.

Table 1. The results of estimation in simulation

parameter	α_0	α_1	α_2
Bias	0.014	0.022	0.041
ASE	0.649	0.581	1.233
parameter	σ_{11}	σ_{12}	σ_{22}
Bias	0.045	0.021	0.035
ASE	1.823	0.873	1.842

Furthermore, we also present the convergence process of algorithm and the nonparametric estimation in Figure 1. It can be easily found from Figure 1 that the algorithm is converged after about 50 iterations and the nonparametric function is fitted well by p-spline generally.

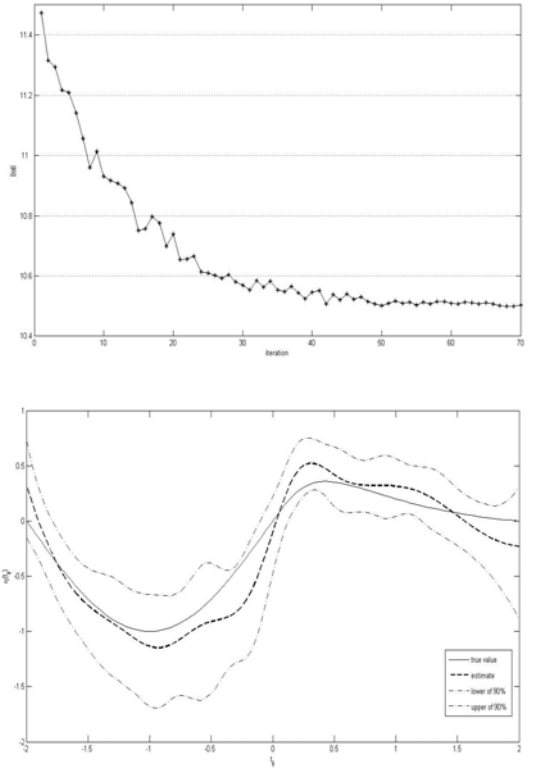


Fig. 1. The convergence process and the nonparameteric estimation

6 Conclusions

In this paper, we propose a novel approximate estimate method for a new class of semiparametric nonlinear mixed effect models which assume that the response is from a reproductive dispersion model. The main contributions of our paper are that firstly, the model we considered is rather general as it allows for a variety of distributions which is much wider than exponential family for response and secondly, our method is computationally effective as it completely avoids intractable high-dimension integration. Furthermore, our future research is to consider the issues of inference for the semiparametric nonlinear mixed effect models with nonparametric assumption for random effect.

References

1. Ke, C., Wang, Y.: Semiparametric nonlinear mixed-effects models and their applications. J. Amer. Stat. Asso. 96, 1272–1298 (2001)
2. Wu, H., Zhang, J.: The study of long-term HIV dynamics using semi-parametric non-linear mixed-effects models. Stat. in Medi. 21, 3655–3675 (2002)

3. Liu, W., Wu, L.: Simultaneous inference for semiparametric nonlinear mixed-effects models with covariate measurement errors and missing responses. Biometrics 63, 342–350 (2007)
4. Jorgense, B.: The theory of dispersion models, pp. 281–294. Chapman and Hall (1997)
5. Tang, N.S., Wei, B.C., Zhang, W.Z.: Influence diagnostics in nonlinear reproductive dispersion mixed models. Statistics 40, 227–246 (2006)
6. Tang, N.S., Chen, X.D., Wang, X.R.: Consistency and asymptotic normality of profile-kernel and backfitting estimators in semiparametric reproductive dispersion nonlinear models. Sci. in China Series A 25, 757–770 (2009)
7. Chen, X.D., Tang, N.S., Wang, X.R.: On confidence regions of semiparametric nonlinear reproductive dispersion models. Statistics 43, 583–595 (2009)
8. Lindstrom, M.J., Bates, D.M.: Nonlinear mixed effects models for repeated measures data. Biometrics 46, 673–687 (1990)
9. Wu, L.: A joint model for nonlinear mixed-effects models with censoring and covariates measured with error with application to AIDS studies. J. Amer. Stat. Asso. 97, 955–964 (2002)
10. Wu, H.L., Zhang, J.T.: Nonparametric Regression Methods for Longitudinal Data Analysis, pp. 212–234. John Wiley & Sons (2006)

Continuous Control of Lagrangian Data

Pierre Allain[1], Nicolas Courty[1], and Thomas Corpetti[2]

[1] VALORIA, University of Bretagne Sud, 56000 Vannes, France
[2] CNRS, LIAMA (TIPE project), ZhongGuanCun DongLu 95, Beijing 100190, P.R. China

Abstract. This paper addresses the challenging problem of globally controlling several (and possibly independent) moving agents that together form a whole, generally called swarm, which may display interesting properties. Applications are numerous and can be related either to robotics or computer animation. Assuming the agents are driven by their own dynamics (such act as Newtonian particles), controlling this swarm is known as the particle swarm control problem. In this paper, the theory of an original approach to solve this issue, where we rely on a centralized control rather than focusing on designing individual and simple rules for the agents, is presented. To that end, we propose a framework to control several particles with constraints either expressed on a per-particle basis, or expressed as a function of their environment. We refer to these two categories as respectively *Lagrangian* or *Eulerian* constraints.

1 Introduction

Planning trajectories is the task that consists in estimating the best path of a set of particles in order to reach a desired configuration/scenario. This issue has many applications such as robot swarms [1], simulation of schools and herds of animals [2], human crowd simulation [3] or even mobile networks with switching topology [4]. Many competitive approaches have been proposed in the literature to solve this tricky problem. Most of the work on particle swarm control focuses on individual control laws, usually based on potential fields (see e.g. [5,6]), to define the control policy. The latter induces an emerging and global behavior of the swarm also known as *the swarm intelligence*, which has raised several important stability issues [7,8]. A good review of those problems and the common associated recipes, along with stability results, can be found in [6].

In this paper, unlike most of the previous studies, particles are not examined individually but we rather prefer to define the problem as a *centralized control*. We indeed assume that each particle evolves under a continuous low and particles are therefore dependent each others. Hence, the swarm control operates globally to enforce consistency at the scene level.

This approach enables flexibility since the commands can then either be expressed

a) at a particle level: position of particles, spatial relations between particles, ... at a given time t. This is referenced as *Lagrangian constraints*, or

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 307–310.
springerlink.com

b) at a domain specific level: spatially-continuous density of particles, velocity, ... at a given time t . This is identified as *Eulerian constraints*.

It is obvious that such properties in a unified framework enable a rich range of applications.

The general idea to deal with this problem is introduced in this paper. Roughly speaking, the technique consists in rewriting a completely continuous control theory (based on the variational assimilation framework) in order to allow situations where the system state, the observations and the dynamical model do not belong to the same space of representation. The mapping between the different spaces involved (eulerian and lagrangian) is performed thanks to operators inspired from the graph theory [9].

The paper is organized as follows. The section 2 presents the basic concepts of optimal control theory based on variational data assimilation. In section 3, we introduce the key points enabling a representation of lagrangian data and their relationships with the eulerian space. Due to space limitations, only some video screenshots are drawn but videos can be seen on supplementary files.

2 Variational Data Assimilation

In this section we present the main principles of variational data assimilation. We refer the reader to [10] for complete methodological aspects.

The general problem consists in recovering, from an initial condition, a system's state $\mathbf{X}$ partially observed and driven by approximately known dynamics $\mathbb{M}$. This can be formalized as finding $\mathbf{X}(t)$ at time $t \in [t_0, t_f]$, that satisfies the system:

$$\frac{\partial \mathbf{X}}{\partial t} + \mathbb{M}(\mathbf{X}) = \boldsymbol{\epsilon}_{\mathbb{M}} \quad ; \quad \mathbf{X}(t_0) = \mathbf{X}_0 + \boldsymbol{\epsilon}_0 \quad ; \quad \mathbb{H}(\mathbf{X}) = \mathbf{Y} + \boldsymbol{\epsilon}_{\mathbb{H}} \tag{1}$$

where $\mathbb{M}$ is the non-linear operator relative to the dynamics, $\mathbf{X}_0$ is the initial vector at time t_0 and $(\boldsymbol{\epsilon}_{\mathbb{M}}, \boldsymbol{\epsilon}_0)$ are (unknown) additive control variables relative to noise on the dynamics and the initial condition respectively. In addition, noisy measurements $\mathbf{Y}$ of the unknown state are available through the non-linear operator $\mathbb{H}$ up to $\boldsymbol{\epsilon}_{\mathbb{H}}$. These observations can either be issued from *real data* (in this context, we try to estimate some hidden parameters $\mathbf{X}$) but they can also be defined *by the user*. This latter situation is more related to a pure control process. To estimate the system's state, a common methodology relies on the minimization of the cost function that takes into account the three above equations. Unfortunately, for practical reasons due to the complexity of the dynamical systems and to the dimension of the system state, the direct minimization of such a cost function is in practice unfeasible. A known way to cope with this difficulty is to write *an adjoint formulation* of the problem. It can indeed be shown (see [10] for details) that recovering $\mathbf{X}$ can be easily done using a direct and a backward model integration, which considerably simplify the complexity.

This framework is then an appealing solution for large system states and complex dynamical models, as the ones we are likely to deal with a high number of

swarm particles submitted by continuous physical laws. However with swarm particles, the system state is Lagrangian (i.e. defined for particles) whereas the command can either be defined at a Lagrangian level (configuration of particles) or in an Eulerian context (any continuous data like the density). It is therefore of primary interest to design some mathematical tools to switch from Lagrangian/Eulerian spaces. The general idea is presented in the next section.

3 Graph-Based Data Representation

We consider the swarm as a connected graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ composed of a set of N particles, seen as vertices. The population $\mathcal{V} \in \mathbb{V}^N$ has interconnections, seen as edges $\mathcal{E} \in \mathbb{E}^Z$, where Z is the number of edges connecting two vertices of $\mathcal{V}$. One can easily understand that we are not interested in connecting every particle to all others, and therefore we aim at having $Z \ll N^2$. We obtain the definition of the graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ defining in a row particles and their interactions. For example, the position of particles, also called configuration, can be written as:

$$\mathbf{y}(\mathcal{V}) = [y(p_1), ..., y(p_i), ..., y(p_N)]^T = [y_1, ..., y_i, ..., y_N]^T, \tag{2}$$

where $y(p_i) = y_i$ are the spatial locations of the particles. In order to write graph quantity expressions, we introduce the diagonalisation operator $\underline{.}$ which creates a pure diagonal matrix made of the vector values. For instance, $\underline{\mathbf{f}}\mathbf{y}$ gives us a new graph quantity being the product of $\mathbf{f}$ and $\mathbf{y}$. The connectivity between the different vertices $\mathcal{V}$ is defined by the adjacency matrix $\boldsymbol{A} \in \mathbb{Z}^{N \times N}$ which captures the structure of the graph $\mathcal{G}$, i.e. $\boldsymbol{A}_{ij} = 1$ if there exists an edge between p_i and p_j and $\boldsymbol{A}_{ij} = 0$ elsewhere. Depending on applications, this adjacency matrix can also be balanced by weighted $w(f(p_i), f(p_j)) = w_{ij}^{\mathbf{f}} \in \mathbb{W}$ to promote some specific relations, which when combined to the adjacency matrix leads to $\boldsymbol{A}_{w^{\mathbf{f}}} = [w_{ij}^{\mathbf{f}}] \circ \boldsymbol{A}$, $\circ$ being the Hadamard, or entrywise, product. The aim being to deal, on the one hand, with a system state defined on such a graph and on the other hand, with a continuous dynamic model and eventually eulerian observations, we have to define a Lagrangian to Eulerian operator (able to extract continuous quantity from a graph) as well as an Eulerian to Lagrangian projector (that evaluates a continuous quantity on a graph).

Lagrangian to Eulerian projection. This operation can be viewed as the transcription of a graph of N particles on an other graph composed of M vertices, M being the dimension of the Eulerian grid (i.e. the number of pixels). With the help of an adjacency matrix, the operation consisting in deriving an eulerian value $\boldsymbol{\rho}_{\mathcal{C}}$ on a grid $\mathcal{C}$ from a graph $\mathbf{q}$ reads $\boldsymbol{\rho}_{\mathcal{C}} = \mathcal{A}_{\boldsymbol{G}^{\mathbf{y}}}\mathbf{q}$. In this case, $\mathcal{A}_{\boldsymbol{G}^{\mathbf{y}}} \in \mathbb{W}^{M \times N}$ is an Eulerian adjacency matrix such as $\boldsymbol{G}_{i,j}^{\mathbf{y}} = f(\mathbf{y}(i) - \mathbf{y}(j))$ is a kernel function, such as a gaussian one for instance.

Eulerian to Lagrangian projection. To evaluate an Eulerian quantity on a graph, several possibility are available. An efficient and simple, used in this paper, is the bilinear interpolation.

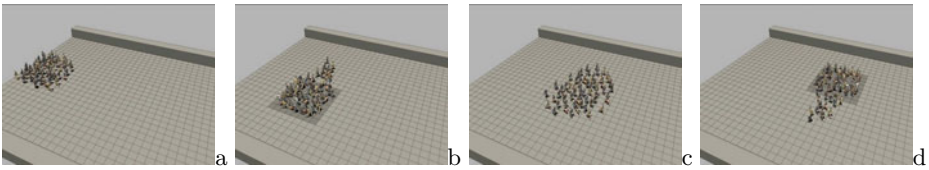


Fig. 1. Example: a classical crowd simulation model is used in this example. The crowd is given the goal to reach the opposite part of the corridor. Meanwhile, this dynamic model is controlled to fulfill two density constraints represented as a grey zone in figure (b) and (d), thanks to our methodology

The control of a Lagrangian system state defined on a graph submitted to a continuous dynamical model can then been managed by mixing the continuous theory of section 2 with the above operators able to switch from Eulerian/Lagrangian spaces. All technical aspects, and especially questions regarding the differentiation, are accessible in [11]. An example of a control with an Eulerian command is shown in figure 1 where a classical crowd model is controlled to reach two given density observations.

4 Conclusion

In this paper we have proposed an original framework to deal with the control of dynamic particle swarm using the variational assimilation theory with a system state defined on graphs. This theoretical framework is able to manage command/observations both issued from Lagrangian or Eulerian quantities.

References

1. Feddema, J., Schoenwald, D.: Decentralized control of cooperative robotic vehicles: Theory and applications. IEEE Trans. on Rob. and Aut. 18, 852–864 (2002)
2. Reynolds, C.W.: Flocks, herds, and schools: A distributed behavioral model. In: Proc. of SIGGRAPH 1987, pp. 25–34. ACM (1987)
3. Helbing, D., Farkas, I., Vicsek, T.: Simulating dynamical features of escape panic. Nature 407, 487–490 (2000)
4. Olfati-saber, R., Murray, M.: Consensus problems in networks of agents with switching topology and time-delays. IEEE T. on Aut. Cont. 49, 1520–1533 (2004)
5. Gazi, V., Fidan, B.: Coordination and control of multi-agent dynamic systems: Models and approaches. In: Swarm Robotics, pp. 71–102 (2006)
6. Martinez, S., Cortes, J., Bullo, F.: Motion coordination with distributed information. IEEE Control Systems Magazine 27, 75–88 (2007)
7. Olfati-saber, R.: Flocking for multi-agent dynamic systems: Algorithms and theory. IEEE Trans. on Automatic Control 51, 401–420 (2006)
8. Gazi, V., Passino, K.M.: Stability analysis of swarms. IEEE Trans. on Automatic Control 48, 692–697 (2003)
9. Diestel, R.: Graph Theory. Graduate Texts in Mathematics. Springer, Heidelberg (2005)
10. Lions, J.L.: Optimal control of systems governed by PDEs. Springer, Heidelberg (1971)
11. Allain, P., Courty, N., Corpetti, T.: Graph-based optimal control of particle swarm. IEEE Trans. on Automatic Control (submitted, 2011)

Research on Global Path Planning for AUV Based on GA

Yu Sun[1] and Rubo Zhang[2]

[1] College of Computer Science and Technology, Harbin Engineering University, Harbin, 150001, China
yusun0815@yahoo.com.cn
[2] College of Information, Guangdong Ocean University, Guangdong, 524088, China
zhangrubo@hrbeu.edu.cn

Abstract. Electronic charts as a specialized electronic map become increasingly mature. By using the electronic charts to establish the model of the environment and considering ocean current in GA search strategy, a new method for AUV global path planning in large-scale marine environments is proposed. The information provided by electronic charts is used in this paper to establish the model of the environment of AUV workspace, and GA as a search strategy is used for the AUV global path planning. The ocean current as a factor that affects the navigation of AUV was considered in the evaluation function of GA, so that the AUV can make full use of the energy of the currents. The model of the environment of AUV workspace and path planning is validated by simulation test.

Keywords: global path planning, AUV, GA, electronic chart.

1 Introduction

Global path planning technique is one of key issues in the field of intelligent robot. Global path planning refers to planning a best or better path from the beginning to the end in the known environment, which is a constrained optimization problem. Generally, AUV has several available paths to complete a given task, we have to choose a optimal (or near optimal) path under a certain criterion in practical applications. The common criterions include: the shortest path, the least energy consumption and the least time consumption etc. Global path planning includes two sub-problems: environmental modeling and path search strategy. The main approach of environmental modeling includes Visibility Graph method, free-space method and grid method. The main approach of path search includes A^* algorithm, Genetic algorithms, artificial potential field method and topology method etc [1].

AUV usually works in complex and variable marine environment. So the impact of currents on AUV is a problem can not be ignored [2]. “Free Area Approach Based on Grids”[3] has solved path planning for AUV in crowded three-dimensional underwater environment. In paper [4], Activation value propagation algorithm which solved path planning for AUV has taken the impact of ocean currents on AUV into account. This method combined grid method with the basic idea of configuration space according to the environmental structure information. The evaluation function considered both energy consumption and path length, which keep the heading of AUV opposite to the direction of the ocean current as possible. In paper [5], a hierarchical path planning

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 311–318.
springerlink.com

algorithm is put forward to solve path planning for AUV in wide range of marine environment. This algorithm considered energy consumption and path length as evaluation factors. The basic idea of this algorithm is to decompose the workspace of AUV into white region (obstacle area) and grey region (non-obstacle area), then find a path only through the white region. If the path passes through the grey region, then this region will be decomposed into smaller regions.

This paper proposed a method of environment modeling using electronic chart information, and genetic algorithm is used to search global path. The influence of ocean currents on AUV is considered in path planning. Experimental results show that the global path planned by this method have a better performance.

2 Analysis and Modeling Based on Electronic Chart

Electronic chart is a set of ordered set of geospatial data as a specialized electronic map with identified coordinates and attribute flags [6]. Electronic charts use discrete data to describe marine geographic information and navigation information. The electronic charts in shapefile format [7] are used in this paper, which contain more than 30 layers. We can choose to show the layers of ocean, obstacle, channel, regional boundaries and residential areas on the monitoring platform. Detailed data in electronic charts can be got by analyzing chart properties database and finding the corresponding graphics information in graphics files, then use GIS control—MapObjects to analyze this data.

Each place in charts has a certain property, such as land and sea. According to the needs of AUV in actual work, we prescribe impassable area and the restricted area that affects the safe navigation of the AUV as obstacle region, such as land, Islands, peninsulas and all kinds of beach, other area as free region. The chart is decomposed into many grids. The resolution of the chart is 18”. The square area of 18”*18” is considered as a grid node. Each grid node is done intersection test with the corresponding region in the chart, so that the property code of intersection area is got. Different properties have their own code. A gird can be determined to be obstacle region or free region according to the property code of the grid by consulting the corresponding property in shapefile electronic chart data dictionary. The well-converted grid files are used as chart information files in global path planning for AUV. Figure 1 is an electronic chart of a certain area, Figure 2 shows the grid of the electronic chart.

Fig. 1. Electronic chart of a certain area

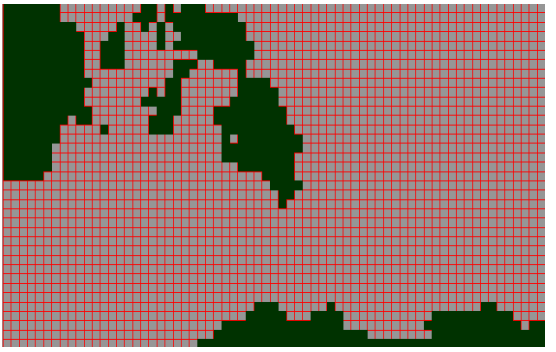


Fig. 2. Grid of the electronic chart

3 Global Path Planning for Auv Based on Ga

3.1 The Scheme of Path Encoding

A path from starting point to ending point consists of a series of grid nodes. Assume that the starting point is s and the ending point is d. Then the path can be expressed as $path = \{s, \ldots x_i, x_{i+1}, \ldots, d\}$, Adjacent nodes are connected by straight line. The obstacle grids are considered as impassable points. Assume the velocity vector of ocean current is $v_c(x, y)$. According to the former assumptions, the definition of path planning can be stated as follows: Given that the following information is known: the starting position, ending position, current velocity, AUV velocity and the position of obstacles. the target of path planning is to find a safe path which minimize the energy consumption of AUV by utilizing the ocean current. We prescribe that the path in X-axis direction is strictly monotone increasing, that is $h_{i+1} = h_i + 1$. It is clear that not all paths are monotonous without roundabout. But in a wide range of marine environments, the obstacles are quite sparse, so it is reasonable to assume that the path is monotonous in environment with few obstacles.

3.2 Determine the Fitness Function

Each path planned by algorithm need to be evaluated. Here energy consumption is considered as the fitness, we need to calculate and accumulate the energy consumption of AUV in Each segment of the path with the influence of the current. Then the cumulative result can be used as the fitness function of the path. Suppose the ith segment of a path is connected by node xi-1 and node xi, denoted as $X_{i-1}X_i$. And the unit vector in direction of $X_{i-1}X_i$ is e_i, e_i is the direction of AUV in actual navigation. Assume the velocity of AUV in actual navigation is c, the velocity vector of current in a certain position (x, y) in line segment $X_{i-1}X_i$ is $v_c(x, y)$, the velocity vector provided by AUV itself is $v_i(x, y)$. Then the following relationship can be obtained:

$$v_i(x, y) = ce_i - v_c(x, y) \quad (x, y) \in X_{i-1}X_i \tag{1}$$

When AUV navigates along the direction of currents at a speed of c , the velocity provided by AUV itself is less than c, so that the energy consumption is saved with the speed of ocean current. When against the direction of current, to navigate at the speed of c, AUV needs to overcome the influence of the current and provide a speed higher than c, which consume more energy of AUV. In the other words, if the velocity provided by AUV is constant, navigation along the direction of current will provide AUV with a higher speed. So AUV could cover a longer distance with the same energy consumption. The navigation distance is increased by saving the energy. On the contrary, if AUV navigates against the direction of current, the navigation distance will be shorter with the same energy consumption, which need to consume more energy to cover the same distance.

So the navigation of AUV should make full use of currents to reduce the energy consumption, and avoid the navigation which is against the direction of current, or to go against the current of which the velocity is lower as possible.

The evaluation function of energy consumption is as follow:

$$fit_1 = \sum_{Xi-1Xi} \frac{k}{\sqrt{(v_x \bullet d_x)^2 + (v_y \bullet d_y)^2}} \tag{2}$$

where v_x and v_y is the sub-speed of AUV along the direction of X axis and Y-axis respectively, and d_x and d_y is the navigation distance along the direction of X axis and Y-axis respectively. k is a constant. It can be seen from equation (2) that the fewer energy AUV consumes, the higher the fitness value is.

If the path pass through a obstacle, the grids in the obstacle is assigned a punishment value, which is defined as follow:

$$fit_2 = \begin{cases} 1, if\ \ no\ \ obstacle \\ 0, others \end{cases} \tag{3}$$

The final evaluation function is:

$$fit = fit_1 \times fit_2 \tag{4}$$

3.3 Genetic Operators

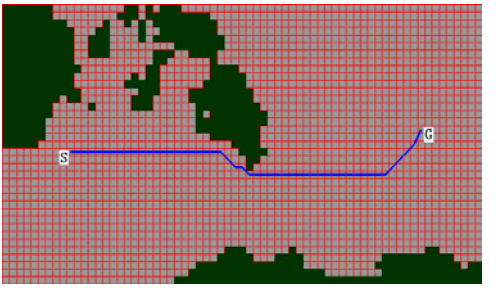
1) Selection operator: the population is ranked according to the fitness, this operator selects the individual with high fitness to procreate the next generation, the individual with low fitness is replaced by that with high fitness.

2) Crossover operator: Every two individuals are paired for two-point crossover operation. This operator compares the fitness of the individual after crossover with its parent, if the fitness of parent is lower than that of offspring, the parent is replaced by its offspring. Else the parent is still retained.

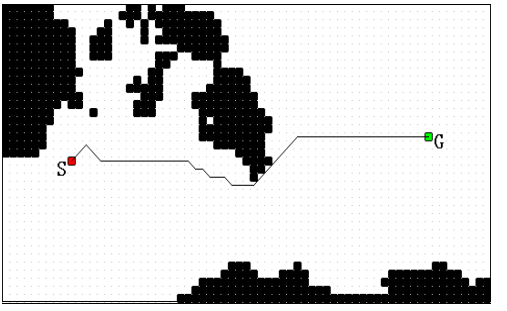
3) Mutation operator: A small part of individuals are mutated with a certain probability of mutation. To ensure that the excellent individual is not lost inadvertently during the mutation, we prescribe that the part of the best individuals are not mutated.

4 Experimental Results and Analysis

In figure 3, the starting point *S* in navigation is (120° 37′37″ E, 37° 53′49″ N); the ending point *D* is (120° 52′16″ E, 37° 54′36″ N). The initial population of GA is 100 individuals, The termination conditions of GA is the iteration for 800 generation. The probability of crossover is 0.8, the probability of mutation is 0.4. Assume the ocean current is weak, we don't take the influence of current into account. The planning result is shown in figure 3(a). Figure 3(b) shows the result planned by A* algorithm. The starting point and the ending point are the same with figure 3(a), the heuristic information is the optimization criterion of shortest path.



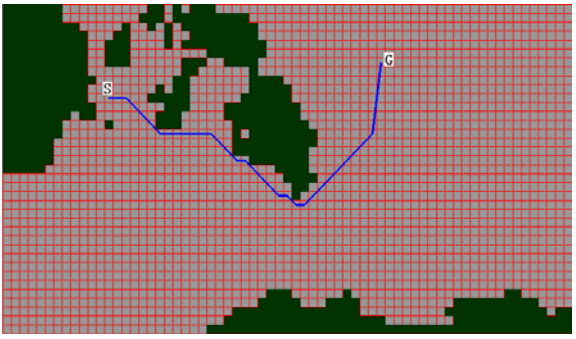
(a) GA



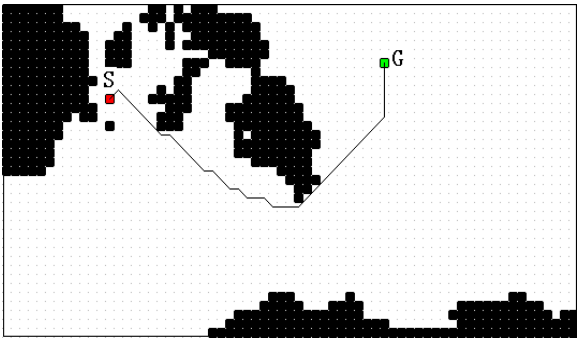
(b) A* algorithm

Fig. 3. Global path planning test 1

In figure 4, the starting point *S* is (120° 38′24″ E, 37° 56′24″ N); the ending point *D* is (120° 48′0″ E, 37° 57′36″ N). The results planned by A* algorithm and by GA are shown in figure 4(a) and figure 4(b) respectively.



(a) GA



(b) A* algorithm

Fig. 4. Global path planning test 2

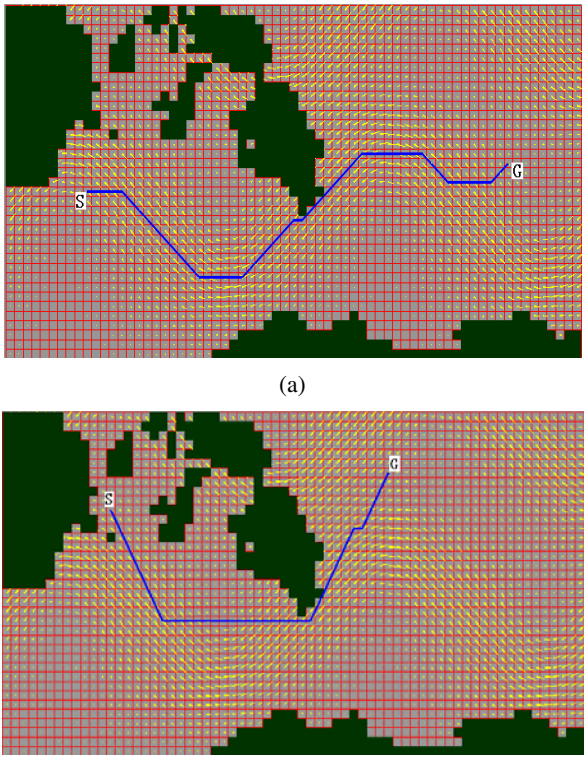
Table 1 shows the comparison of results between two methods of the global path planning for AUV, regardless of the current factor in the case of the weak ocean current.

Table 1. Compare of the Results

	algorithm	The length of the path(km)	The number of inflection point	The planning time(sec)
test 1	GA	26.16	5	3
	A*	27.27	10	0.5
test 2	GA	23.64	10	3
	A*	24.68	12	0.5

It can be seen from above results that the length of the path planned by GA is shorter than that planned by A* algorithm, and the planning result of GA is smoother than that of A* algorithm, which is reflected in the fewer inflection points. But the planning time of GA is longer than that of A* algorithm. The global path planning has been done before the navigation of AUV, so it is not strict with the requirement of the planning time.

If the ocean current is strong in marine environment, the current factor can't be ignored. The planning result of GA with ocean current is shown in figure 5. The size and direction of ocean currents is indicated by arrows.



(a)

(b)

Fig. 5. Global path planning with ocean current

As shown in experimental results, AUV could navigate along the direction of the current as possible to make full use of the energy of currents. So that the distance of navigation will be longer by consuming a certain energy.

5 Conclusion

The paper proposed a method of environment modeling based on the environmental information provided by electronic chart which is utilized in the global path planning for AUV, and solved the problem of environment modeling in large-scale marine environment. GA is used as the search strategy of the global path planning, and compared to A* algorithm. The path planned by GA has the advantages of the shorter length and the fewer inflection points. If it is necessary to consider the influence of

current on AUV, the navigation of AUV will make full use of the energy of current. This method is applied to the navigation of AUV with the consideration of energy consumption in large-scale marine environment.

References

1. Zhang, Y., Wu, C., Yuan, B.: Progress on Path Planning Research for Robot. Control Engineering 10(21), 152–155 (2003)
2. Alvarez, A., Caiti, A., Onken, R.: Evolutionary path planning for autonomous underwater vehicles in a variable ocean. IEEE Journal of Oceanic Engineering 29(2), 418–423 (2004)
3. Du, M., Gu, G.: Path Planning with Free Area Approach Based on Grids. Journal of Harbin Engineering University 16(2), 49–53 (1995)
4. Dong, X., Gu, G.: AUV global path planning allowing for ocean current. In: Proceedings of the 3rd World Congress on Intelligent Control and Automation, Hefei, China, pp. 1230–1234 (2000)
5. Zhang, Q., Zhang, B.: The Application of Genetic Algorithm in Path Planning for AUV. Control and Automation 22(11), 240–243 (2006)
6. Liu, Z., Yu, Y.: The production of digital chart and geographic information system of chart (1), 6–10 (2001)
7. ESRI Shapefile Technical Description- An ESRI White Paper. Environmental Systems Research Institute, Inc., USA (1998)
8. Han, P.: The development of geographic information system—a method of MapObjects. Wuhan University Press, Wuhan (2004)

Research on Operation Mechanism of Enterprise Technology Innovation

Dong Peiwu and Lin Xiaohong

School of Management and Economics,
Beijing Institute of Technology, Beijing, China
lxhxpz1984@163.com

Abstract. Operating mechanism is the most complex mechanism in the enterprise technology innovation. And it plays a crucial role in promoting the technology innovation process. By analyzing the elements which impact on the technology innovation of enterprises operating, the paper presents the operation structure model of enterprise technology innovation based on ISM (Interpretative Structural Modeling). According to that, this paper derived enterprise technology innovation operating mechanism model and structure which is conductive to innovation activities effectively.

Keywords: enterprise technology innovation, operation mechanism, ISM.

1 Introduction

Operating mechanism of the technology innovation refers to the relationships among all the elements and the roles which they play in the technological innovation activity. Operating mechanism of the technology innovation covers the whole process from investment of the enterprise, the internal activities to transforming into commodities. Therefore, it is very important to improve the ability of technological innovation.

2 ISM Model

ISM (Interpretative Structural Modeling) is a structured modeling technique that developed by professor J.Warfield[1] from United States in 1973, this method is mainly used for complex Socio-economic system problems analysis, it decomposes complex system into several subsystems (elements) by using people's practical experience and knowledge, so that the complex relationship among many elements can be laid and principled, thus show the internal structure of the system and the dependency between the elements, eventually to construct the system into a multilevel hierarchical structure model[2].

3 Setting Up the Technological Innovation Elements

As the operation mechanism of technological innovation presents for the whole process of technological innovation activities, the technology innovation is so closely related to the impact of internal and external factors.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 319–325.
springerlink.com

(1)Government Intervention (S_1): The government intervention refers to the macro-economic policies, legal measures and the relevant government agencies conduct.

(2)Technology Development (S_2): The environment of technology is composed of technology level, science level and technology system. [3]

(3)Market Environment (S_3): The market which could achieve the ultimate innovation is composed of market demand and market competition.

(4)Market Evaluation (S_4): It means the investigation of new ideas in potential customers, distributors and retailers. [4]

(5)Intellectual Property Protection (S_5): To the point of view, it is the fundamental countermeasures of the overall plan and goal to achieve innovative.

(6)Technology Infrastructure (S_6): Clayton Christensen defined the technological innovation as the development of new technology resources and skills. [5]

(7)Technology Intermediary (S_7): Technology Intermediary provides the technical exchange environment for innovation activities.

(8)Innovative Consciousness of Entrepreneur (S_8): Entrepreneurial innovation directly affects the behavior of enterprises in technological innovation.

(9)Innovation Fund (S_9): Enough money to get into innovative R&D resources, so as to make the business more successful innovation activities. [6]

(10)Innovative Advantages (S_{10}): Innovation advantages are composed of technical innovation and talent advantage.

(11)Project Decision-Making (S_{11}): the use of the information obtained to do scientific decision-making. [7]

(12)Import, Absorption and Innovation (S_{12}): A type of technology innovation.

(13)Cooperation Innovation (S_{13}): It is used to describe the research cooperation.

(14)Original Innovation (S_{14}): Enterprises mainly rely on their own resources, human resources and technology to research and develop.

(15)Risk Aversion (S_{15}): It is more conducive to business decision-making.

(16)Innovators (S_{16}): Innovation and development of talent is the core issue.[8]

(17)Staff Incentives (S_{17}): Companies need to establish an effective incentive mechanism of the enterprise innovation.

(18)Manufacturing Capacity (S_{18}): It determines the ability to implement technological innovation.

(19)Marketing Capability (S_{19}): It is the ability to sell the technology innovations to market.

(20)Technology Innovation Strategy (S_{20}): It has become the fundamental components of development strategy.

(21)R&D Institutions (S_{21}): The organizations of technological innovation [9].

(22)Innovation Benefit (S_{22}): It is the profit of technological innovation.

4 Construct Operation Mechanism Model of Enterprise Technology Innovation Based on ISM

4.1 Establish ISM Model Steps

(1) Based on the above identified 22 risk factors, again discussed by the experts to determine the relationship between them; and in accordance with the binary relations between elements, as shown in Table1.

Table 1. Relationship between the Elements

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1
2	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0
3	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
4	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
5	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	1
6	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
7	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
9	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	1	1
10	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1
15	0	0	0	1	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	1	0	0
16	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	1	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
20	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0		0	0	0	0	0	1
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(2) Establish reachable matrix. Using Boolean operations, get $(A+1)^{k-1} \neq (A+1)^k = (A+1)^{k+1} = M$, that is $A_k = A_{k+1}$, so reachable matrix is $M = A_k = A_{k+1}$ [10].

(3) Determine the reachable set and advanced set, carry out inter-class division. Set *R(Si)* is a reachable set of S_i, which is combined by all the factors that in all lines factor equal to 1 in the S_i row of matrix *M*, that is $R(S_i) = \{S_i \in N / R_{ij} = 1\}$. *A(Si)* is the advanced set of S_i, which is combined by all the factors that in all rows factor equal to 1 in the S_i line of matrix *M*, that is $A(S_i) = \{S_i \in N / R_{ij} = 1\}$ [11]. Identify the reachable and advanced set of 22 risk factors in turn, as shown in Table 2.

Table 2. Reachable Matrix

i	22	18	19	12	13	14	10	11	20	21	4	6	8	9	16	1	2	3	5	7	15	17
22	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
11	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
20	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
21	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
4	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
6	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0
8	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0
9	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0
16	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	0	0	0	0	0	0
2	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	1	0	0	0	0	0
3	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	1	0	0	0	0
5	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	1	0	0	0
7	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	0	1	0	0
15	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	1	0
17	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	0	0	0	0	0	1

(4) Reachable matrix inter-class division[12] .According to the calculation, the first advanced elements set is S_{22}, L_1={ S_{22}}, cross out L_1 from reachable set, get new reachable and advanced set, and elements can be divided into different levels step by step, we get L_1={ S_4,S_6 ,S_8,S_{18}}. Therefore, the original risk factors can be divided into six levels from top to down.

(5) ISM model establishment. According to the result of reachable matrix, the structure model is established, as shown in Fig. 1.

4.2 Analysis on Operation Mechanism Model of Enterprise Technology Innovation

The operation mechanism model reflects the relationship among the 22 elements which impact on the technology innovation comprehensively and systematically. The model is divided into six levels from top to bottom.

Government intervention, technological development, marketing environment, IPR(intellectual property protection), technology intermediary, risk aversion and staff incentive are at the bottom of the enterprise independent innovation structural model, as shown in Fig.1. Market evaluation and funds of innovation are directly affected by the government intervention policy, and technical development has direct influence on market evaluation, technological base and innovators.

In addition, market evaluation and innovative conscious of entrepreneurs is influenced by marketing environment. IPR has effects on market evaluation, technological base, innovative consciousness of entrepreneurs and innovators. Technical intermediary has direct influence on market evaluation, technical base, innovation funds and innovators. Incentive of staff has effects on training of innovators. Furthermore, results of market evaluation, innovative consciousness of entrepreneurs and accumulation of innovation funds are affected by risk aversion.

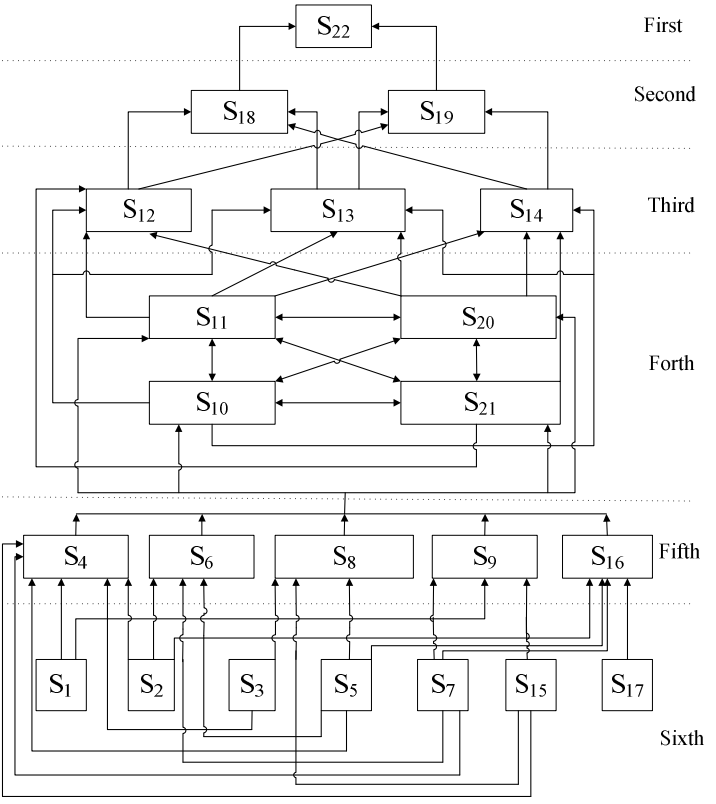


Fig. 1. Operation Mechanism Model of Enterprise Technology Innovation

Construction of advantage about enterprise innovation, project decision-making, technological innovation strategy and R&D behavior are directly affected by market evaluation, technical base, innovative consciousness of entrepreneurs, innovation funds and innovators. However, preponderant, project decision-making, technical strategy and R&D behavior depends on each other as well as affecting the R&D behaviors such as innovation based on technology import, cooperation among industries, universities and research institutes and original innovation. All of the three types of R&D activities have influence on productivity, marketing management capability, which finally constitutes the creative profits.

Innovation based on technology import, cooperation among industries, universities and research institutes and original innovation has direct influence on productivity of enterprise. It is very important that enterprise manufacture and marketing policy to be the second level of the structure. In the intersection of the structure, such as internal and external construction are affected by government intervention, only the government policy is good for enterprise technical innovation, can enterprise technical innovation process be boosted by enterprise and external environment. R&D activities and production is affected by technical environment. Market environment is composed of market requirement and market competition. Market requirement is the source of enterprise technical innovative consciousness and market competition is the essential factor considered by entrepreneurs in technical innovative activities. Technical environment, technical intermediary have influence on effectiveness of R&D process, and innovative consciousness of entrepreneur is affected by IPR environment.

The model can be divided into five aspects by analyzing and summarizing the whole technical innovative process based on the reasoning results, as shown in Fig.2, which constitute the enterprise technical innovation operation mechanism.

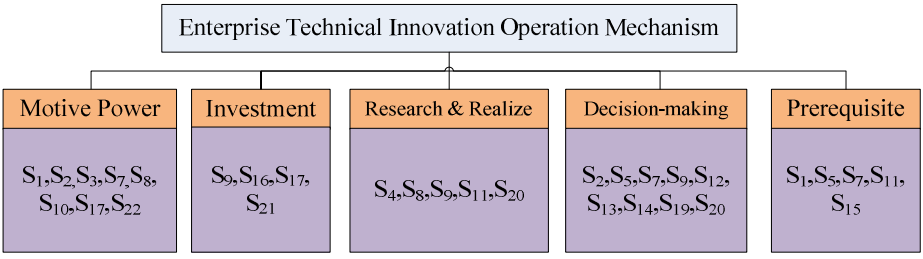


Fig. 2. Operation Mechanism of Enterprise Technology Innovation Structure

Conclusion

Analysis on the enterprise technology innovation operating mechanism by using ISM confirms the relationship and influence levels of these elements which impact on the enterprise technology innovation. Based on the results of the model and structure of enterprise technology innovation operating mechanism, internal factors and external factors depend on each other in the way that: firstly the government should construct environment with complete policies and regulations, maintain a stable political environment and economic system; secondly, the internal factors should ensure the operation of innovation, and should be combined with external factors to accelerate the technology innovation; thirdly, the government should play the role of supporting intermediary organizations, increasing innovation funds, standardizing the operations, so as to play the role of bridge to boost technology innovation. According to the function and relationship of the factors, the model can be divided into five aspects by analyzing and summarizing the whole technical innovative process: motive power, investment, research & realize, decision-making and prerequisite.

References

1. Warfield, J.N.: Participative Methodology for Public System Planning. Computers & Electrical Engineering 1(1) (1973)
2. Wang, Y.L.: Systems Theory, Methods and Applications. Higher Education Press (1998)
3. Casper, S., Whitley, R.: Managing competences in entrepreneurial technology firms: a comparative institutional analysis of Germany, Sweden and the UK. Research Policy 33, 89–106 (2004)
4. Maurer, S.D., Zugelder, M.T.: Trade secret management in high technology: a legal review and research agenda. The Journal of High Technology Management Research 11, 155–174 (2000)
5. Sparrow, P., Hiltrop, J.M.: European Human resource Management in Transition, pp. 124–125. Prentice Hall, London (1994)
6. Huergo, E.: The role of technological management as a source of innovation: Evidence from Spanish manufacturing firms. Research Policy, 1377–1388 (2006)
7. Cainelli, G., Mancinelli, S., Mazzanti, M.: Social capital and innovation dynamics in district-based local systems. The Journal of Socio-Economics 36, 932–948 (2007)
8. Fei, L., Zhang, L., Hua, L.K.: High-tech Zone of independent innovation system operating mechanism. Technology and Management 1, 20 (2008)
9. Gong, J., Wang, F.D.: On the small high-tech enterprises operating mechanism of technological innovation. Economic Issues 9, 40 (2000)
10. Bo, L., Tianjun, H., Li, L.: Systems engineering, 1st edn., vol. 1, pp. 15–16. Northern Jiaotong University Press (2003)
11. Yang, J., Guo, L., Sen, L., Jingyan, S.: Introduction to systems engineering, vol. 1, pp. 153–158. Wuhan University of Technology Press (2002)
12. Wang, Y.-L.: Systems engineering theory, methods and applications, 2nd edn., pp. 49–81. Higher Education Press

Evolution Research on Organization and Management Mode of Major Scientific and Technological Projects

De-cheng Kong, Guang-ming Hou, and Jun-peng Wang

School of Management and Economics, Beijing Institute of Technology,
Beijing, 100081, China
decheng-kong@163.com

Abstract. In this paper, taking some typical major scientific and technological projects as examples, we analyzed the characteristics and operation mechanism of major scientific and technological projects operated in different historical periods. According to the role of government in the project management process, the organization and management mode of major scientific and technological projects could be divided into three basic types: government master mode, enterprises to participate in government-led model and model of enterprise as the main under the government guidance. Through analysis, the main factors that affect selection of organization and management mode were economic system, technical level and engineering target. By farther analysis, it is found that organization and management mode characterized by enterprise as the main under the government guidance would be the main mode in the future.

Keywords: major scientific and technological project, organization and management, management mode, evolution path.

1 Introduction

Major scientific and technological projects (MSTP) are different from general mass production of goods. Their implementation require the core technology research, and engineering products are important to the economic and social development, national security, international status and they belong to complex product and system(CoPS). The complexity of such systems is mainly reflected in the number of custom components, span of knowledge and skills, and the extent of new technologies. High cost, strong coupling, intensity of information and technology, customer customization, characterize projects. Since the founding of New China, we has organized and implemented a series of major scientific and technological projects, such as "bombs and one satellite" project, manned spaceflight, lunar exploration, China-made large aircraft programs. They need not only core technology breakthroughs but also organization and management mode that are scientific and reasonable, which is decided by essence. Only in this way, can we enhance project success rate, improve project quality and shorten the development cycle. In this paper, we discussed the characteristics and operation mechanism of major scientific and technological projects operated in different historical periods and extracted the factors affecting the selection of project organization and management mode, which would provide reference to the organization and management of major scientific and technological projects in the future.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 327–334.
springerlink.com

2 Evolution Path Analysis of Organization and Management Mode of Major Scientific and Technological Projects

Since the 1950s and 1960s, China has organized and implemented MSTP. Great achievements has been made, such as bombs and one satellite project which making China to be the fifth country that possession nuclear weapons in the world, the success of manned spaceflight project make China the third country which master space walk technology, and China-made large aircraft program's success will make our aviation industry as good as Boeing and Airbus. These achievements have not only enhanced our defense capabilities greatly but also actively promoted and will continue to promote the China's economic construction. Taking the organization and management mode, which adapt with the times is important to ensure the projects successful and efficient. In recent decades of China, major scientific and technological projects organization and management model can be divided into three basic types: government master mode, enterprises to participate in government-led model and model of enterprise as the main under the government guidance according to the role of government in the project management process.

2.1 Government Master Mode

2.1.1 Characteristics of Government Master Model

China has successively organized and implemented a series of typical MSTP, such as "bombs and one satellite" project, intercontinental rockets, satellite communications and so on between 1960s and 1980s. These projects had been managed directly by core leaders of the nation (such as "bombs and one satellite" project is been managed directly by Premier Zhou [1]). The overall design department was established around the goal of the project for target decomposition and co-ordination, and the administrative organization on the implementation of the paramilitary management, the experts responsible for technology, then the administration and technology coordination with each other. The organization and management mode stated above came to be known as "one general design department, two command-line" project management system (Fig. 1), and which led to the theory of systems engineering formation and development in China [2].

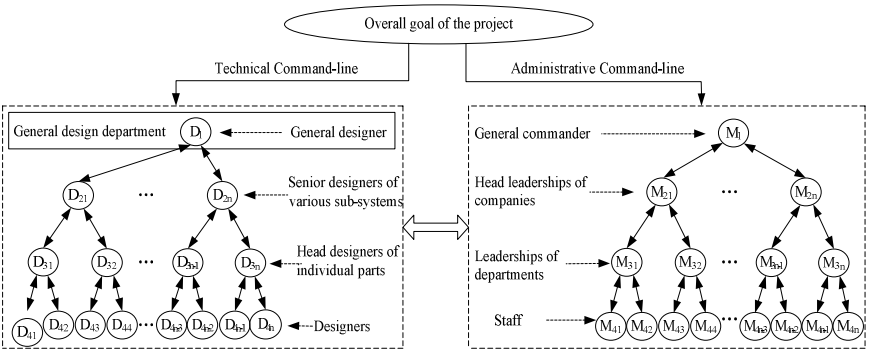


Fig. 1. Systems engineering management diagram

2.1.2 Operation Mechanism of Government Master Model

Operation mechanism of government master mode means "one center, two command-line" systems engineering management. The general design department is responsible for the overall design and system coordination. Two command-line indicate the technical command line in charge of general designer and the administrative command line in charge of general commander. Technical command line is composed of general designer, senior designers of various sub-systems, head designers of individual equipment and parts and the other designers. Administrative command line is composed of general commander, head leaderships of each research department, planned control systems and the relation staff of function department in machinery. "Two command line" work as a organically linked to centralized and flow-sensitive production management development by division of labor with individual responsibility, mutual penetration and mutual support. At present, the project management model of "one center, two command-line" has become a symbol of management with Chinese characteristics in complex major systems engineering.

Take satellite communications engineering for example, it is composed of five systems, the launch vehicle, satellite body, launching site, monitoring and control network and ground station system. Each one of the five systems has independent technical and administrative command systems. For example, in launch vehicle system, the technology commander is Xie Guangxuan who is the chief architect of the Long March III. The administrative commander is Miao Fengchen who is the vice president of Academy of Launch Vehicle. Technical commander of satellite body is Sun Jiadong who is the general designer of the satellite. The administrative commander is Liu Chuanshi who is the president of Space Technology Institute. The whole communication engineering have another overall command system, the general technology commander is Ren Xinmin who is the overall engineering designer and the administrative commanders are Chen Bin and Ma Jie Miao who are the leaderships of National Defense[3]. The technical command line and the administrative command line make departments accountable, compact development steps, human, financial and efficient organization.

2.2 Enterprises to Participate in Government-Led Model

2.2.1 Characteristics of Enterprises to Participate in Government-Led Model

China has successively organized and implemented a series of MSTP, such as the technology project of launch vehicles, manned spaceflight and so on between 1960s and 2000s. Based on the "one center, two command-line" systems engineering organization and management model, these works take a special management manner which focus on special organization and implementation of country-led, the overall responsibility of the relevant ministries, and actively absorb large enterprises involved in the development. These research institutes and companies constitute a cross-industrial, highly centralized organization and management system in accordance with engineering science and technology processes and functions.

2.2.2 Operation Mechanism of Enterprises to Participate in Government-Led Model (Taking Manned Space Projects for Instance)

Chinese Party Central Committee and the State Council lead Chinese manned space project. And it is directly leaded by the General Armament Department, Ministry of Industry and Information Technology (under the National Defense Science and Industry Council), Chinese Academy of Sciences and the China Aerospace Science and Technology Corporation and other industries. At the operational level, General Armament Department is responsible for four systems includes astronauts, launching site, monitoring and control communication, and landing site and a number of their corresponding sub-systems. Chinese Academy of Sciences is responsible for spacecraft application, China Aerospace Science and Technology Corporation is responsible for the manned spacecraft, launch vehicles, space laboratory system and the corresponding sub-systems. The system and the corresponding subsystems are set up corresponding overall department. The two vertical lines of general commander and designer run through from top to bottom, and manned space-engineering office at all levels work in horizontal management. Departments at various levels fix their positions and responsibilities, and weave into a matrix-style organizational system and network together. In coordination mechanism, organization and management principles are combined policies with management, administrative management with technical responsibility, decentralized management and unified coordination. The management model diagram as follows:

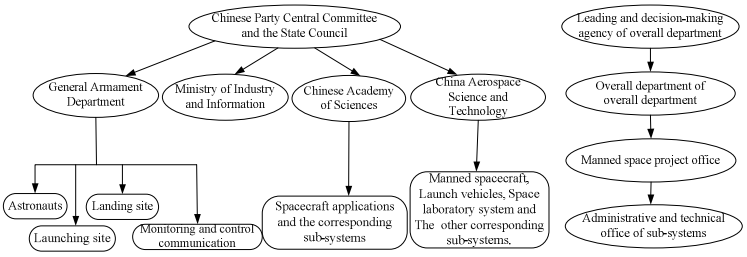


Fig. 2. Diagram of enterprises to participate in government-led model (taking manned space projects for example)

2.3 Enterprise as the Main under the Government Guidance Model

2.3.1 Characteristics of Enterprise as the Main under the Government Guidance Model

In 2006, the State Council issued the "Long-term Scientific and Technological Development (2006-2020)", that established 16 major science and technology special projects. Some of them have made substantial progress and organization and management approach is mainly reflected in two modes: one is taking the "one center, two command-line" organization and management mode in project development, and in product manufacturing stage and application phase is relies on large companies or professional firms to form the engineering products industrial chain, such as high-resolution earth observation system. The other one is that a large holding company is set up by the nation, then the project will be work directly by the holding company,

the government only play a supervisory role and not responsible directly. The company operates in market-oriented to pursuit of economic interests based on ensuring national interests. "One center, two command-line" systems engineering organization and management mode is completely to the company inside, such as China-made large aircraft programs.

2.3.2 Operation Mechanism of Enterprise as the Main Under the Government Guidance Model (Taking China-Made Large Aircraft Programs for Example)

In China-made large aircraft project, the organizational model is innovated, and ACAC is established. As the newly established central business, the shareholder composition has been a breakthrough. This form of organization is taken into account in R&D, manufacturing, marketing, airworthiness certification for the shareholder to play their role, to create a good external environment for the management. In the operating mechanism, the company implements the "one center, two command-line" systems engineering organization and management model around the large aircraft program, chairman and general manager of the company is the Director and Deputy Director of National Defense respectively. They are in charge of administrative commander and sectoral coordination. WU Guang-hui who was the ARJ aircraft general designer will served as the vice-general manager of the joint-stock company, and worked as general designer responsible for the technical command [4]. There are three subsidiary integration platforms: the design and development platform, the platform assembly manufacturing and service support platform [5]. This governance structure is in accordance with the systematic management principles of respective merits, cooperation; each performs its responsibilities and unity of purpose. That is helping to eliminate barriers in the previous organization and streamline relationship between the parties.

3 Evolution Factors Analysis of MSTP Organization and Management Model

China has organized and implemented a series of MSTP in different historical periods, and the scientific, reasonable and timely organization and management mode played a pivotal role on project's success. A lot of factors affect the organization and management model selection, including economic structure, technological level and the goal of the project which are the main factors in mode selection. Therefore, the evolution path of MSTP organization and management model can be explored through these three elements.

3.1 Institutional Factors

China has been in the planned economy era and implemented the planned economic system before 1990s. Economic system still followed the planned economic system in spite of the reform and opening up policy had been raised at Third Plenary Session. All domestic enterprises are public-owned, essentially a "factory" and do not have autonomous rights. Under this institutional situation, organization and implementation of MSTP must be directly involved in national command and control, which is the government master mode.

From 1992 the research institutes have implemented restructuring in pace with the market economy system establishment. The organizational systems of the previous research become into production and business organization systems existed on the form of company, such as the First Office of a branch of the Old Fifth Institute gradually transformed in to company, and ultimately form Aerospace Corporation [6]. However, the companies are not mature by the restructuring of research institutes and there are many problems. Based on this, when the MSTP is organized and implemented by state, the companies are encourage to participate in and will be given some scientific research and production tasks that adapt to their abilities. That is the model of enterprises to participate in government-led.

China's socialist market economic system has been further established since the twenty-first century, the restructuring of enterprises have increased steadily more market-oriented, corporate governance institutions have been further optimized, the field of defense industry enterprises and the civil companies gradually formatted a cooperative win-win situation. Based on this, when the MSTP is organized and implemented by state, it is rely on large group organization and management, and the role of government change gradually from direct leader to supervises the inspector. Then it is formed the enterprise as the main under the government guidance model.

3.2 Technical Factors

In fifties of the twentieth century, China's military technology was relatively backward and there were almost no stock resources of cutting-edge technology. And the development of civilian technology is far behind military technology. In 1980s, China's national defense science and technology has made great improvement. In this situation, government should guide MSTP.

Civilian technology had made great progress since 1990s. There were some reasons for it: on the one hand, the transfers of High-end technical talent promote the horizontal transfer of technology. Some research institutes became large groups, which had autonomy. On the other hand, the quantity of high-tech technical talent increased rapidly. For example, up to 2001, there were 21.69 million professional, 2.07 million scientists and engineers. There was 0.95million population who joined in research and development (R&D), and 0.74million was scientists and engineers. Compared to 1991, the rate raised by 42.6% and 57.8% [7]. These talents had made civilian technology be improved greatly. Under this situation, civilian companies should be encouraged to take part in the organization of MSTP and enterprises to participate in government-led model should be put into use.

In 21st century, the world's new technological revolution has been growing more rapidly and information technology has become an important engine that drives economic growth and dissemination [8]. Many organizations prefer to develop sophisticated technology and emphases technology is irreplaceable. Global production has become a new mode of production and for this cause any cannot complete a program by itself. Mode of enterprise as the main under the government guidance is put into use in the organization and management of MSTP.

3.3 Objective Factors

Since the beginning of the founding of China, the goal of MSTP is to improve the country's defense capabilities and to eliminate the military threat to Western countries. Under the guidance of this goal, the management of MSTP is paramilitary and the military generals of country's core leadership were directly responsible for it. Implementation and management of the project was pure government action, and this model would help to gather all the necessary resources and achieve goal of the project in the minimum time.

Since the 1990s, Military and civilian-style development trend gradually arised. Meanwhile, research institutes, which were led by government, began to integrate resources, and became stakeholders of market economy, for example, China Aerospace Science and Technology Corporation was evolved by the fifth Research Institute of Department of Defense [9]. To achieve military and civilian-style development, companies should be encouraged to join MSTP.

In the new century, ideas of integrating the armed forces and civil-military integration deepened further in order to stimulate economic development and fully enhance the "military + financial resources" integrated defense capability and ultimately relived the wealthy and powerful military defense strategy [10]. The organization of high-resolution earth observation system and China-made large aircraft program were a concrete manifestation of this typical idea. Under the guidance of above thinking, large national holding companies were established, which were used to be platform for management of MSTP, so mode of enterprise as the main under the government guidance was put into work.

4 Conclusion

The success of major scientific and technological projects that were operated in different historical periods not only enhanced our defense capabilities but also promoted China's economic development. As a Systems Engineering in itself, the success was based on breakthrough of cutting-edge technology, organization, and management model, which was scientific timely. With the development of China's economic system, technology level, target of engineering and so on, organization and management mode of major scientific and technological projects had changed from government master mode, enterprises to participate in government-led model to mode of enterprise as the main under the government guidance, today. With the further deep implementation of China's market economy system, labor specialization and internationalization, and implementation of "integrating the armed forces, civil-military integration" type of development strategy, mode of enterprise as the main under the government guidance will be the main mode used to guide major scientific and technological projects in the future.

References

1. Huang, C.-P., Hou, G.-M.: Manned space launch vehicle system development management. Science Press, Beijing (2007)
2. Chen, M.-Z.: Aerospace systems engineering and overall design department-Mr. Qian's contributions to our country in the system engineering theory and practice of contribution. Qian Xuesen Scientific Contribution and Academic Thought Seminar Collection, 225–228, http://d.g.wanfangdata.com.cn/Conference_3409189.aspx
3. Zhang, J.: Contemporary China's Space Industry, p. 440. China Social Science Press, Beijing (1986)
4. http://money.163.com/11/0412/09/71E9PLT700253B0H_5.html
5. http://www.comac.cc/nljs/sdpt/fwzc/index.shtml
6. Liu, J.-D.: The practice of systems engineering for the development of China's missiles and rockets Inspiration Based on "bombs and one satellite". China's First High-Level Forum of Historical Research Paper Bombs and One Satellite, 222–227, http://d.g.wanfangdata.com.cn/Conference_7223356.aspx
7. Sun, L.-J.: Flow of Scientific and Technological Personnel in China and Related Environment. Impact Science on Society 1, 10–13 (2004)
8. Hu, J.-T.: Stick to the Road of Independent Innovation with Chinese Characteristics to Building an Innovative Country and Work Hard - in the National Science and Technology Conference Speech. Qiushi 2, 3–9 (2006)
9. http://www.spacechina.com/jtgk.shtml
10. Hou, G.-M.: Military-Civilian Technology Transfer: Organization and Policy Research. Sciences Press (2009)

Simulation and Control of OWC Wave Power Generation System*

Hongwei Fang[1], Jiajia Chen[1], Yongqin Ren[2], Ning Wang[2], and Lin Cui[2]

[1] School of Electrical Engineering and Automation, Tianjin University, Tianjin 300072, China
[2] National Ocean Technology Center, Tianjin 300112, China
Hongwei_fang@tju.edu.cn

Abstract. An oscillating water column (OWC) wave power generation system, including wave energy converter, asynchronous generator, synchronous condenser and varied load, was built. Consequently, by using the frequency control and reactive power compensation techniques, an islanding power supply scheme was proposed. A comparative performance analysis in speed, torque, frequency and output voltage of the wave power generation system was done with single wave inputs and multiple wave inputs. The simulation results show that the proposed model can satisfy different running conditions of the wave power generation system. Moreover, the proposed power generation scheme can ensure the normal running of the system. It can provide good reference values for the actual construction of the wave energy system.

Keywords: asynchronous generator, wave power generation system, oscillating water column, frequency control, reactive power compensation.

1 Introduction

The continued global economic crisis has aggravated the deterioration of world energy crisis, especially accelerated the exhaustion process of the oil, natural gas, coal and other fossil energy. While the growth in the living standard and the accelerated development of modern science and technology make the demand for energy increased. In order to solve the contradiction between the supply and demand of energy, people are trying to find new alternative renewable energy, such as solar energy, wind energy, bio-energy, geothermal energy and ocean energy that including wave energy which are gradually developed and utilized in recent years [1, 2].

Wave energy is one of the energy storage ways that contain the most energy but are the least likely to use in ocean energy. Therefore, although Britain, Japan, Norway and other countries have been studying the wave power generation since 1970s, and have developed Gather Wave Reservoir, Oscillating Water Column, Duck, Point Absorber and many other multiform wave power generation devices, the wave power generation still has a long way to go before realizing its scale and commercialization [3-10].

* This work was supported by a grant from the National Natural Science Foundation of China (No. 51007060) and the Innovation Foundation of Tianjin University.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 335–342.
springerlink.com

China has vast sea areas, covering a total area of more than $4.7 \times 100 km^2$, having over 7000 islands, and the length of its coastline is more than 18000 km. According to the observation data of Chinese Marine Department, the wave density of our waters is about 2~5 kW/m, which means a lot of wave energy. If wave energy can be converted into electrical energy, it can supply for certain conditons that lack of electricity or can not be easily provided such as developed coastal areas, island residents and deserted island, then the unbalanced energy distribution problems in our country can be solved and the coastal defence of our country may be enhanced. However, at present, some technical problems in the exploitation of wave power are still difficult to overcome because of the complexity and inclemency of the marine conditions and the lack of research and investment. Therefore the commercial application of wave power technology and the large-scale grid generation still have a long way to go.

Because the rotation mechanism does not contact with the sea, the Oscillating Water Column generation device has a good anti-corrosion performance. It is safe, reliable and convenient to maintain, so that it can resist huge waves, typhoon and all kinds of bad weather. Therefore it is the most widely used conversion way of wave energy [7]. Fig. 1 shows the schematic diagram of the OWC wave power generation device.

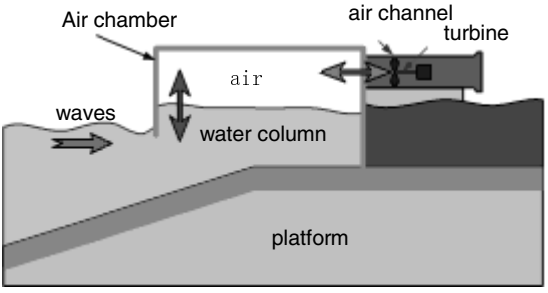


Fig. 1. Diagram of OWC-WEC

This paper focuses on the modeling of the OWC wave power generation device, and by using asynchronous generator, corresponding frequency control and reactive power compensation technique, the normal operation of wave power generation system is realized.

2 Mathematical Model of Wave Energy Converter

The wave energy converter is expressed by a mechanical oscillation system with damper as shown in Fig. 2 [7]. The basic principle of this system is that when external force acts on a mass block supported by the spring, the damper absorbs the motion energy of the mass block. The external force is provided by wave power, when the

generator is working, the damper will produce a reaction force on the mass block. In Fig. 2, motional mass block oscillates with driving frequency. Combining the mass block and spring system, the kinetic energy of mass block is converted into the potential energy of spring, and vice versa.

When the exciting force F is a sinusoidal signal, the motion equation of the motional mass block can be expressed as

$$m\ddot{x} + c\dot{x} + kx = F_0 \sin \omega t . \tag{1}$$

where m is the mass of motional object; x is the displacement of mass block; c is the damping coefficient; k is the spring coefficient; F_0 is the amplitude of F; ω is the angular frequency.

The solution of equation (1) can be expressed as the displacement movement of mass block as

$$x = A_0 \sin(\omega t - \varphi) . \tag{2}$$

Where

$$A_0 = \frac{F_0}{\sqrt{(k - m\omega^2)^2 + c^2\omega^2}} . \tag{3}$$

and

$$\varphi = \tan^{-1} \frac{c\omega}{k - m\omega^2} . \tag{4}$$

We can see that the mass block does sinusoidal movement with the amplitude A_0. Compared with the exciting force, its phase lags φ. Assuming the natural frequency of the mass block-spring system is ω_0, then

$$\omega_0 = \sqrt{\frac{k}{m}} . \tag{5}$$

When $\omega = \omega_0$, resonance happens. Then the amplitude of the mass block is maximizing with $\varphi = 90^\circ$. This is exactly the optimal working conditions for the external force to motivate the movement of mass block. At this time, all the energy of external force is converted into movement energy. Wave energy converters are also using this rule, trying to make wave and converter resonate in order to obtain higher energy conversion efficiency. The mathematical model of input wave force $F_0 \sin \omega t$ and output energy W_a in the typical OWC wave power generation device is shown in Fig. 3.

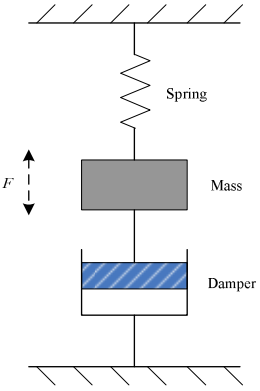


Fig. 2. Equivalent model of WEC

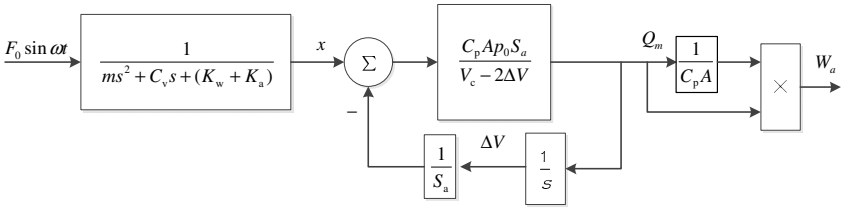


Fig. 3. Mathematic model of OWC-WEC

In Fig. 3, C_v is the damping coefficient of wave; K_W is the elasticity coefficient of water; K_a is the elasticity coefficient of air; p_0 is the air pressure when x=0; V_c is the chamber air volume when x=0. C_p is the flow coefficient; S_a is the superficial area of the water in chamber; ΔV is the volume variation of chamber; A is the equivalent orifice area; Q_m is the air mass flow.

3 Structure and Control of Wave Power Generation System

In order to set up an independent wave power generation system that based on the wave energy converter and asynchronous generator, the wave power control system is proposed as shown in Fig. 4. When the overdrive gear drags the asynchronous machine into running, the machine's speed exceeds its synchronous speed, then the machine turns into generator state, and the conversion from wave energy to electrical energy will be realized.

In the wave power generation system as shown in Fig. 4, the reactive power compensation unit meets the demand by using capacitors and synchronous condenser in order to ensure the stability of the output voltage.

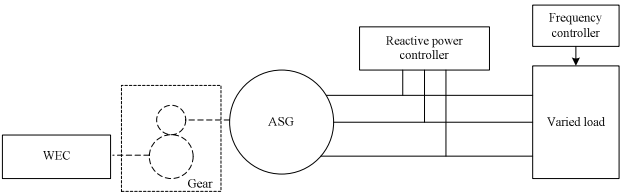


Fig. 4. Structure diagram of wave power generation system

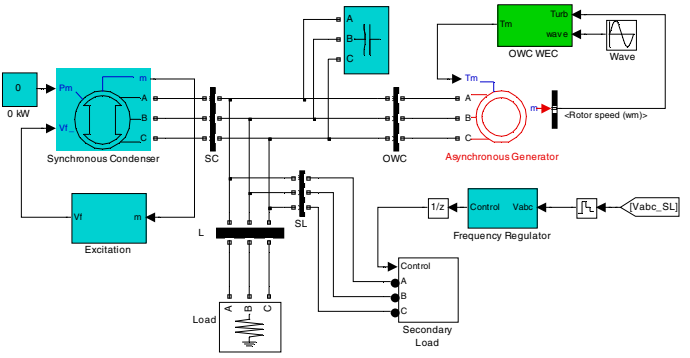


Fig. 5. Simulation diagram of wave power generation system

The paper uses the varied load frequency control strategy to ensure the stability of the output frequency. The frequency control unit uses PLL to measure the frequency of generating system. Comparing it with the reference frequency 50Hz ,we get the frequency error, whose integral result is the corresponding phase difference, finally we can get the demanding value of switching load through the PD controller. In addition, when the load is switching, in order to reduce the fluctuations of voltage, the switching happens at the zero crossing point of voltage.

Through the above reactive power compensation and frequency control strategy, we can ultimately ensure that the unstable wave energy is converted into stable electrical energy. It's worth noting that this islanding wave power generation system has certain requirements for the input values of wave energy, so it is suitable for the stable sea areas, and when the input wave forces are too small or too large, the system will be automatically out of service.

4 The Simulation Results

Fig. 5 shows the corresponding simulation system chart built in MATLAB. As for the wave input in Fig. 5, the paper uses the ideal sine wave function input to start the simulation, the wave thrust F_0=8×10^{5}N, the height of incident wave H=2m, the period of incident wave T=6s. The rated power, voltage and frequency of asynchronous generator are respectively set as 20kW, 380V and 50Hz. The corresponding output

results are shown in Fig. 6. As shown in Fig. 6, at this time the input torque ripple is big, so the corresponding motor speed oscillation and frequency fluctuation are also big, while both of them and the output voltage of the generator remain stable. This basically meets the demand of the load.

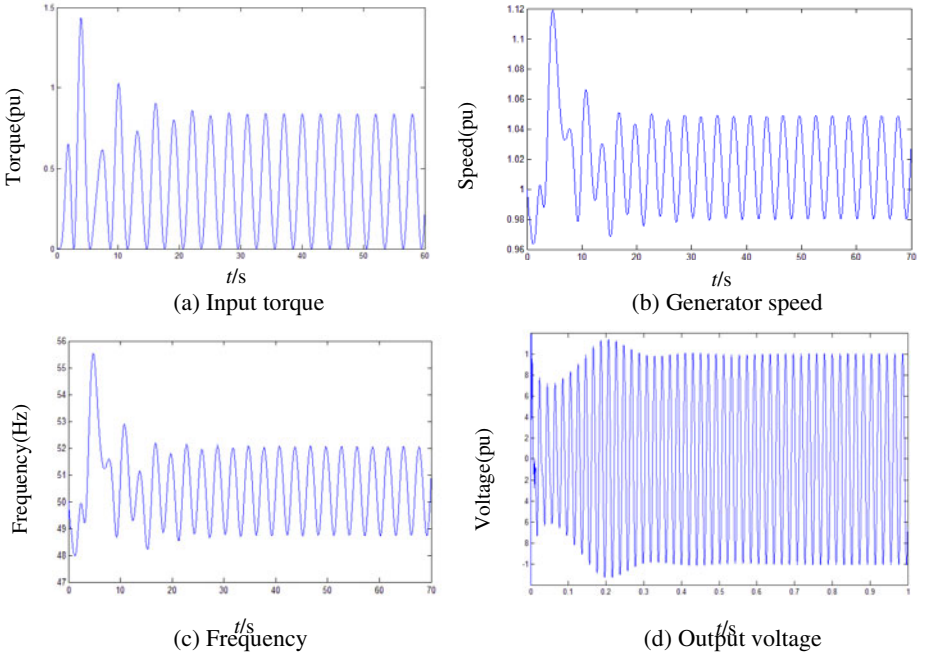


Fig. 6. Simulation results of single wave input system

The periodic zero points (shown in Fig. 6 (a)) existed in the single wave input system cause the insufficient power input, which affects the reliability and stability of the wave power generation system. This problem can be solved by increasing the volume and area of the wave energy converter (multiple openings and air chambers mode can be used in OWC-WEC). At this time the wave that actually acts on the converter can be approximately regarded as the superposition of multiple wave inputs with different phases, and then compositing the corresponding outputs to simulate the complex output energy of the OWC. Fig. 7 shows the simulation model of multiple wave input system, this model uses four sine wave inputs with the same amplitude 8×10^5N, same period 6s and successively phase difference 0.8rad. The corresponding simulation results are shown in Fig. 8.

As shown in Fig. 8 (a), in this condition, the input torque ripple becomes smaller. When it is stable, the fluctuation range is 0.41~0.43. This fundamentally ensures the reliability and steady operation of the system. While Fig. 8 (b), (c) indicates that the stable effect of the corresponding motor speed and frequency is better. In steady state, the speed is 1.00~1.02, the slip ratio is small and the frequency maintains 49.95~50.05Hz, which is in the allowable range. Meanwhile Fig. 8(d) shows that the output voltage quality of the generator is also improved.

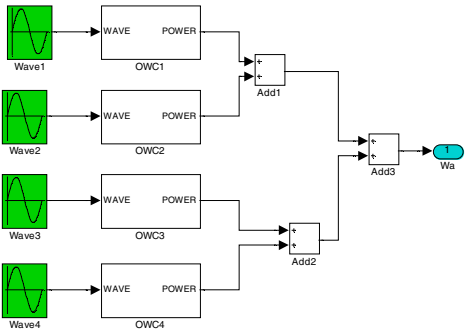
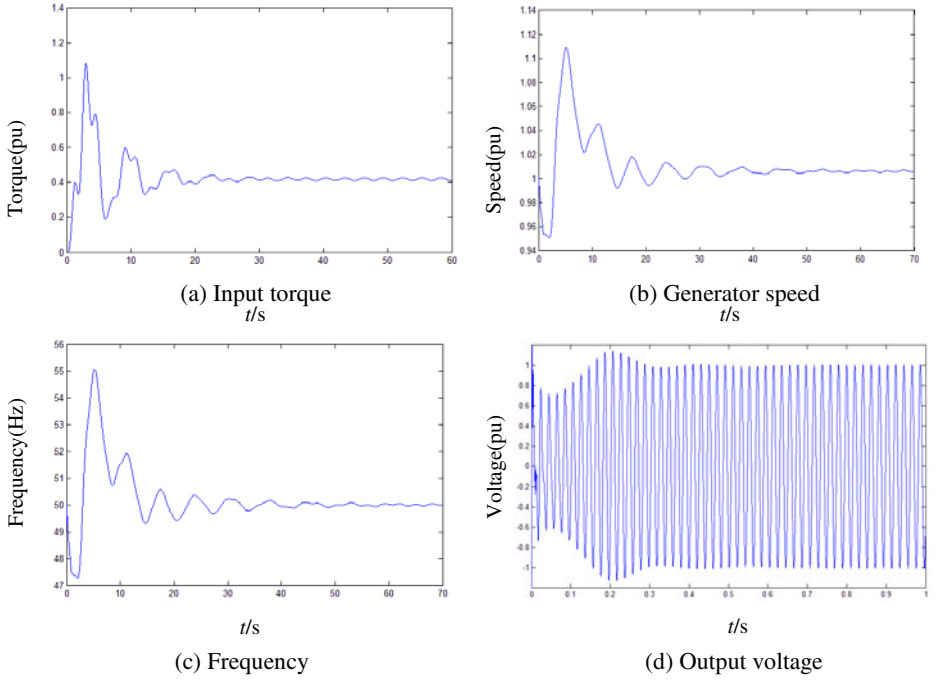


Fig. 7. Simulation model of multiple wave input system



(a) Input torque (b) Generator speed

(c) Frequency (d) Output voltage

Fig. 8. Simulation results of multiple wave input system.

5 Conclusion

This paper has proposed an independent wave power generation system based on asynchronous generator and OWC. By using the varied load frequency control method and the reactive power compensation strategy of combining the nearby capacitors and synchronous condenser, it realized the well control of system voltage and frequency, and ensured the reliability and stability of the system. Meanwhile the

corresponding simulation block diagrams and results are also provided. The analysis shows that the proposed topological structure of power generation and system control method is suitable for the stable sea areas. It can convert the unstable wave energy into stable electrical energy, so it will have certain potential application values on the development of the deserted islands and the independent power supply for the frontier defense islands.

At the same time, the simulation results of multiple wave input system show that one of the solutions for the stability problem of wave power generation system is trying to enlarge the wave energy converter or use multiple distributed wave energy converters (it is better to use optimal location distribution) to generate jointly, because this can fundamentally solve the instability and low efficiency problems of the input. In addition, since the structure proposed in this paper only considered the stable sea areas, there is no AC-DC-AC power conversion unit at the generator's output side.

References

1. Twidell, J., Weir, A.: Renewable Energy Resources. Taylor & Francis, London (2006)
2. Boyle, G.: Renewable Energy, Power for a Sustainable Future. Oxford University Press, Oxford (2004)
3. Thorpe, T.W.: A Brief Review of Wave Energy. ETSU Report Number R-72 (1999)
4. Falnes, J.: Ocean Waves and Oscillating Systems, Linear Interaction Including Wave-Energy Extraction. Cambridge University Press, Cambridge (2002)
5. Cheng, Y.L., Dang, Y., Wu, Y.J.: Status and Trends of the Power Generation from Wave. Applied Energy Technology 12, 26–30 (2009) (in Chinese)
6. Cui, L., Xiong, Y., Wang, H.F.: Study on Analysis Method of Conversion Efficiency of Wave Power Generation System Based on Onsite Testing. Ocean Technology 28(3), 111–113 (2009) (in Chinese)
7. Yan, Y.B., Tomiji, W.: Principle and Device of the Ocean Wave Energy Conversion. Shanghai Science & Technology Press, Shanghai (2011) (in Chinese)
8. Wu, F., Zhang, X.P., Ju, P., Sterling, M.J.H.: Modeling and Control of AWS-Based Wave Energy Conversion System Integrated into Power Grid. IEEE Transactions on Power Systems 23(3), 1196–1204 (2008)
9. Zhou, Z., Knapp, W., MacEnri, J.: Permanent Magnet Generator Control and Electrical System Configuration for Wave Dragon MW Wave Energy Take-off System. In: IEEE International Symposium on Industrial Electronics, Cambridge, pp. 1580–1585 (2008)
10. O' Sullivan, D.L., Lewis, A.W.: Generator Selection for Offshore Oscillating Water Column Wave Energy Converters. In: The 13th Power Electronics and Motion Control Conference, Poznan, pp. 1790–1797 (2008)

Phase-Based RMFDA Fault Diagnosis Method Using Bootstrap Technique

Shu Wang[1,2,*], Zhen Zhao[3], Yu-qing Chang[1,2], and Fu-li Wang[1,2]

[1] Information Science and Engineering School, Northeastern University, Shenyang, China
[2] Key Laboratory of Integrated Automation of Process Industry, Northeastern University, Shenyang, China
[3] School of Information & Control Engineering, Liaoning University of Petroleum & Chemical Technology, Fushun.
alicews5@163.com

Abstract. The fault diagnosis performance of Fisher Discriminant Analysis (FDA) method is superior to Principle Component Analysis (PCA) by taking into account both normal and fault data for modeling. For the cases with insufficient fault data, a diagnosis strategy is developed based on Bootstrap and phase-based Recursive Multi-way Fisher Discriminant Analysis (RMFDA). By this method, modeling data was constructed by Bootstrap. Besides, the diagnosis information of the previous phase was introduced in the next phase for MFDA modeling by combining recursive method. The effectiveness of the proposed method is demonstrated by applying it to the hydraulic tube tester.

Keywords: Bootstrap, sub phase, RMFDA, fault diagnosis.

1 Introduction

To guarantee the safety and reliability of batch process production, the online monitoring and fault diagnosis has become a key challenging problem in these processes. Earlier fault diagnosis methods are almost based on precise system models. However, it is difficult to build precise mathematical model with the increasing system complexity and characteristics of high dimension, nonlinearity, strongly coupling, random noise and time delay. With the development of industrial automation, more and more historical data information has been retained in modern industrial production. Because of that, fault diagnosis based on multivariate statistical method has attracted more and more attentions by scientific researchers. Considering that FDA utilizes historical data not only under normal condition but also under fault conditions, further provides fault reason according to priori knowledge, it is obvious that FDA yields better performance than other multivariate statistical technologies [1].

Since most batch processes have multi-phases behavior, it is inevitable to weaken this phase characteristic by building a single MFDA model using all the data in a whole batch as a statistical sample, and decreases precision of diagnosis. For this

* Corresponding author.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 343–350.
springerlink.com

reason, it is necessary to build phase-based sub-models for batch process monitoring and diagnosis. Meanwhile, it is impossible to obtain enough and complete fault samples for that some measure is implemented as soon as a fault.

To solve these abovementioned problems, this paper proposes a phase-based Recursive Multi-way Fisher Discriminant Analysis (RMFDA) fault diagnosis method using bootstrap technique to diagnose faults of hydraulic elements in a hydraulic tube tester. In the first step, the modeling data is augmented using bootstrap technique. Based on the augmented data, the whole process is divided into several sub-phases. Fault diagnosis is carried out by building MFDA model for each sub-phase. While building the models for the second sub-phase to the last sub-phase, the main fault information extracted in the fore sub-phase is also used in the current modeling process.

The remainder of this paper is organized as follows. First, Bootstrap technique is applied to construct the modeling data in Section 2. Section 3 describes phase division of a batch process. Subsequently, the proposed fault diagnosis method is formulated in detail. In Section 5, a hydrostatic testing process simulation shows the effectiveness of the proposed method. Finally, some conclusions are drawn in Section 6.

2 Construct Modeling Data Using Bootstrap

Based on sufficient modeling data, including normal data and fault data, FDA performs well in fault diagnosis. However, fault data is always scarce in practical production processes, which may greatly decrease the diagnosis precision using FDA method. In this paper, modeling data for building the MFDA models are augmented using Bootstrap technique.

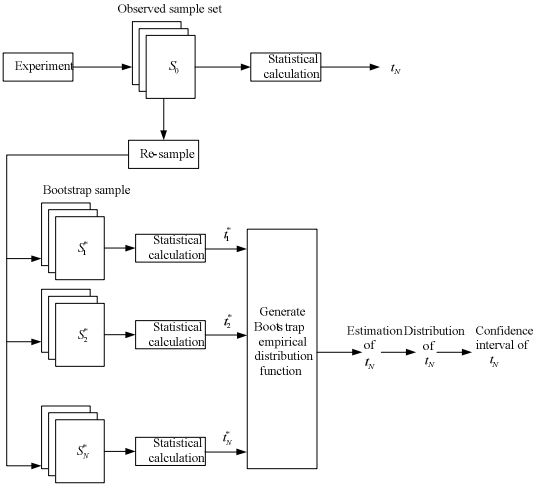


Fig. 1. Flow diagram of Bootstrap

Bootstrap, which is a statistical method using re-sampling technique to estimate population parameters, was proposed based on Jackknife method by B. Efron in 1979 [2,3]. Consider the cumulative distribution function of a random variable X, with observed samples $S_0 = \{x_1, x_2, \cdots x_n\}$, which is indicated by the symbol F. F is an unknown underlying distribution. In fact, Bootstrap technique is to repeatedly re-sample data from the original sample set, i.e. generate B samples, denoted by x_i^* (i is the order of the re-sampling), each of size m drawn with replacement from the n observations $S_0 = \{x_1, x_2, \cdots x_n\}$. Fig. 1 shows the flow diagram of the Bootstrap technique. In Fig. 1, $S_x^* = \{x_1^*, x_2^*, \cdots x_n^*\}$, also called a Bootstrap dataset, is a simple sample generated from S_0.

In this paper, the construction of modeling data using Bootstrap technique is divided into two cases: sufficient data and limited data. Considering limited samples, Bootstrap datasets are generated using replacement re-sampling to make up the insufficiency of data. When the reference data is sufficient, the bootstrapping is relaxed, in which re-sampling is implemented without replacement so that the modeling samples in each bootstrap modeling set are not repeated. Generally, replacement re-sampling is used here for that it is difficult to obtain enough fault samples. Multiple models obtained from different calibration samples generated by the bootstrap technique may be expected to evaluate the candidate descriptors more robust, reasonable and precise than a single model. The detailed steps are as follows:

a) The data is preprocessed as the same as MFDA. The three-way matrix $\bar{\mathbf{X}}$ is unfolded time-wise into a two-way matrix $\mathbf{X}(I \times KJ)$, each row of which is the observation data in a whole batch. It is noted that the number of batches for each type of data is the same to each other.

b) Bootstrap technique is adopted on the unfolded data $\{\boldsymbol{X}(I \times KJ)\}$ to uniformly re-sample each type of data to obtain N_b Bootstrap datasets $\{\boldsymbol{X}\}_i$ ($i = 1, 2, ..., N_b$), in which they show the same number of batches I between Bootstrap datasets and the same number of batches for each type of data in a single Bootstrap dataset.

c) For each Bootstrap dataset, every type of data is partitioned into training part and testing part, and consequently forms I_{tr} batches of training data and I_{te} batches of testing data ($I = I_{\text{tr}} + I_{\text{te}}$). Hence, there are N_b training datasets $\{\mathbf{X}(I_{tr} \times KJ)\}_i$ ($i = 1, 2, ..., N_b$) and N_b testing datasets $\{\mathbf{X}(I_{te} \times KJ)\}_i$ ($i = 1, 2, ..., N_b$), which contain different data samples and represent different input spaces.

From the above procedure, the use of Bootstrap technique produces more Bootstrap datasets and thus provides more reference modeling data for the MFDA modeling, which has advantages of reducing dependence on a single modeling dataset and thus enhancing the diagnosis performance of the MFDA.

3 Phase Division

Multi-phase is an inherent feature of many batch processes, for instance the batch fermentation process can be divided into lag, log (exponential), stationary, and death/decay in line with the bacteria growth cycle; the injection molding production process can be split into 3 stages of injection, pressure maintaining and cooling. Kosanovich [4] et al employed the MPCA method into the industrial process of polymer reaction in 1994, raising the point that the separate building of MPCA models in accordance with the reaction phases with different characteristics can more accurately and effectively monitor the running state during the production process. Therefore, the multi-phase batch process operation and troubleshooting of the statistical modeling process not only to analyze the overall health are normal, but should be in-depth analysis of each sub-phase operation is normal. In other words, the batch process is divided into several sub-phases to establish different phase models will be able to better reveal the running state of the process. Besides, the sub-phases modeling methods will facilitate the subsequent fault diagnosis and fault prediction research.

Processes can be divided into sub-phases by clustering or indicator variable directly [5,6].

Let N_b Bootstrap training data sets from P types data be $\{X(I_{\text{tr}} \times KJ)\}_i$ ($i = 1,2,...,N_b$), it can be denoted by $\boldsymbol{X} = [\boldsymbol{X}_1 \quad \boldsymbol{X}_2 \quad \cdots \quad \boldsymbol{X}_C]_i$ after division, where $\boldsymbol{X}_1 \in R^{I_{tr} \times k_1 J}$, $\boldsymbol{X}_2 \in R^{I_{tr} \times (k_2 - k_1) J}$, $\boldsymbol{X}_C \in R^{I_{tr} \times (k_C - k_{C-1}) J}$ are the measurement data matrix in the i^{th} datasets respectively.

A MFDA model is built in each phase when establish the fault diagnosis models. Modeling data in each model is denoted as $\{\boldsymbol{M}_c\}_i$ ($c = 1,2,\cdots,C$) ($i = 1,2,...,N_b$), such as demonstrated in Fig. 2.

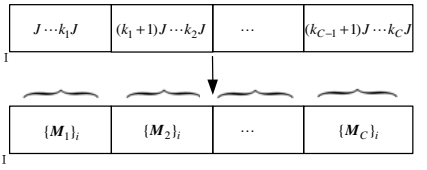


Fig. 2. Phase division

4 Bootstrap and Phase-Based RMFDA

As the batch process has a strong dynamic performance, single phase fault information may not be fully described. The previous phase information is introduced to next phase modeling can take full advantage of the limited data resources, to further improve the performance of fault diagnosis. Therefore, a bootstrap and phase-based recursive multi-way Fisher discriminant analysis method is proposed in this paper for fault diagnosis. The method is described as follows:

Trained N_b Bootstrap training data sets as fault diagnosis models as following:

The first Bootstrap data sets $\{\boldsymbol{X}(I_{\text{tr}} \times KJ)\}_1$ is described as $\boldsymbol{X} = [\boldsymbol{X}_1 \quad \boldsymbol{X}_2 \quad \cdots \quad \boldsymbol{X}_C]_1$. Built C sub MFDA models with $\{\boldsymbol{M}_c\}_1$ $(c = 1, 2, \cdots, C)$.

The first sub MFDA model is built with the first phase data matrix $\boldsymbol{X}_1$, i.e. $\boldsymbol{M}_1 = \boldsymbol{X}_1$. The first model's discriminate matrix $\boldsymbol{W}_1$ can be gotten by FDA. Project $\boldsymbol{X}_1$ to the first MFDA model to get the first model's discriminant score matrix $\boldsymbol{T}_1$:

$$\boldsymbol{T}_1 = \boldsymbol{X}_1 \boldsymbol{W}_1 \tag{1}$$

The average discriminant score vector of each sort in the 1st MFDA model is $\overline{\boldsymbol{t}}_{1,p}\,(p = 0, 1, \cdots P-1)$.

All the other sub models besides the first one are built by using not only the data of present phase but also the fault information extracted form former phase. That is, for the c^{th} phase of MFDA model, its off-line modeling data is a augmented matrix composed of $\boldsymbol{T}_{c-1}$ and $\boldsymbol{X}_c$, where $\boldsymbol{T}_{c-1}$ is the discriminant score matrix of phase $c-1$.

$$\boldsymbol{M}_c = [\boldsymbol{X}_c \quad \boldsymbol{T}_{c-1}] \tag{2}$$

Consequently, discriminant matrix $\boldsymbol{W}_c$ of the c^{th} model and discriminant score matrix $\boldsymbol{T}_c$ is obtained as

$$\boldsymbol{T}_c = \boldsymbol{M}_c \boldsymbol{W}_c \tag{3}$$

Average discriminant score vector of each sort is derived as

$$\overline{\mathbf{t}}_{c,p} = \left[\overline{\mathbf{x}}_{c,p} \quad \overline{\mathbf{t}}_{c-1,p}\right] \mathbf{W}_c, \quad p = 0, 1, \cdots P-1 \tag{4}$$

where $\overline{\boldsymbol{x}}_{c,p}$ is average vector of p sort in c^{th} model. Repeat the above steps to establish each model in order.

Repeat the above steps to N_b data sets. Built c sub models for every dataset and calculate all sorts' discriminant score vector to every sub models.

Average the N_b group RMFDA sub models to obtain a comprehensive model as the final sub-phase RMFDA model. Where the number of latent variables (determine the matrix dimension) can be determined using Akaike information method.

Once a fault diagnosis model has been built, it can be used for online fault diagnosis. If $\boldsymbol{x}_{new,c}$ is measured variables of production process at c^{th} phase, project it to corresponding c^{th} KFDA. Thus, the discriminant score vector $\boldsymbol{t}_{new,c}$ can be obtained as follows:

$$\begin{aligned} \boldsymbol{t}_{new,1} &= \boldsymbol{x}_{new,1} \boldsymbol{W}_1 \\ \boldsymbol{t}_{new,c} &= \left[\boldsymbol{x}_{new,c} \quad \boldsymbol{t}_{new,c-1}\right] \boldsymbol{W}_c \end{aligned} \tag{5}$$

Based on the Euclidean distance, it can be diagnosed which type of fault has been happened. The Euclidean distance between $\boldsymbol{t}_{new,c}$ and each sort of average discriminant score vector can be expressed in the form

$$d_{new,p}(k) = \left\| \boldsymbol{t}_{new,c} - \bar{\boldsymbol{t}}_{c,p} \right\|, \quad p = 0,1,\cdots P-1 \tag{6}$$

The least one of $d_{new,p}(k)$ is the fault diagnosis result which the current process belongs to.

5 Application Studies and Discussion

5.1 Process Introduction

The proposed method is validated by hydrostatic testing process which is a typical batch process. Hydraulic tube tester is a machine designed and manufactured to do hydraulic test for steel tube, which is the last and key step of quality detection for steel tube. During the hydraulic testing process, the pressure load ability and shortcoming of tube is showed. Due to the high pressure in the hydraulic testing process, it may involve grave danger to human life or property when possible faults happen. Thus, fault diagnosis on hydraulic tube tester will effectively ensure process safety, greatly improve the productivity of steel tubes, and promote the development of steel tube industry. The illustration of hydraulic tester is shown in Fig. 3.

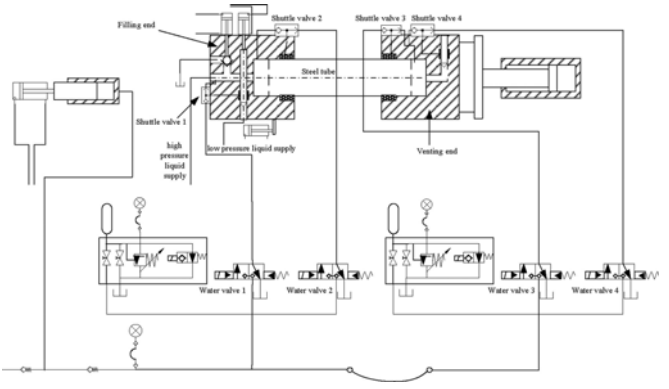


Fig. 3. Illustration of hydraulic tester components

5.2 Experiment Verification

In order to validate the above method, it is applied to a complex hydraulic testing machine fault diagnosis. In this paper, shuttle valve failure, for example, the pressurized side of the two shuttle valves set at No. 1 and No. 2, the support side of the two shuttle valves set at 3 and 4, shuttle valve, select the pressurized cylinder pressure, Support cylinder pressure, pre-seal pressure, support the four variables of these four-cylinder displacement of the fault diagnosis of valve locks. 10 normal

batches and 40 other four kinds of fault batches are collected respectively as the training datasets. Bootstrap four Bootstrap datasets all contains 50 batches data. In this paper, the approach requires the same production cycle, that is, equal to the number of sampling points. Because we were limited to the failure types of shuttle valve failure, the number of sampling points does not appear too large difference. Namely, the average length of production cycle as the benchmark, if the production cycle is shorter than the base period, while a truncation of the batch, if the production cycle is shorter than the benchmark short period, then fill the batch, so that it has the same sampling points. In this example, a sample taken every 0.2s, each batch of 64 sampling points, time division of the situation shown in Fig. 4.

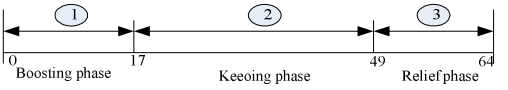


Fig. 4. Three stages of process data

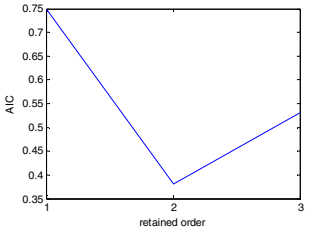


Fig. 5. AICof 1st sub-phase

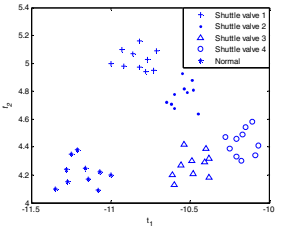


Fig. 6. Discriminant score of 1st sub-phase

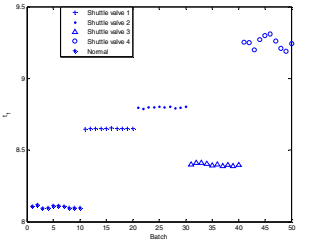


Fig. 7. Discriminant score of second sub-phase

According to the proposed method, three sub RMFDA models can be built based on each bootstrap datasets. The optimal orders of the first and third models are reduced to 2. The optimal order of the second model is reduced to 1. Fig. 5 shows the Akaike information criterion (AIC) of first sub-phase. Fig. 6 shows the discriminant score obtained from the first sub- phase data in the first MFDA models. Fig. 7 shows the discriminant score obtained from the second sub- phase data in the second MFDA models. To test the performance of the method, the rate of misclassification is used as

the evaluation of fault diagnosis performance indicators, where the rate of misclassification = (number of misclassification) / (sum of samples). The diagnosis performance of the proposed method and MFDA method is showed in Table 1. The misclassification rate of MFDA is 32% and the misclassification rate of the proposed Bootstrap based RMFDA method is 14%.

Table 1. Compared results of the two methods

	Average misclassification rate
MFDA	0.32
The proposed method	0.14

6 Conclusions

A Bootstrap based RMFDA method is presented to resolve the fault diagnostic problem of batch process. The performance of the proposed diagnosis method is tested by using some kinds of faults data which are obtained from the production process of Hydraulic Tube Tester. The diagnostic results show that the method proposed in this paper is effective and has high performance in fault diagnosis.

Acknowledgement. This work was supported in part by the National 973 project (2009CB320603), the National Natural Science Foundation of China (60974056).

References

1. Jiang, L.: Researches on Fault Diagnosis for Process Industry with FDA/DPLS Methods. Doctor dissertation. Zhejiang University, Hangzhou (2005)
2. Bradley, E.: Bootstrap methods: another look at the jackknife. The Annals of Statistics 7(1), 1–26 (1979)
3. Quenodille, M.H.: Approximate tests of correlation in time - series. Journal of Royal Statistical Society, B (11), 68–72 (1949)
4. Kosanovich, K.A., Piovoso, M.J., Dahl, K.S.: Multi-Way PCA Applied to an Industrial Batch Process. In: Proc. of American Control Conf., p. 1294 (1994)
5. Zhao, C., Wang, F., Lu, N., Jia, M.: Stage-based soft-transition multiple PCA modeling and on-line monitoring strategy for batch processes. Journal of Process Control 17(9), 728–741 (2007)
6. Lu, N.Y., Gao, F.R., Wang, F.L.: A sub-PCA modeling and on-line monitoring strategy for batch processes. AIChE J. 50(1), 255–259 (2004)

Design of Vision System Based on Varying Lighting Condition for Multi-robots

Wuxin Huang[1,2], Shili Tan[2], and Xiang He[2]

[1] Faculty of Mechanical and Electronic Engineering,
Jiangxi University of Science and Technology, Ganzhou 341000, China
[2] College of Mechatronics Engineering and Automation, Shanghai University,
Shanghai 200444, China
hwx0819@sohu.com, tanshili@shu.edu.cn

Abstract. A robot collects environment information mainly by its vision system. In order to improve recognition accuracy and positioning precision under variable lighting conditions, this paper describes a new pattern design and proposes a method to adjust the color threshold according to the credibility of current color threshold. Image morphology is used to optimize the recognition results. The vision system is tested in the project Family Life Support Multi-Robot Systems.

Keywords: machine vision, threshold value, confidence level, vision system.

1 Introduction

Vision system of robot mainly includes image processing, pattern recognition and visual track, etc. Its primary tasks are to obtain real-time image, and to analyze and determine the coordinates and configuration of robot to provide the data for control module. In this paper, the vision system of multiply service robot are studied and designed. Because the service robot has some characteristics such as broad action range, unstable environment illumination, there are the following difficulties during constructing the vision system: ① unstable environment illumination condition; ②large processing data; ③many objects to be recognized; ④distortion image.

In view of the above question, the researchers from the domestic and foreign have been proposed lots of solutions. In 2003, the team CMUDragon from CMU has proposed a butterfly-shaped color code strategy in the robot soccer competition, and it has obvious advantages of positing accuracy and quantity of target discrimination [1]. Wang shuan from Tsinghua University proposed a multi-objective track method based on difference image which compute the difference between two images before and after such that it can reduce the data to handle [2]. In order to solve the uncertainty of image division color threshold of the vision system, the color search tables, color division based on margin check and HSI threshold value etc. have been presented successively by many researchers. These methods have been widely applied in the soccer robot vision system, and have made more improvement [3]. In this paper, considering wide

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 351–356.
springerlink.com

action range and unstable environment illumination of multi-service robot, a novel color code design is proposed to adjust automatically the color threshold value. Finally, the project on multi-service robot is carried on so that the correction and reliability of the proposed method can be validated.

2 Robot Vision Subsystem

The vision system is a main sensor which can sense external environment in the robot system. It consist of the camera, the image acquisition card, image processing program, the computer and the network service equipment. The vision system is shown in Figure 1.

Fig. 1. Ceiling cameras and trinocular vision

3 Set the Color Threshold Value

Choosing color image division method to discriminate target, we should choose a proper color space, such as RGB, YUV, HSI (HSV),CMY, which may influence directly image division effect. HSI color space is close to human eye sensation color space, H is tone, I is brightness, S is a degree of saturation. Tone attribute H can reflect the color type quite accurately, and its sensitive of outside illumination condition is low. Color characteristic H has quite stable and narrower variable range for the same object, and it can be chosen as primary distinguishing parameter. But when the RGB value is small, namely, brightness I is small, the H value is uncertain, therefore brightness I and degree of saturation S is also very important [4].

Before dividing the image, the threshold value scope should be determined first, for example, red H value is probably [0, 0. 5]. Due to the effect of unstable environment light source, the same red code cannot be distinguished completely in the same environment, thus it can influence the recognition result. Considering the environment light source is unstable and distribution of light is uneven, the four color blocks method and automatic threshold adjust based on its confidence level are proposed, respectively. In view of the unstable environment illumination, the Otsu threshold value method by using four color blocks(red, yellow, green, blue) and K-L transform method are applied to adjust automatically the threshold value of color space.

The initial color space threshold values of the vision system can be obtained by calculating HSI values of four colors in its color blocks. When environment photo source change, we only need calculate the HSI value of four color blocks (red, yellow, green, blue), such that the color threshold values can be adjusted automatically during dividing image. Its computational process is as follows:

(1) The RGB values of four colors in color blocks are transformed to HSI.

(2) Uses the Otsu threshold value method to choose appropriate threshold value. Divides the four color blocks into four connected domains.

(3) Computer the Otsu threshold values of four domains, respectively.

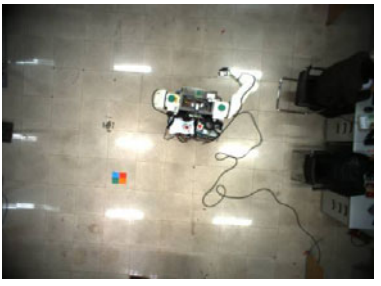
(4) Make the histogram of I and C, and calculate the threshold value of I and C by using Otsu multi-threshold method.

(5) Divide locally the four connected domains according the threshold values obtained from above step.

(6) The interference regions influenced by light resource can be eliminated by handling of step 5, and the connected regions become more impact. Then, the H, S, I values can be acquired by using the theory of probability, and the best HSI threshold value of red, yellow, green and blue can be obtained. Table 1 is HSI threshold value statistics in two kind of illumination situation. Figure 2 is two kind of different illumination conditions.

Table 1. Threshold determined automatically

Conditions	Color	H	S	I
1	red	[0, 0. 2]	>0. 50	>60
	yellow	[0. 3, 0. 6]	>0. 60	>75
	blue	[3. 2, 4. 8]	>0. 60	>80
	green	[1. 1, 3. 2]	>0. 50	>50
2	red	[0, 0. 4]	>0. 40	>45
	yellow	[0. 5, 1. 0]	>0. 35	>45
	blue	[2. 8, 4. 6]	>0. 45	>60
	green	[0. 9, 3. 1]	>0. 45	>55



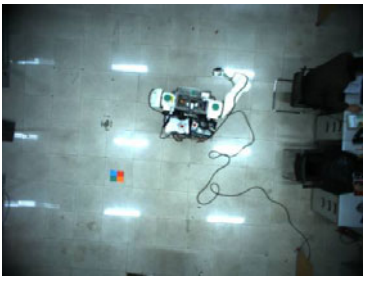
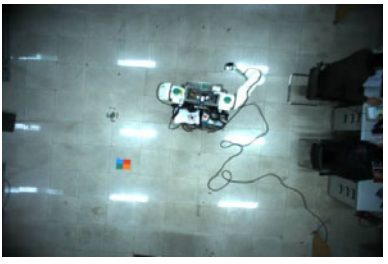


Fig. 2. Comparison of two lighting conditions

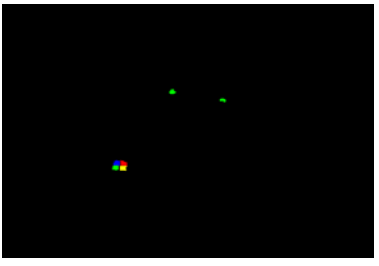
In the case of uneven distribution of optical fiber, the threshold value confidence level method is applied to recognize the results [5]. All information of target can be obtained during the process, is called as complete information, and confidence level is relatively higher. Otherwise, if the information on target is lost or cannot be obtained, it is called as the fragment information, and the confidence level is lower. Whether the threshold value is available and the degree of its change is determined confidence level. The confidence level α of the current threshold value is determined according to the integrity of recognition results. If the current recognition result is the complete information, Alpha is 100; If the current recognition result is the fragment information, it is $\alpha=\alpha\times\lambda$, λ= pixel number of the target discrimination / pixel number of target to be identified [6].

The confidence level has the bound namely its scope, the threshold value confidence level scope is [80, 102]. When threshold value confidence level is not in scope of the system, the system will choose automatically a new threshold value according to initial value of confidence level and corresponding table of HSI threshold value. Using the confidence level, the fragment information is used as far as possible, and real-time adjusts the color HSI threshold value, so that the recognition result's continuity and the accuracy can be guaranteed, and improve efficiency and the precision of the system recognition. Figure 3 is recognition goal result through the automatic threshold value adjustment.

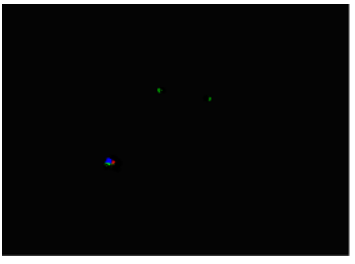
The massive experimental results have shown that the color block and confidence level of threshold value method can greatly improve the efficiency and the precision under the case of unstable and uneven light source as shown in table 2.



(a)Condition 2



(b)With threshold of the credibility



(c)No threshold of the credibility

Fig. 3. Results of identifying the target with threshold of the credibility of the system

Table 2. Recognition rate by new measures %

	Unstable light source	Uneven light source	Recognition rate
no measure	85.2	92.0	90.6
measure	91.3	96.5	98.8

4 The Results and Analysis of Recognition

The coordinate of object is obtained in the image, and the image coordinate can be transformed the world coordinate by using their relation, and the three dimensional information of service robot can be acquired. It can be shown as the following diagram:

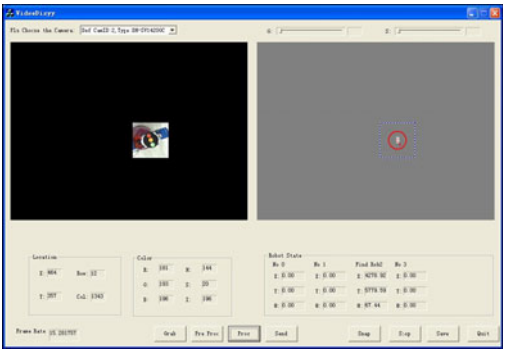


Fig. 4. Vision system and recognition result

The original point of the image coordinate system is left bottom, X axis direction is toward right, Y axis direction is upward. Let the vector direction between the center point of two circular marking and another circular marking point is parallel to axis X, the direction of robot is 0 degree. The erroneous coordinate between (X:250.40cm, Y:252.50, angle is 178.0 degrees) and the actual environment measurement result (X:250, Y:250, angle 180 degrees) is within $\pm 5cm$, the angle is within ± 5 according to recognition results of top vision system.

5 Object Recognition of Short Distance

If service robot need catch the object, the current position information need be obtained by top vision system, and move the target point according to the position of the target so that the object can be captured reaching the object point. The robot still need send out the request to trinocular vision system to obtain the precise target position, that is, it is named the recognition of short distance.

The fingernail of the service robot is made by the characteristic of jar, the data on the coordinate of the rotation center need be gained. The data of outflow boundary of the bottle cap can be obtained using the feature of bottle cap, and the three dimensional data of rotation center can be acquired by the radius of bottle cap. In order to gain a more precise data, all similar characteristic points are obtained and used to assess overall

after the recognition of object characteristic region. As for the bottle cap, the outflow boundary close to camera is variable along with the longitudinal height(Y value), and the distance between outflow boundary and the camera is constant, that is, the value Z is invariable. But this outflow boundary is the position between the center line of two bottle cap under the camera photography. The data by combining right camera and left one are as follows:

Table 3. The three dimensional data of center line of bottle cap using right and left camera (mm)

X	11.4	11.4	11.4	11.3	11.3	11.4	11.3	11.3	11.3	11.3
Y	-126.5	-126.6	-125.3	-125.8	-121.9	-124.4	-119.3	-118.7	-117.2	-116.0
Z	635.7	636.0	634.2	631.7	632.0	632.4	633.2	628.6	630.0	630.2

The data show the X values is invariable basically, and the depth of field Z values have the greatest change, it also most unstable. Y values represent the vertical motion direction of fingernail, and it has small influence on the capture of object. The Z values are key to catch object. The average value from 10 same Z value obtained from above results is 632.4mm, the actual measurement value is 636.5mm, and the error range is $\pm 1cm$.

6 Conclusion

In this paper, considering wide action range and unstable environment illumination of multi-service robot, a novel color code design is proposed to adjust automatically the color threshold value. The massive experimental results have shown that the color block and confidence level of threshold value method can greatly improve the efficiency and the precision under the case of unstable and uneven light source.

Acknowledgements. This project was supported by National High-Tech Research and Development Program of China (863 Program) under Grant No. 2007AA041604, and the Science and Technology Project of Education Department of Jiangxi Province under Contract No.GJJ10491.

References

1. Bruce, J., Veloso, M.: Fast and accurate vision based pattern detection and identification. In: Proceedings of the 2003 IEEE International Conference on Robotics and Automation, pp. 1277–1282 (2003)
2. Kimb, G., Shimj, I., Parkd, J.: Fast image segmentation based on multi-resolution analysis and warelets. Pattern Recongnition Letters 24(16), 2995–3006 (2003)
3. Thrun, S., et al.: Probabilistic Robotics. MIT Press (2005)
4. Triggs, B., Laugier, C.: Automatic camera placement for robot vision tasks. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 1732–1737 (1995)
5. Lee, J.H., Akiyama, T., Hashimoto, H.: Study on optimal camera arrangement for positioning people in intelligent space. In: IEEE/RSJ Intl. Conference on Intelligent Robots and Systems, pp. 220–225 (2002)
6. Ohara, K., et al.: Ubiquitous Robotics with Ubiquitous Functions Activate Module. In: Proc. INSS 2005, pp. 97–102 (2005)

A Software Reliability Evaluation Method Based on Rough Set

Qiuying Li and Jian Wang

School of Reliability and Systems Engineering, Beihang University
100191 Beijing, China
li_qiuying@buaa.edu.cn

Abstract. Quantitative software reliability evaluation method depends on the failure data of software. And most evaluation work is started only after obtaining the failure data in the later period of software development, which leads to a problem that people could not obtain the information of software reliability level in the early stage of software development, which could be used for guiding the work of software reliability design, analysis and testing. Therefore, the qualitative software reliability evaluation methods based on the reliability corresponding factors in each stage of the software development are studied recently. In this paper, a qualitative software reliability evaluation method based on rough set is proposed, which takes use of the data mining technology based on rough set to analyze the data related to the software reliability, and finds the factors and the correlation rules which have an important influence on the software reliability. Then it could be used to evaluate the software reliability.

Keywords: rough set, software reliability, reliability evaluation, qualitative evaluation.

1 Introduction

Software reliability evaluation which is one of the important contents of software reliability engineering could give a guidance of software reliability in software development process to judge whether the various software design and development technologies have a positive role in improving the software reliability, and to evaluate the reliability of the final delivery of software products. It also could give the quantitative result of the software products' final reliability level.

In recent years, the quantitative software reliability evaluation method based on failure data has been developed rapidly. However, its defect is that the method must depend on the failure data, which leads to that they can only be used in the later period of software development. Consequently, various reliability comprehensive evaluation methods have been put forward, such as software reliability comprehensive evaluation based on grey system theory[1], fuzzy comprehensive evaluation model[2], early software reliability prediction method based on fuzzy neural network [3] and software reliability metrics selecting method based on analytic hierarchy process[4] etc. They all try to take use of the various correlative factors during the whole software product life cycle to give a qualitative evaluation and analysis of software reliability. IEEE standard of 982.1 also points out that the correlative factors of

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 357–365.
springerlink.com

various products, processes and resources in the whole software life cycle have a very important effect on software reliability [5]. Reference [6] has proposed the factors by questionnaire survey. Since then, the problem becomes that we need to find a more excellent method which could more accurately give the quantitative relationship between the factors and the software reliability. The rough set theory can effectively dispose the inaccurate, inconsistent and incomplete information, and find their implicit knowledge to reveal the potential law [7]. Therefore, it has been widely used in machine learning, data mining and artificial neural network etc. In this paper, we analyzed the reliability measurement data by using the methods and models of rough set, and explored a new way of software reliability evaluation.

2 The Analysis of the Factors Affecting Software Reliability Evaluation

Reference [6] surveyed 13 organizations' software development process, such as AT&T, BellCore, Chrysler and MCI International etc., and proposed 32 main factors which have a great influence on software reliability, and sorted them according to their degree of influence on the reliability. Based on [6] and [2] [8], 30 common environmental factors shown in Table 1 are divided into three categories, such as product factors, resources factors and process factors.

Table 1. Common factors affecting the software reliability

categories	No.	Name of the environmental factors
Product factors	1	Software complexity
	2	Reuse code number
	3	The quality objectives of users
	4	Software programming language
	5	The attributes of failure and defects
	6	Developers' skills and comprehensive quality
Resources factors	7	The scale and skills of development and test team
	8	Testers' skills and comprehensive quality
	9	The comprehensive quality of users
	10	Testing tools
	11	Testing environment
	12	Hardware resources
	13	Software development environment
	14	Development and test cost
Process factors	15	Testing methods
	16	Testing coverage
	17	Test cases
	18	Fault checking and debugging process
	19	Testing effort
	20	Data collection
	21	Software reliability design method and technology
	22	Software development design method and technology
	23	Software operations
	24	Software quality management activities
	25	Documents and phase evaluation standard
	26	The quality of demand analysis and the detailed design
	27	Software changes activities
	28	The difficult of programming and development
	29	Integral scheduling
	30	Development effort

Software product factors are the attributes or features of the software itself which are relative with the software size, documents and structure. They are static characteristics of software, and can be collected from the software design documents.

Resources factors are divided into three categories: people, the reusable software components and hardware and software environments.

Process factors are the characteristics and behaviors which are existed in the processes of development, testing and maintenance etc..

It is thus clear that the factors affecting the software reliability are various. If the subordinate factors could be removed, it can help to not only improve the accuracy of qualitative evaluation, but also reduce the difficulty of the data collection. In this paper, we try to take use of the data mining technology based on rough set to find the key factors influencing the level of reliability, and to analyze their relationship with the reliability evaluation result.

3 Data Preprocessing

In order to simply and intuitively illustrate the method, and to reduce the computation complexity, seven key factors are selected from Table 1, as shown in Table 2.

Table 2. Factors and decision list

No.	Name of the environmental factors
A	Software reliability design method and technology
B	The scale and skills of development and test team
C	Test cases
D	Testing coverage
E	Software operations
F	Documents and phase evaluation standard
G	Software complexity
R	Reliability evaluation result (decision attribute)

When using the rough set theory to dispose the information table which is also called decision table, the attribute values in information table must be discrete. Therefore, the values of the factors and the decision attribute should be discretized before carrying through attribute reduction. The details are as follows:

A: 3-using a suitable fault tolerant, avoidance, correction and check design; 2-using a limited fault tolerant, avoidance, correction and check design; 1-almost no use of any software reliability design methods and techniques.

B: 3-with a large-scale and experienced; 2-with a general scale and general experience; 1- with a small scale and almost no experience.

C: 3-with high quality and good implementation; 2-with general quality and general implementation; 1-with poor quality and poor implementation.

D: 3-all requirements of software functions and performance, and all paths, branches and interfaces have been covered; 2-the main software functions and performance and paths, branches, interfaces have been covered; 1-only a few software functions and performance and paths, branches, interfaces have been covered.

E: 1-software running in good condition and no failure occurs; 0-software running condition is not stable, and some failures occur.

F: 3-documents are complete, and have been reviewed seriously; 2-documents are not complete, and have not been reviewed seriously; 1-almost no documents.

G: 3-software complexity is low; 2-software complexity is in general; 1-software complexity is high;

R: 3-reliability is high; 2-reliability is in general; 1-reliability is low.

Table 3 is a decision table about the relationship of above factors and reliability level given by the reviews of experts.

Table 3. Decision table of software reliability evaluation

No.	A	B	C	D	E	F	G	R
1	2	3	2	3	1	3	2	3
2	3	2	3	3	1	2	3	3
3	3	2	3	3	0	2	3	3
4	3	3	2	3	0	2	3	3
5	3	2	3	3	1	2	1	2
6	2	2	3	2	0	2	2	2
7	3	3	3	1	0	2	3	2
8	1	2	2	3	1	2	1	1
9	2	2	3	1	0	2	2	1
10	3	3	2	1	0	2	3	2
11	1	2	2	1	0	2	1	1
12	3	2	2	1	1	2	1	2

4 The Attribute Reduction Method Based on Discernibility Matrix

Reduction is an important concept in rough set theory. In the information system, there are often a large number of redundant attributes. The attribute reduction is very necessary to reduce the redundancy and improve the intelligibility of knowledge contained in the information system.

In table 3, the seven factors are all important on the final software reliability evaluation. However, according to the importance of attribute and the reduction algorithm, the attributes of decision table should be reduced.

Here, using the discernibility matrix [7] presented by Skowron to obtain the best attribute reduction.

4.1 Discernibility Matrix

Discernibility matrix is defined as follows [7]: Assuming $S=(U,R,V,f)$ is an information system, and U is the domain and $U=\{x_1,x_2,\cdots,x_n\}$. $R=C\bigcup D$ is a set of attributes. C is the set of condition attributes and D is the set of decision attributes. V is the set of range of the attributes, and $f:U\times R\to V$ is an information function which specifies attribute values of each object in U. $a(x)$ is the value of the object x in terms of the condition attribute a.

Then the discernibility matrix can be expressed as follows:

$$(c_{ij}) = \begin{cases} \{a \in C : a(x_i \neq x_j)\}, D(x_i) \neq D(x_j) \\ 0, D(x_i) = D(x_j) \\ -1, \{\forall a, a(x_i) = a(x_j), D(x_i) \neq D(x_j)\} \end{cases} \quad (1)$$

4.2 The Attribute Reduction Algorithm

The discernibility matrix definition declares that: the elements whose attribute combination equals to 1 in the matrix indicate that except for these condition attributes, the rest condition attributes couldn't distinguish between two different records, which means that this attribute must be retained. The concept is consistent with the core attributes in the decision table of rough set. Therefore, all the attributes whose attribute combination equals to 1 are the core attributes of the decision table (the core attribute may be nonexistent).

Since the discernibility matrix contains all attribute distinguish information in the decision table, the remaining attributes which should be retained need to be obtained from those whose attribute combination is not equal to 1 except for the core attributes. Assuming that there are only two attribute combinations remain except for the core attributes, like $a_1 a_2 \cdots a_p$, $b_1 b_2 \cdots b_q$. Using Boolean value to judge whether this attribute combination contains a certain condition attribute. For example, $a_1 = 0$ means that it doesn't contain the condition attribute a_1, and $a_1 = 1$ means that it contains the condition attribute a_1. According to the theory of discernibility matrix, if all different decision records are required to be identified, at least one attribute must be retained in each $a_1 a_2 \cdots a_p$ and $b_1 b_2 \cdots b_q$. Constructing an expression $P = (a_1 \vee a_2 \vee \cdots a_p) \wedge (b_1 \vee b_2 \vee \cdots b_q)$, and $P = 1$ from the above analysis. Making P into a disjunctive normal form and letting the value of any conjunct of P is equal to 1, then the attribute combinations could distinguish all the decisions of the original decision table combined with the core attributes.

Then the application can be adopted based on discernibility matrix reduction algorithm for attribute reduction in Table 3. The discernibility matrix is shown in Table 4.

Table 4. Discernibility matrix

	1	2	3	4	5	6	7	8	9	10	11	12
1	0											
2	0	0										
3	0	0	0									
4	0	0	0	0								
5	ABCFG	G	EG	BCEG	0							
6	BCDEF	ADEG	ADG	ABCDG	0	0						
7	ACDEFG	BDE	BD	CD	0	0	0					
8	ABFG	ACG	ACG	ABEG	AC	ACDEG	ABCDEG	0				
9	BCDEF	ADEG	ADEG	ABCDG	ADEG	D	ABG	0	0			
10	ADEFG	BCDE	BCDE	D	0	0	0	ABDEG	ABCG	0		
11	ABDEFG	ACDEG	ACDG	ABDG	ACDE	ACDG	ABCG	0	0	ABG	0	
12	ABDFG	CDG	CDEG	BDEG	0	0	0	AC	ACEG	0	AE	0

According to Table 4, the core attributes set is{D,G}, and the attribute combinations which do not contain the core attributes are AC and AE. Then the corresponding logical expression derived from Table 4 is:

$$P = \mathrm{D} \wedge \mathrm{G} \wedge (\mathrm{A} \vee \mathrm{C}) \wedge (\mathrm{A} \vee \mathrm{E})$$

Then, we have

$$P = (A \wedge D \wedge \mathrm{G}) \vee (\mathrm{C} \wedge \mathrm{D} \wedge \mathrm{E} \wedge \mathrm{G})$$

Therefore, the corresponding attribute reduction is: {A, D, G} or {C, D, E, G}.According to the principle of the least number of attributes, the best combination of attribute reduction selected is: {A, D, G}.The decision table after reduction is shown in Table 5.

Table 5. The decision table after reduction

No.	A	D	G	R
1	2	3	2	3
2	3	3	3	3
3	3	3	3	3
4	3	3	3	3
5	3	3	1	2
6	2	2	2	2
7	3	1	3	2
8	1	3	1	1
9	2	1	2	1
10	3	1	3	2
11	1	1	1	1
12	3	1	1	2

By making the attribute reduction of decision table of software reliability evaluation, we get the key factors impacting on the software reliability evaluation. They are software reliability design methods and techniques, test coverage and software complexity. These three factors play a key role on software reliability evaluation.

5 Attribute Value Reduction and Rule Generation

Rough set theory could extract rules from the decision table. Actually, the process of extracting rules in rough set theory is the process of value reduction to decision table. Attribute value reduction algorithm starts from the core value. Core value is a very important concept in the process of attribute value reduction. It is the attribute value which could not be deleted in decision table. The basic idea is that after deleting a certain attribute value for each instance, and testing whether this record is conflict with other records. If a contradiction exists, it shows that this attribute value could not be deleted; otherwise, this attribute value could be deleted. The specific algorithm is as follows [9]:

Step1. Deleting the possible repeated records and making an on-the-spot investigation by column for the condition attribute in decision table. After removing the column, if there exist conflict records, the attribute value of the conflict records

must be kept as original; if there are no conflict records, but contain repeated records, marking "*" as the attribute value of the repeated records; for other records, marking "?" as the attribute value.

Step2. Deleting the possible repeated records and reviewing each record which contains "?". If the decision could be judged by the unmarked attribute values, then marking "?" with "*" . Otherwise, the "?" must be changed into the original value; if all the condition attribute values of a record are marked, then the "?" must be changed into the original value.

Step3. Removing all the records whose condition attribute values are marked with "*" and the possible repeated records.

Step4. If only one condition attribute value are different from two records, and one of them is marked with "*", for this record, if the decision could be judged by the attribute values which have not been marked, then deleting the other record; otherwise, deleting this record.

After attribute value reduction, we get a new decision table. Thus, all attribute values are the core values, and all records form the rules of the decision table.

For example, bring the above steps into decision Table 5 to get the core values and rules.

Firstly, using step 1 we have the Table 6. Then, using step 2 we have the Table 7.

Table 6. The decision table

No.	A	D	G	R
1	?	3	?	3
2	?	3	3	3
5	3	*	1	2
6	?	2	?	2
7	?	1	*	2
8	1	*	?	1
9	?	1	?	1
11	1	*	?	1
12	3	*	*	2

Table 7. The decision table

No.	A	D	G	R
1	2	3	*	3
2	*	3	3	3
5	3	*	1	2
6	*	2	*	2
7	3	1	*	2
8	1	*	*	1
9	2	1	*	1
11	1	*	*	1
12	3	*	*	2

In Table 7, comparing the record 5 with the record 12, because the decision could not be derived only by A=3, then the record 12 must be deleted. Then, delete the possible repeated record 11. Finally we get the decision Table 8 below after attribute value reduction.

Table 8. Decision table after attribute value reduction

No.	A	D	G	R
1	2	3	*	3
2	*	3	3	3
5	3	*	1	2
6	*	2	*	2
7	3	1	*	2
8	1	*	*	1
9	2	1	*	1

The potential rules of the original decision table reflected by table 8 are as follows:

Rule1: if (A=2) and (D=3), then R=3; Rule2: if (D=3) and (G=3), then R=3; Rule3: if (A=3) and (G=1), then R=2; Rule4: if (D=2), then R=2; Rule5: if (A=3) and (D=1), then R=2; Rule6: if (A=1), then R=1; Rule7: if (A=2) and (D=1), then R=1.

The explains of the above rules are as follows: such as Rule1: If using a limited fault tolerant, fault avoidance, fault correction and fault check design, and testing covered all software functions and performance requirements along with all of the program paths, branches and interfaces, then the software reliability level is high; Rule5: If using a suitable fault tolerant, fault avoidance, fault correction and fault check design, however, only a few software functions and performance and paths, branches, interfaces have been covered, then the software reliability level is in general; Rule6: If almost no software reliability design method and technology are used, then the software reliability level is often low. It shows that in the software development, adopting appropriate software reliability design methods and techniques, and covering every function point and performance requirements, and all of the branches, paths and interfaces as far as possible during testing are good for improving the level of software reliability.

6 Conclusion

This paper presents a software reliability evaluation method based on Rough Set. At first, to determine factors that may have influence on the software reliability. Then, using attribute reduction and value reduction methods in Rough Set theory, and through analyzing the relationship between these factors and software reliability to identify factors that have a significant impact on the software reliability, and to determine their correlation rules with reliability. Thus, get a comprehensive evaluation result of software reliability. In the terms of the implementation process of this method, not only the software reliability evaluation result could be achieved, but that the feedback of factors determination could guide software development and testing. It has practical engineering value and the future work should be about appropriate factors' determination and further analysis and refinement of the method.

References

1. Li, H., Lu, M., Wang, Z., Li, Z.: Framework for software reliability comprehensive evaluation based on grey system theory. Journal of Beijing University of Aeronautics and Astronautics 34(11), 1261–1265 (2008) (in Chinese)
2. Wang, T., Li, M.: A Fuzzy Comprehensive Evaluation Model for Software Reliability. Computer Engineering and Applications 38(20), 23–26 (2002) (in Chinese)
3. Liu, B., Lu, M.-Y., Ruan, L.: Early Software Reliability Prediction: an Approach Based on Fuzzy Neural Network. Journal of Beijing University of Aeronautics and Astronautics 27(2), 237–240 (2001) (in Chinese)
4. Li, H., Lu, M., Li, Q.: Software Reliability Metrics Selecting Method Based on Analytic Hierarchy Process. In: Proceedings of QSIC, pp. 337–346 (2006)

5. IEEE Std 982.1-1988: IEEE Standard Dictionary of Measures to Produce Reliable Software, IEEE (1988)
6. Zhang, X., Hoang, P.: An analysis of factors affecting software reliability. The Journal of Systems and Software 50, 43–56 (2000)
7. Wang, G.: Rough set theory and knowledge acquisition. Xi'an Jiao tong University Press, Xi'an (2001)
8. Pham, H.: Software Reliability. Springer, Singapore (2000)
9. Chang, L., Wang, G., Wu, Y.: An Approach for Attribute Reduction and Rule Generation Based on Rough Set Theory. Journal of Software 10(11), 1206–1211 (1999)

Kinematics Models and Development of Control System of 2-PRR Parallel Machine

Ming Hu

School of Mechanical Engineering & Automation, Northeastern University, Shenyang
Minghu@mail.neu.edu.cn

Abstract. It is very important to design motion control system for parallel machine. A 2-DOF in-parallel machine tools based on a tripod mechanism is developed and studied, the kinematic is analytically performed, the workspace is derived, The design idea to the motion control system of the parallel robot was presented. This paper deals with the methodology for developing the ADV400 Numerical Control System for the parallel machine tools. The effectiveness of this methodology has been exemplified by the successful application to a parallel mechanism.

Keywords: Parallel mechanism, Kinematics equation, Workspace, CNC.

1 Introduction

Kinematical analysis is a means to reveal the nature of body movement and it is the basis of analysis of other agencies. 2-PRR parallel mechanism is a typical lower-mobility parallel mechanism; less free mechanism are the most widely used family members in the mechanism [1]. Therefore, fewer degrees of freedom kinematics in-depth study will not only help understanding of existing character as a small degree of freedom, but also to grasp and develop new parallel mechanism is of special importance [3]. This study focused on the content of the space model of parallel mechanism, size, type, structural design of parallel mechanism, workspace, kinematics of the positive obtained, The agency for further analysis, research and development and application provides a theoretical basis. When the servo drive rod coordinated motion, the moving platform of the parallel machine tool is moved along the space coordinate. Only each driving rod length change, the required trajectory of moving platform is able to get. Therefore, in the motion control, the position inverse solution model must be used. Will be given in advance of the moving platform position and attitude and velocity information is transformed into the servo system control instructions, and drive the moving platform of parallel mechanism to achieve desired movement, control method is by the means of kinematics method, in paper Control law is simple, easy to implement to the 2-DOF parallel machine tool, so ADV400 CNC machine tool control system is selected as the hardware platform. The CNC system also has a good open and is relatively easy to implement multi-axis control.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 367–372.
springerlink.com

2 Parallel Mechanism and the Degree of Freedom

The 2-PRR Parallel Machine Tool body diagram is shown in Fig. 1, in the kinematic diagram. A, B, C, D respectively indicates the location of each hinge of the Mechanism, x_A, x_D indicates hinge A, D respectively in the Cartesian coordinates of the X-coordinate; $P(x_p, y_p)$ indicates that the location of the end of the actuator; Given that $L_i(i=3,4,5,6,7)$ indicates the actual length of the rod respectively, and $L_3 = L_4, L_6 = L_7$. Given that θ_i $(i = 3,4,5,6,7)$ respectively indicates the positive angle of the Link i and the X-axis, with $\theta_3 = \theta_4, . \theta_6 = \theta_7$, $\theta_5 = 0$ H indicates the distance between $P(x_p, y_p)$ and L_5.Taking into account the existence of mechanism of local freedom and virtual computing the impact of constraints on freedom, with freedom and virtual constraints of local mechanism, in the calculation of degrees of freedom when you can first draw to remove redundant constraint and local degrees of freedom body diagram, then calculated by the formula[2]. A total of six events in which components, 8 lower pairs (including two prismatic pairs and 6 revolute pairs) with no higher pair and no local constraints, and then set n=6, $P_L = 8$, $P_H = 0$ into the equation, then the body degrees of freedom for：

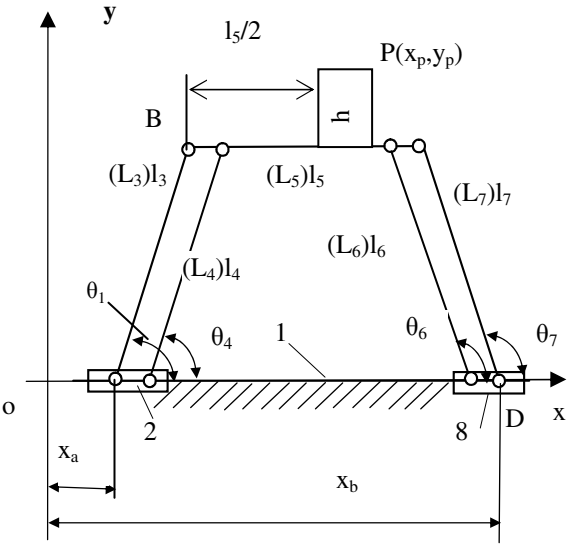


Fig. 1. 2-PRR parallel kinematic diagram

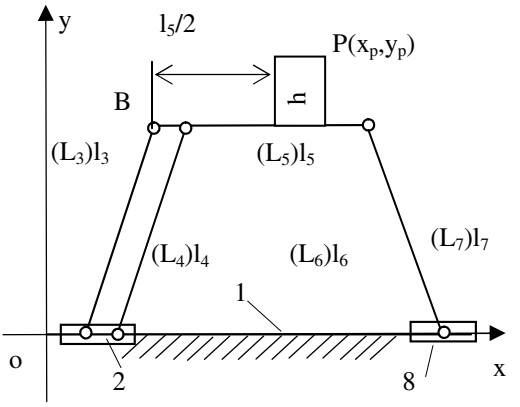


Fig. 2. 2PRR simplified schematic of parallel mechanism

$$F = 3n - (2 \times P_L + P_H) = 3 \times 6 - 2 \times 8 = 2 \tag{1}$$

The agency consists of two motor drives, motor drove screw rotation, two separate slider bodies as an input, and meet the conditions as the main movement module for the two-DOF mechanism.

3 Kinematic Equation

2-PRR parallel kinematic machine tool main movement diagram shown in Fig.1, $L_i (i = 3,4,5,6,7)$ respectively indicates the actual length of rod, and $L_3 = L_4$ $L_6 = L_7$; $\theta_i (i = 3,4,5,6,7)$ respectively indicates the positive angle of the Link i and the X-axis, and $\theta_3 = \theta_4$, $\theta_6 = \theta_7$, $\theta_5 = 0$, $\theta_3 = \pi - \theta_7$. The coordinates of the two sliders are respectively (x_A, y_A),(x_D, y_D), According to the construction of the coordinate system, in which $y_A = y_D = 0$. By the coordinate system shown in Fig. 1 the relationship between the size and structure of export we can overturn the positive solution equation of P point with the end of the implementing agencies:

$$\begin{cases} x_p = \dfrac{x_A + x_D}{2} \\ y_p = \sqrt{L_3^{\,2} - \left(\dfrac{x_D - x_A - L_5}{2} \right)^2} + h \end{cases} \tag{2}$$

the negative solution equation for:

$$\begin{cases} x_A = x_p \mp \sqrt{L_3^{\,2} - (y_p - h)^2} - \dfrac{L_5}{2} \\ x_D = x_p \pm \sqrt{L_3^{\,2} - (y_p - h)^2} + \dfrac{L_5}{2} \end{cases} \tag{3}$$

In which the size of the sign by the decision of θ_3, when it is less than ninety degrees, x_A take the minus sign, x_D take the plus sign. When the angle is greater than ninety degrees, take the contrast. As the article itself actually 2PRR parallel mechanism and the effect of gravity limit the movement of θ_3 not more than ninety degrees, so:

$$\begin{cases} x_A = x_p - \sqrt{L_3^{\,2} - (y_p - h)^2} - \dfrac{L_5}{2} \\ x_D = x_p + \sqrt{L_3^{\,2} - (y_p - h)^2} + \dfrac{L_5}{2} \end{cases} \tag{4}$$

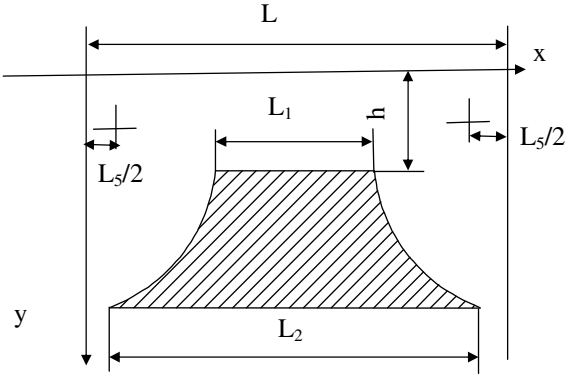


Fig. 3. The Parallel mechanism's workspace

As the 2-PRR parallel mechanism in the X-Y plane for translational, two sliders do the same straight line, so drivers will not interfere with blocks, so this body is mainly due to track length and angle between two factors affecting rod work space. Taking into account the moving platform for translation, space can be calculated by tip-point $p(x_p, y_p)$, Given that the length of the guide is L, $L_3 = L_7 = l$ $\theta_3 = \pi - \theta_7$, the distance between Point P and L_5 is h. So the workspace of the 2-PRR parallel mechanism is as shown in Fig.3: By the diagram of workspace we can see that their work space for a graphic similar to the isosceles trapezoid, non-empty, non-

consecutive points or lines. The moving platform is only changes in position, no attitude change, for a rectangular operations, milling, grinding, turning, grinding and other processing of Billet; the work intuitive, simple space, and the conditions of the given, basic size work space can be easily calculated to facilitate the processing of the calculation Workspace size is determined by L, L_3 and L_5, Location by h and L_5. The width of workspace is $L_1 = L - 2L_3 - L5$, $L_2 = L - L_5$, therefore, the theoretical area of work space is:

4 Control Systems

For consistency with conventional CNC programming and the traditional CNC machine code directly applied to the control of parallel. to reduce the development effort, Advantage 400 CNC is used in the parallel machine, which is shown in Fig 4 and built The PC104 computer system. The platform embedded Windows CE NET, the CNC Use the PC for display, information input and other non-real-time processing, support G-code, Lower machine multi-axis controller card for the PMAC2, The equivalent of a real-time multi-tasking computer, PMAC can run independently stored in the internal procedures and perform multiple tasks simultaneously and prioritize correctly. The interpolation operation, servo control, real-time tasks completed by the PMAC motion control, the user can operate according to equations of motion to space agency into a joint space motion movement. the two AC servo motors with pulse mode control can be use in the machine.

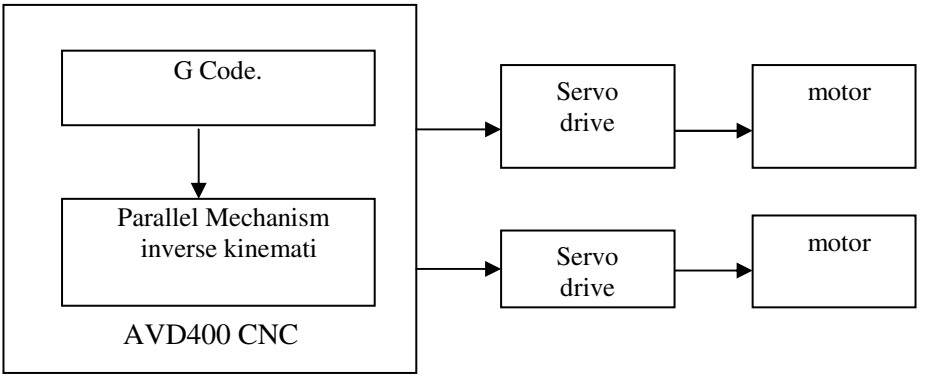


Fig. 4. Control system block diagram

5 Conclusions

In this paper, the 2-PRR Parallel mechanism of the kinematic properties is first introduced. the parallel machine tool has two planar degrees of freedom, and their

work space for a graphic is similar to the isosceles trapezoid, no empty, non-consecutive points or lines. The moving platform is only the change in position, no attitude change, for a rectangular operation, milling, grinding, turning and other processing of Billet. The 2-PRR-type parallel kinematic machine tools equations and Jacobian matrix for a simple display of expression is clear computing simple, convenient control, and space objects in the whole movement interference and coupling. Parallel mechanism and its easy location of the singularity analysis, the actual processing can effectively prevent the singular position of occurrence of manipulability well. ADV400 CNC system is able to control 2-PRR parallel machine tool to achieved 2 degrees of freedom movement by mean of G-code programming. It is a very broad application of less-DOF Parallel Machine Tool.

References

1. Zhen, H., Yiyuan, Q.: Journal of Yanshan University 13(2), 2–6 (1989)
2. Xin, Z., Shun-Cheng, Xu, X.-Y.: Mechanical Design (Supplement), 75–76 (2002)
3. Wang, X.M., Fu, Y., Cai, G.Q., Hu, M.: Journal of Northeastern University(Natural Science) (6), 685–687 (2001)
4. Honegger, M., Codourey, A., Burdet, E.: Proc. of the 1997 IEEE International Conference on Robotics and Automation, Albuquerque, New, Mexico, pp. 543–548 (1997)
5. Yiu, Y.K., Li, Z.X.: Proc. of International Conference on Mechatronics Technology, Singapore, pp. 339–344 (2008)

Development and Realization of Crude Oil Measuring Technology

Gongchang Ren, Bo Chen, Qing Ye, and Yongfei Wang

College of Mechanical & Electrical Engineering,
Shaanxi University of Science & Technology, Xi'an, China, 710021
rengc@sust.edu.cn, chen_bo_16@163.com,
419727812@qq.com, yongfeio@126.com

Abstract. Currently oil-water interface measurement technique in China falls behind the other countries, due to hanging oil, high cost and large errors. Thus it's necessary to develop a real-time, low cost, reliable, cost-effective on-site measuring device. Electromagnetic wave shows different characteristics when spreads in multiphase media. Base on this unique identity, we adopted matrix electromagnetic sensors to distinguish crude oil layer, oil-water mixed layer and water layer.

Keywords: matrix electromagnetic sensors, oil-water measurement, single-chip microcomputer (SCM).

1 Introduction

Generally the newly extracted crude oil contains large amount of water, it must be sent into the oil-water separation tank first standing for some time. Under the action of gravity, mixture water-drops would gather into big ones and precipitate to the bottom of tank, meanwhile, the oil is floating on the water. The interface detection of the crude oil layer, oil-water mixed layer and water layer is crucial during production process. However, the physical property of oil-water mixture is complex and oil-water interface detection is difficult, the commonly used detection method in our country is still the original manual work.

According to the characteristics of oil-water interface, based on the different characteristics electromagnetic wave propagation in heterogeneous media. We adopted the detection method of using matrix electromagnetic sensors to distinguish crude oil layer, oil-water mixed layer and water layer [1].

2 Oil-Water Interface Detection Principle

Since certain changes of the media in the space would cause a lot of physical quantity changes. For example, the changes in sensitive electric field and magnetic field. The physical quantity changes of electric and magnetic fields will indirect reflect and represent the media unit volume and changes in media properties. It is indicated that the dielectric constants of oil and water vary widely. Different media leads to a huge

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 373–377.
springerlink.com

discrepancy in parameters. At room temperature, the relative dielectric constant of water is 80 but that of pure crude oil is about 2.3. There are nearly 40 times away from each other. Based on different media has different absorbed degrees for electromagnetic wave, we adopt electromagnetic wave absorption liquid level gauge to detect water content in oil and liquid level [2]. The liquid level gauge is consisted of a transmitter and measuring rod. Electromagnetic waves radiated outward through the measuring rod. The ability to absorb electromagnetic waves is different in different media. The consumption of electromagnetic energy in a certain medium shows the depth of the liquid level [3].

According to this principle, we have designed a matrix electromagnetic probe with different bands of microwaves launching and connecting their emission electrodes, forming several level detection cross-sections with the tank wall. The intensity of microwave transmission attenuation in each cross-section is connected with the nature of the medium. It is the strongest in water, then the mixed and then the pure oil and last the air. Under the control of the CPU, oil-water interface gauge, through the matrix probe structure, signal a constant horizontal scanning microwave with error prevention identification code. Meanwhile, the receiver starts receiving echo signal strength decayed by the medium. Then the gauge turns the weak echo signal into a current pulse signal through a special anti-jamming circuit filters out invalid and environmental superimposed current signal, the remaining pure and effective measurement signal is restored through the demodulator to pulse voltage signal with error prevention identification code which has the same wave form but the different amplitudes compared with the launching period, and finally through the high-speed and high-precision analog-digital conversion circuit turned into a digital signal and stored. Then the interface gauge continues to collect data of the next matrix point until complete. We analyses and calculates these data to get the information of liquid level and location of oil-water interface.

According to the above measuring principle and method, we constructed the oil storage tank oil-water interface gauge. Matrix electromagnetic probe is used with different bands of microwaves launching and connecting their emission electrodes, and forming with the tank wall several level detection cross-sections. The SCM scans and detects each electromagnetic quantity and outputs the results by standard signal 485 to achieve the detection of oil, water and oil-water mixture.

Schematic diagram of the sensor is shown in Figure 1. The probe is a standard length component. Because the media between Polyethylene PTFE Tube and tank wall is different, so the sensor acquires location of the liquid level through estimating the decrement of microwave attenuation during the testing process.

3 Overall Design

The oil-water interface gauge is consisted of array probe which installed in parallel with the tank wall, the microwave electromagnetic scanning and detection module, the signal conditioning module, digital-analog conversion module, RS485 communication interface and SCM [4].

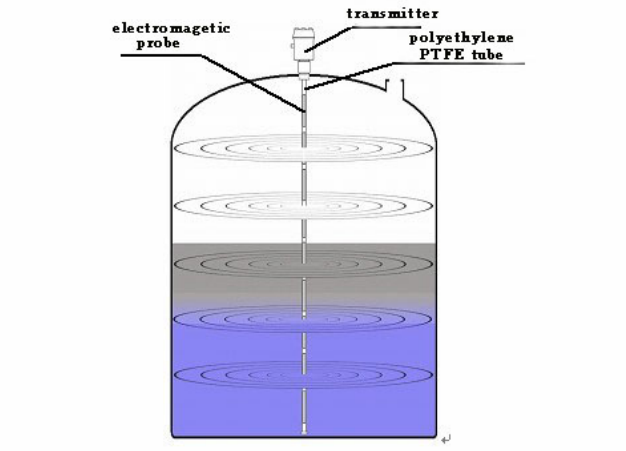


Fig. 1. Principle of the sensor

After receiving signal via RS485 network, SCM controls probe to scan detection module and measures each point on the matrix probe. The collected parameters are transmitted to SCM through the signal conditioning and analog conversion module. After gathering all the electromagnetic information, SCM organizes data in accordance with the communication protocol and transmits the date back to the computer in monitoring center through RS485 network [5]. The overall structure is shown in Figure 2.

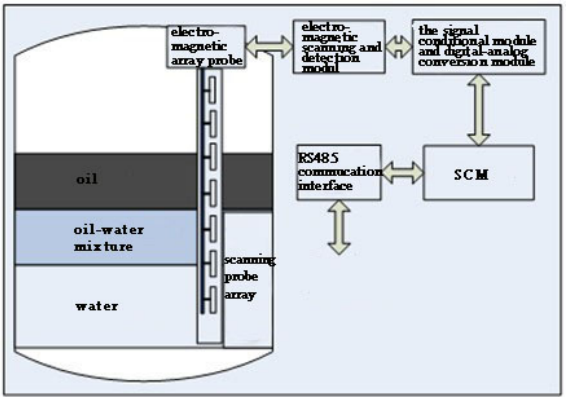


Fig. 2. Overall structure of the oil-water interface gauge

The oil-water interface gauge uses matrix electromagnetic sensors to detect oil-water interface online and turns the changes in oil-water interface into electrical signals changes. The circuit transforms the electromagnetic signal from object into electrical signals. Data acquisition module contains magnetic sensors, magnetic detection circuit, constant-current source circuit, square wave generator circuit formed

by 555 timer, amplifier circuit and A / D converter circuit. The whole measurement system uses AT 89C51 SCM which is strong-functioned and low-priced as processor [6]. Overall system block diagram is shown in Figure 3.

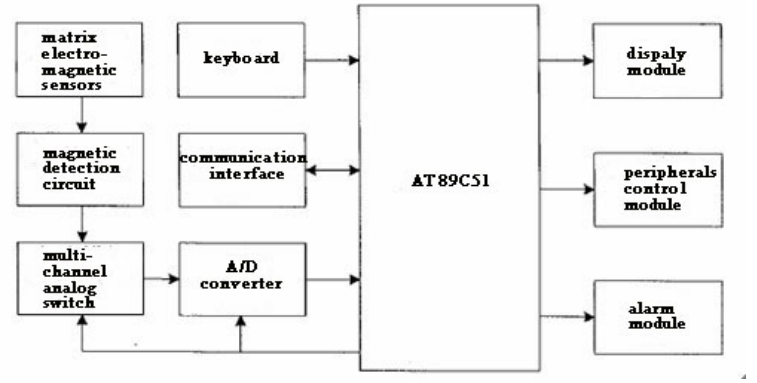


Fig. 3. Overall system block diagram

Installation impression drawing of new oil-water interface detection system is shown in Figure 4.

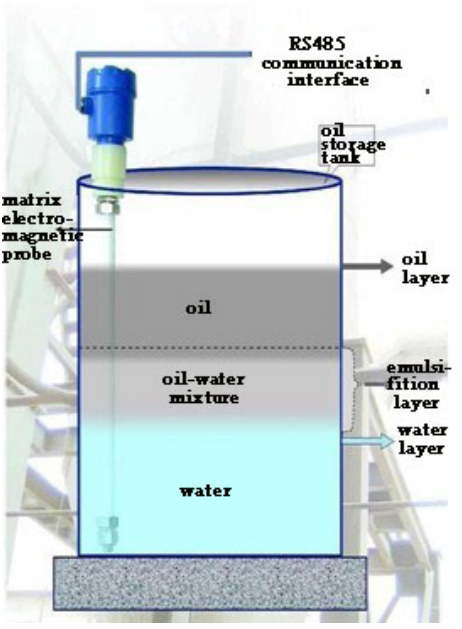


Fig. 4. Installation impression drawing of the oil-water interface gauge

4 Conclusion

In contrast to the traditional differential pressure interface measurement and ultrasonic interface measurement, this design enjoys higher accuracy, lower cost, and easy to install and maintain. It saves both effective human resources and makes the measurement results more timely, covers shortcomings of traditional measurement with a bright future for promotion.

References

1. Hamelain, J.: Liquid Level Control Using a Electromagnetic Sensors (2005)
2. Chang, Q.: Oil field Branch Company Gas Production Factory. Single Well Measuring Technology of Changqing Gas Field. J. Petroleum Planning & Engineering 7, 13–14 (2008)
3. Able Instruments & Controls Ltd, Sensors for Level Measurements (2003)
4. Weilin L., Guansheng Y., Shan H.: SCM—AT89S51 & AVR (2008)
5. Bo, Y., Tao, X., Tianwen, D.: Application and Discussion of Oil Field Metering Station Automation System. J. Petroleum Planning & Engineering 7, 37–38 (2009)
6. Clifford, M.: Water Level Monitoring (2006)

A Machine Vision System for Chinese Chess-Playing Robot

Jianjun Fang

School of Automation, Beijing Union University, Beijing, 100101, China
fangjj1947@126.com

Abstract. This paper introduces a machine vision system for Chinese chess-playing robots which is usually regarded as a form of recreation. The machine vision system with two color cameras takes simultaneously two images of a chessboard and round pieces on the chessboard from different angles (views). Firstly, original images are handled by a series of image processing operations such as color conversion, binarization and denoise. Secondly, a hierarchical Hough transform algorithm was taken to detect lines and circles in the binarized image. Circles with reasonable radii nearly centered on the crossings of chessboard will be considered as pieces on the chessboard and its corresponding Chinese character inside the circle would be recognized based on BP neural network and ring intersection points. The 3D coordinates of pieces can be calculated through computation and then transformed to coordinates expressed by row and column. Finally, results of image processing and pattern recognition procedures will be sent to robot control system to manipulate end-effector of a robot to move round pieces from one place to another desired. Experimental results reveal that the designed machine vision system in this paper can work well with higher reliability.

Keywords: Machine vision, Chinese chess-playing robot, Hierarchical Hough transform, BP neural network.

1 Introduction

The third generation of robots is called intelligent robots and becoming more attractive to researchers and high-tech companies from any corners of the global. Some robots can be found in home application fields [1] with the help of high-tech development and expanding future market from an upcoming aged society. As a form of intelligent robots, Chinese chess-playing robots are potential to be a successful application because of a larger Chinese population who enjoys playing Chinese chess as a form of recreation. Studies show that playing the game is helpful to prevent aged players from senile dementia and brings players joyful experience when he/she can defeat a robotic competitor which is a symbol of high technology. The Chinese chess-playing robots have a good future in China because whether you are an aged person or not, you will need a robotic player who has good temper and enough time to accompany you to play at any time and anywhere.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125,, pp. 379–385.
springerlink.com

The machine vision system is essential to Chinese chess-playing robots. The vision system can answer two questions. On one hand, machine vision system should have an ability to recognize Chinese characters on the top surface of round pieces correctly and efficiently. On the other hand, the vision system can calculate three dimensional (3D) coordinates of round pieces on the chessboard through camera calibration and indirect computation and then manipulate robot's end-effector to move the desired round piece. Literatures show some researchers began to design and develop machine vision system for Chinese chess-playing robots [2, 3]. This paper places emphasis on two questions mentioned above with new ideas and techniques.

This paper is organized as follows. We describe related background for developing Chinese chess-playing robots and construction of machine vision system of this robot in Section 1. Further, we discuss algorithm for detecting lines and circles using hierarchical Hough transform in Section 2. In Section 3, we describe Chinese characters recognition based on ring intersection points and BP neural network. Finally, We conclude this paper in Section 4.

2 Hierarchical Hough Transform (HHT) to Detect Lines and Circles

Hough transform (Hough, 1962) has been proven to be successful in terms of accuracy, convenience and robustness [4]. Since its invention, Hough transform had been widely used in many applications. Actually, any objects which are represented by parametric equations can be detected using Hough transform. The standard Hough transform can't provide length of line segment and coordinates of two endpoints. However, improved Hough transform can do that [5, 6]. Circles in real-world images are frequently found to be discontinuous. The Hough transform is also well-known to be an efficient tool for detecting discontinuous circles, even though embedded in real-world noisy images. Discrimination of circles in image using HT has been discussed in [7, 8].

Hough transform is a voting process which each feature point votes for all possible patterns passing through it. The votes are accumulated in an accumulator array and pattern which gets maximum votes is the desired one to be detected. Although Hough transform is an efficient and robust tool to detect patterns in images, it also has fatal drawbacks of time-consuming computation and huge storage demands which prevent it from widely application in industry. Many researches had been devoted to efficiency and storage improvement of Hough transform. Some researchers put forward hardware solutions [9] to overcome huge computational overhead and others paved another way to improve the performance of HT through software evolution [10].

2.1 Hierarchical Hough Transform(HHT)

Because Hough transform is a voting process, every image point is mapped to one parameter point, memory storage and computation time increase significantly. Hierarchical Hough transform (HHT) is an improvement by applying coarse-to-fine searching strategy[11,12]. An original image was divided into a set of small sub-images and then HT is performed on these sub-images. Due to small sub-images, size of accumulator array can be kept small, which will lead to an efficient and robust process.

For standard HT, there need two accumulators for lines and three ones for circles. The size of accumulator depends on detection accuracy. As far as detection of circles is concerned, analytical equation of a circle can be expressed as

$$(x-a)^2+(y-b)^2=r^2 \tag{1}$$

Where (a,b) represents coordinates of circle's center point and r is the radius of a circle. Supposed that the variation of a,b and r is $0\sim g(x)$, $0\sim g(y)$ and $0\sim 0.5\sqrt{g^2(x)+g^2(y)}$ respectively. Discretization resolution of a,b and r of the accumulator array are Δa, Δb and Δr. Hence, the size of accumulator array is

$$\frac{g(x)g(y)\sqrt{g^2(x)+g^2(y)}}{2\Delta(a)\Delta(b)\Delta(r)} \tag{2}$$

The smaller is $\Delta(a)$, $\Delta(b)$ and $\Delta(r)$, the higher is detection accuracy. Correspondingly, huge storage is required to perform HT program.

The idea of hierarchical Hough transform is to divide an original image to a set of small images by a Gaussian sub-sampling process [13]. The factor which an image is segmented at each step is σ. That's, using a $\sigma\times\sigma$ window to reduce the size of an image. An original image was performed L times operation to produce a set of hierarchical images, illustrated as figure 1. The first HT operation makes use of the smallest image from the hierarchical images and accumulates the votes in a small accumulator array. Because of the use of a small image and a small accumulator array in the first HT operation, the values obtained for a,b and r, of course, are very rough estimation of the actual ones. The second HT operation is performed on a larger image and a larger accumulator array than those used before. However, the parameter ranges of investigation are narrowed down during the following operation. The estimated value of a,b and r obtained in the precedent operation can be used to reduce ranges of parameters to be investigated in the successive operation. Repeat the above mentioned procedures until the original image has been analyzed. The coarse-to-fine analysis strategy results in a significant reduction of computation time and storage compared to the standard HT [14].

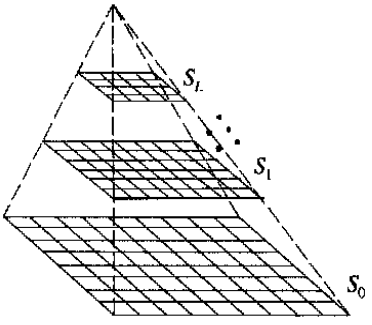


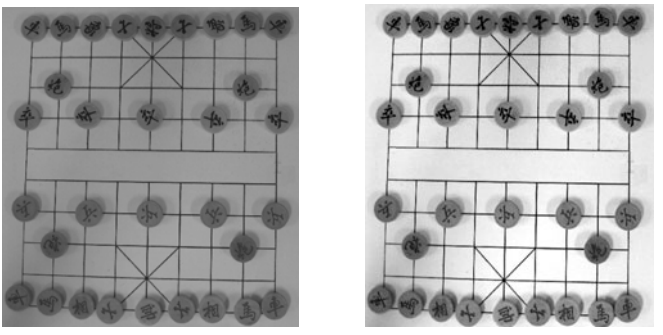
Fig. 1. Hierarchical subimages

Fig. 2. Prototype of a machine vision system

2.2 Lines and Circles Detection

The prototype of machine vision system for Chinese chess-playing robot was illustrated in Figure 2. The system consists of two color cameras mounted on the framework with a certain inclined angle, a chessboard with round pieces and corresponding programs for image processing and pattern recognition. In general, one camera is enough to compute any 3D coordinates if vertical distance between camera and chessboard is known. This can be done in laboratory. However, a real robot with machine vision system has ability to measure a changing distance between camera and objects. Taking this situation into account, a stereo vision system has been designed. Once the distance between camera and chessboard is determined and unchanged, only one camera works to reduce computation time and complexity.

The quality of an original image is essential to successive image processing procedures and pattern recognition. The color image grabbed by camera was converted into a grayscale image. In order to get a hign quality binary image, image enhanced and denoise operation are required. Figure 3 shows difference between original and enhanced grayscale image. Simple analysis shows there are only lines and round shapes in an image. Unfortunately, in some cases, round shapes maybe became ellipse-like one because of shadow and natural illumination. Before HHT was performed on an original image, a necessary procedure is edge detection and thresholding operations. Different operators were performed on a binary image and Laplacian operator was proven to be better. The edge image was dilated and eroded by morphological operations, see figure 4. Noticed that the distance between lines is the same and round shapes have the same radius value. Only a small image taken from center of the whole image was considered to be treated. In this small image, there includes lines and circles. HHT had been devoted to detect a few circles to determine the radius. This can reduce another parameter of circle detection process. There are only two parameters, x and y coordinates of circle's center point, are to be detected. This improvement also reduces computational and storage overhead. The results of HHT detection of lines and circles are illustrated in Figure 5. In order to see clearly, only a small part of image was shown.



(a) Original grayscale image (b) Enhanced grayscale image

Fig. 3. Original and enhanced grayscale image

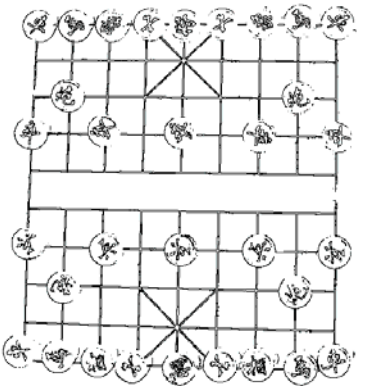


Fig. 4. Improved edge detection

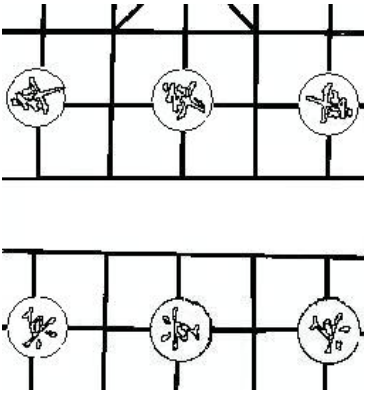


Fig. 5. Lines and circles detection using HHT

3 Recognition of Chinese Characters on Round Pieces

There are 32 Chinese characters which are divided into two groups with 16 characters each in Chinese chess-playing game. However, there exists only ten or eleven different shapes, see figure 6. Recognition of printed Chinese characters has been developed and put into practice with a relative higher accuracy. Chinese characters are stored in memory by means of dot matrix. Template matching approach is widely used in characters' recognition. The dot matrix of a standard printed Chinese character is stored in memory ahead of time. The vision system deals with the image and extracts the dot matrix of Chinese characters. The extracted dot matrix has been compared with predefined matrix. If two matrices are the same, then the Chinese character is recognized. Template matching is very easy to be understood and applied. But this approach is sensitive to direction of extracted image. Once this case happens, there need to rotate an extracted Chinese character to fit the standard one in memory. Of course, this is a time-consuming process. To eliminate bad effect of pieces' placing direction on recognition algorithm, a new approach based on statistical analysis of intersection points has been developed.

Fig. 6. Chinese characters on chessboard

Fig. 7. Concentric circles around a character

Assumed that there're a set of concentric circles around the center of a round piece, see Figure 7. Note that there exist no real circles in Figure 7. That's to say, these concentric circles are imaginary ones. From inner circle to outside circle, intersection points between each circle and the shape of a character have been counted and then a code was formed. In general, different characters have different standard codes, see table 1. Comparing extracted code of a Chinese character on pieces with all standard codes in memory, a corresponding Chinese character will be chosen. Of course, the statistical analysis approach is sensitive to image quality and different style of Chinese characters. To deal with this situation, a three-layer BP neural network was introduced to evaluate which code is best fit to the extracted code. The topology of neural network is shown in Figure 8. The input layer has three nodes and output layer has eleven nodes. The nodes of hidden layer will be decided by training sets. Sample plenty of data as possible as you can, and then group them into two sets. One is training set, the other one is testing set. Experiment shows this approach is better than simple code matching method. It's proven to be robust and reliable.

Table 1. Intersection points between character and virtual circles

	将	帅	帥	车	車	馬	马	炮	砲	卒	兵	相	象	士	仕
R1	5	3	4	3	6	8	4	4	5	5	5	5	6	2	3
R2	8	5	5	6	9	6	4	9	7	6	7	7	10	5	5
R3	2	2	1	1	2	1	0	3	1	0	0	2	1	1	0
Code	582	352	451	361	692	861	440	493	571	560	570	572	6A1	251	350

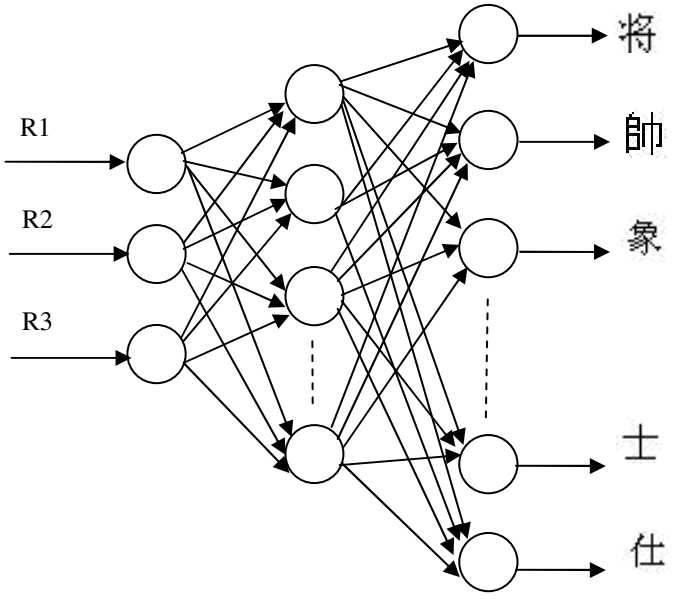


Fig. 8. Topology of three-layer back propagation neural network

Acknowledgments. This paper was fulfilled under the financial support of Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality. The project number is PHR201106149.

References

1. Lin, C.Y., Jo, P.C., Tseng, C.K.: Multi-functional intelligent robot DOC-2. Humanoid, 530–535 (2006)
2. Zhao, S., Chen, C., Liu, C., et al.: Algorithm of location of chess-robot system based on computer vision. In: Chinese Control and Decision Conference (CCDC 2008), pp. 5215–5218 (2008)
3. Du, J.-L., Zhang, J.-F., Huang, X.-H.: Chessboard recognition based on vision. Computer Engineering and Applications 43(34), 220–222 (2007)
4. Chau, C.-P., Siu, W.-C.: Generalized Hough transform using regions with homogeneous color. International Journal of Computer Vision 59(2), 183–199 (2004)
5. Akhtar, M.W., Atiquzzaman, M.: Determination of line length using Hough transform. Electronics Letters 28(1), 94–96 (1992)
6. Atiquzzaman, M., Akhtar, M.W.: Complete line segment description using the Hough transform. Image and Vision Computing 12(5), 267–273 (1994)
7. Chan, R., Siu, W.C.: New parallel Hough transform for circles. IEEE Proceedings Part-E 138(5), 335–344 (1991)
8. Kierkegaard, P.: A method for detection of circular arcs based on the Hough transform. Machine Vision and Applications 5, 249–263 (1992)
9. Nakanishi, M., Ogura, T.: Real-time CAM-based Hough transform algorithm and its performance evaluation. Machine Vision and Applications 12, 59–68 (2000)
10. Ji, Q., Xie, Y.: Randomised Hough transform with error propagation for line and circle detection. Pattern Analysis Application 6, 55–64 (2003)
11. Atiquzzaman, M.: Multi-resolution hough transform – an efficient method of detecting pattern in images. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(11), 1090–1095 (1992)
12. Illingworth, J., Kittler, J.: Adaptive Hough transform. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-9(5), 690–698 (1987)
13. Burt, P.J., Adelson, E.H.: The Laplacian pyramid as a compact image code. IEEE Transactions on Communications COM-31(4), 532–540 (1983)
14. Atiquzzaman, M.: Coarse-to-fine search technique to detect circles in images. International Journal of Advanced Manufacturing Technology 15, 96–102 (1999)

Utility-Maximizing Task Scheduling for Partially Observable Multiagent Systems

Qi-jin Ji, Zhe Yang, and Yan-qin Zhu

School of Computer Science and Technology, Soochow University, Suzhou, China
{Ji,yangzhe,qyzhu}@suda.edu.cn

Abstract. Task scheduling in a multiagent system makes decision on assigning which (and how many) tasks to which neighbour agents, therefore has immediate impact on the efficiency of resource utilization of the multiagent systems. In a large scale partially observable multiagent system, a globally optimal scheduling is more challenging because agents only have a partial view and limited information of the whole system. In this paper, we consider the problem of utility-maximizing task scheduling for a large scale partially observable multiagent system to achieve global optimum in a decentralized manner. We model the task scheduling as a nonlinear discrete optimization problem with coupled objective and constraints. We design a distributed primal-dual algorithm based on Lagrangian relaxation and solve the primal problem with a novel greedy algorithm and the dual algorithm with subgradient method respectively. We demonstrate the efficiency of our solution with a simple network but characterizing the essence of resource competition among agents and show that our algorithm can converge to the global optimum within limited steps of iteration.

Keywords: Task scheduling, Partially observable multiagent systems, Utility maximization.

1 Introduction

Task scheduling in a multiagent system (MAS) makes decision on assigning which (and how many) tasks at client agents to which neighbour agents as servers, therefore has immediate impact on the efficiency of resource utilization of the MASs. This impact is especially significant when agents are heterogeneous in both requirement and capacity. Previous work on task scheduling in a MAS often assumed that the world state is fully observable to the agents. Here we relax this assumption and examine the case where parts of the state are hidden to the agents.

There are several challenges to achieve the optimal goal in a partially observable MAS due to the partial observability and limited information of the whole system. In a *large scale* MAS, e.g., peer-to-peer file sharing networks [1] and distributed sensor networks [2], there might be hundreds to tens of thousand participating agents which locate widely. At this situation, it is very difficult to maintain and distribute global information in a centralized way, which excludes any centralized scheduling

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 387–394.
springerlink.com

algorithms. On the other hand, connections among agents are random and it is very possible that agents share some neighbours with others. Obviously, the shared agents will lead to resource competition and make scheduling decision of client agents interdependently. *There exists a contradiction between partial Observability of the client agents and global optimum of the system.*

In this paper, the scheduling satisfaction of client agents is captured by utility functions and the optimal task scheduling problem is formulated with a theoretical framework of network utility maximization [3]. We model the task scheduling as a nonlinear discrete optimization problem with coupled objective and constraints. We design a distributed primal-dual algorithm based on Lagrangian relaxation and solve the primal problem with a novel greedy algorithm and the dual algorithm with subgradient method respectively. Essentially, the Lagrangian multipliers intermediate the primal-dual interaction, which is an implicit coordinator among agents in a large scale systems. We demonstrate our solution of the problem with a simple network capturing the essential resource competition among agents and show that our algorithm can converge to the global optimum iteratively in a small number of rounds.

The rest of this paper is organized as follows: In Section 2, we present an overview of system model and formulate the utility maximization scheduling problem. Section 3 describes the design and analysis of distributed algorithm for the utility maximization problem in a large scale partially observable MAS. Numerical evaluation of the distributed algorithm is examined in Section 4. After the present of related work in Section 5, Section 6 concludes the paper.

2 Utility-Maximizing Scheduling Problem

Let us consider a partially observable MAS with *N* heterogeneous agents. The system is assumed static during the scheduling process, i.e., there is no agents joining or leaving from the system. Denote C = $\{C_i \mid i = 1,2,...,N\}$ as the set of client agents and S = $\{S_i \mid i = 1,2,...,N\}$ as the set of server agents. C_i and S_i are the *i*th agent in client agent set and server agent set respectively. The client agents periodically assign tasks to the server agents selected from neighbours and the server agents allocate their resource to the clients by accept or reject the tasks according to the resource availability. We assume a continuous backlog of tasks at each agent that needs to be assigned. Denote mc_i and DC_i as the minimum and maximum task number that can be assigned in a scheduling period of agent i and UC_j as the maximum task number that can be accepted in a scheduling period of agent *j*.

Let x_{ij} denote the assigned task number from C_i to S_i and $x_i = \sum_{j \in S} x_{ij}$ is total tasks assigned from C_i. In facts, agents can only send tasks to their neighbours. Denote p_i as the neighbour set of client C_i, and then we have $x_i = \sum_{j \in p_i} x_{ij}$. Correspondingly, let $y_j = \sum_{i \in C} x_{ij}$ represent the total tasks received from the other agents of agent *j*. A utility function U(.) is used to capture the satisfaction of a agent's

task assignment. The utility function is a strictly concave, increasing differentiable function of the agent's chunk task number. Assuming that the agents do not care which agents the tasks are sent to, the utility function of agent *i* can be represented as $U(x_i) = U(\sum_{j \in p_i} x_{ij})$.

Our objective is to make the task scheduling decision so as to maximize the aggregated social welfare subject to the agent capacity constraints and the problem is defined as follows:

$$\max_{x} \sum_{i=1}^{N} U(x_i) \tag{1}$$

subject to $$x_i = \sum_{j \in p_i} x_{ij}\ , \tag{2}$$

$$mc_i \le x_i \le DC_i\ , \tag{3}$$

$$y_j \le UC_j\ , \tag{4}$$

$$x_{ij} \in Z_+ \tag{5}$$

In the above model, Z_+ means the set of non-negative integer. The other denotations are consistent with the former definition. Inequality (3) and (4) represent the capacity limit of client and server agents respectively.

We denote (1)-(5) as problem (P). The difficulties of this problem are as follows: a) Coupled objective function and constraints; b) Requirement at each client agent to assign its tasks to more than one server agent; c) Optimization with discrete decision variables. Since the decision variables are integer, so no optimal conditions of the above problem can be derived from the Karush–Kuhn–Tucke theorem as its continuous counterpart [4, pp.342].

3 Distributed Algorithm

In this section, we will present a synchronous distributed algorithm for the problem (P). The distributed algorithm should be comprised of locally optimal algorithms and the communication mechanism among local processors. Inspired by the seminal work of Kelly [5] for social welfare maximization over a network, we design a distributed algorithm based on the Lagrangian relaxation of the primal problem. By the decomposition, the problem can be decoupled as client agent problem maximizing their utilities and server agent problem maximizing their profit. Since the decision variables of our schedule problem are all integer, and the task have to be assigned to more than one server, we design a novel greedy algorithm to solve the local task assignment problem at client agents. To cope with the non-differentiability of the dual problem, a subgradient method is employed to update the Lagranian multiplier at the server agents.

3.1 Decomposition of the Problem

To solve the problem (P) in a distributed way, we have to look at its dual to decouple the global objective and constraints. A dual problem can be obtained by relaxing the coupled constraint (4) with Lagrangian multiplier $\mu = \{\mu_i \mid i = 1,2,...,N\}$ as follows.

The Lagrangian duality of the model (Lagrangian relaxation of Integer Programming) can be obtained by relaxing the global constraint (4):

$$\begin{aligned} L(x,\mu) &= \sum_{i=1}^{N} U_i(\sum\nolimits_{j\in p_i} x_{ij}) + \sum_{j=1}^{N} \mu_j (UC_j - \sum\nolimits_{i=1}^{N} x_{ij}) \\ &= \sum_{i=1}^{N} (U_i(\sum_{j\in p_i} x_{ij}) - \sum_{j=1}^{N} \mu_j x_{ij}) + \sum_{j=1}^{N} \mu_j UC_j \\ &= \sum\nolimits_{i=1}^{N} (U_i(\sum_{j\in p_i} x_{ij}) - \sum_{j\in p_i} \mu_j x_{ij}) + \mu^T UC \end{aligned} \tag{6}$$

subject to constraints (2), (3) and (5).

And where $x = \{x_{11,}x_{12},...,x_{1N},x_{21,}x_{22},...,x_{2N},...,x_{N1,}x_{N2},...,x_{NN}\}$, $\mu = \{\mu_1, \mu_2,...,\mu_N\}$ and $UC = \{UC_1, UC_2,...,UC_N\}$.

3.2 Primal-Dual Algorithm

According (6), the coupled problem (P) is decomposed by Lagrangian relaxation and the objective of the relaxed problem can be accumulated by the objective of local decoupled problems. The subproblem at the *i*th client agent can be formulated as the following discrete optimization problem:

$$\max(U_i(\sum_{j\in p_i} x_{ij}) - \sum_{j\in p_i} \mu_j x_{ij}) \tag{7}$$

subject to constraints (2), (3) and (5).

We notice that the constraints (2), (3) and (5) all local constraints which can be satisfied with local information. μ_j is the Lagrangian multiplier, which can be obtained by solving the dual problem at server agent *j* and is considered as a constant for the local primal problem. The utility function U(.) is strictly increasing and it will increase with total tasks sent. The second item of the local objective represents the cost of the task. Since the total task number will be limited by DC_i at each client agent *i*, it is easy to understand that more low cost tasks and higher value of objective function. So a simple greedy algorithm can be designed to solve the local integer optimization problem, of which the key idea is to schedule as many as possible tasks to the server agent with lowest cost.

Algorithm 1: Greedy algorithm for maximizing the utility of local task scheduling

Input: p_i, μ_j, UC_j

Output: x_{ij}

$x_i \leftarrow 0;$

while ($x_i < DC_i \&\& \ p_i \neq \phi$)

Select the agent S_j in p_i with minimum μ_j;

$x_{ij} \leftarrow UC_j$; // as many as possible;

Scheduling x_{ij} task to the selected agent S_j;

$x_i \leftarrow x_i + x_{ij}$;

$p_i \leftarrow p_i - S_j$;

endwhile

Algorithm 1 optimally distributes the tasks of a client agent to its neighbours. The complexity of this algorithm is polynomial as $O(\| p_i \|^2)$, in which $\| p_i \|$ denotes the size of set p_i, i.e., the neighbour number of agent *i*.

Besides the available resource of the server agents, it is imperative to know the Lagrangian multiplier vector μ at each client agents to solve the primal problem. The multipliers can be obtained by solving the dual of the primal problem at server agents. According to the Lagrangian duality theory, the dual function $q(\mu)$ of the problem (P) is given as

$$q(\mu) = \sup_x (L(x, \mu) \tag{8}$$

And the dual problem (D) of (P) has the following definition.

$$\min \ q(\mu) \tag{9}$$

$$\text{subject to } \ \mu \in R^N, \mu > 0. \tag{10}$$

In an optimizing algorithm for the primal-dual problem, the dual problem can be solved by iteratively update the Lagrangian multipliers. Since x is discrete variables, the dual function $q(\mu)$ is usually nondifferentiable. The subgradient project method is proved to converge for searching the optimal value of nondifferentiable objective function [6]. It is used here to update the Lagrangian multipliers. The convergence property of subgradient method is determined by the step size of iteration along the subgradient direction towards the optimum value. Generally, the iteration step size should be small enough to ensure the computation converge to the optimum [7].

For each iteration, client agents individually solve (7) and communicate the results x_{ij} to server agents in its neighborhood. These server agents then update their multipliers according to the aggregated tasks received from all the clients and communicate the new multipliers to them, and the cycle repeats. A common economic explanation of the distributed algorithm is that the multiplier is equivalent to the price of the resource and it will increase when the demand increase and vice versa. The

customers will adjust their demand according to the price received. More information on this explanation can be found in [5].

4 Numerical Results

Fig. 1 illustrates a typical subsystem of a large scale MAS. There are four server agents S_0, S_1, S_2, S_3, S_4 and three client agents C_1, C_2 and C_3 in this subsystem. The arrows at the end of the edges show the serving relationship of two type agents. Both server agents and client agents are heterogeneous with different resource provision and demand respectively. To make the system works normally, the resource provision should not be less than the resource demand. In the subsystem client agent C_1, C_2 and C_3 actually compete the resource of the server agent S_0 but without any direct information about the competition of each other. The initial price of S_0 is set lower to illustrate the resource competition (If the price of S_0 is higher initially, no client agent will assign tasks to it according the algorithm 1). S_1, S_2, S_3 and S_4 serve only one client agent respectively. It is assumed that their load is high at the previous scheduling periods so their prices still keep high at the this time. Concretely, we set the initial price of the five server agent as 1.0, 4.0, 6.0, 8.0 and 10.0.

We set the resource provision (in terms of task number) of server agent S_i (i = 0, 1, 2, 3, 4) 10-2i and the resource demand of client agent C_i (i = 1, 2, 3) 14-2i (as labeled in Fig. 1). In this situation, the resource provision just equals to the resource demand. We use a constant step size for the subgradient method and set the step size as 0.1 and 0.05 respectively in the experiments. Fig. 2 shows that the aggregated utility evolving with the iterations. We can find in Fig. 2 that: 1) Our distributed algorithm can converge to global optimum within a limited number of iterations; 2) The converge rate is faster as the step size is larger.

While the aggregated utility implies the collective character of the system, we might be interested in the individual behaviour of the client and server agents. Fig. 3 presents that how the tasks from different client agents are directed to different server agents in the iterating process of the algorithm. From fig.3 it can be found that the tasks from the three clients are collected to S_0 at the beginning stage due to its lower price and then rerouted to other server agents according to their capacity constraints and load level with the evolution of iterations. After a few iterations, they will reach equilibrium at optimum. The dynamics of task distribution is driven by the resource price change controlled by the subgradient method.

5 Related Work

In recent years, there is much effort on task scheduling in a MAS, see for example, [8-10]. However, these works commonly assume that agents know the whole system state, which can not be used in many new applications such as large scale distributed sensor networks. Partial observability may has various consequences to the decision making of the agents. For instance, optimal planning under partial observability can be a hard problem even in the single-agent case [11]. An exact dynamic programming

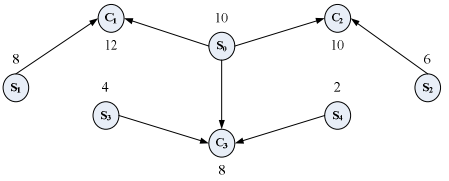


Fig. 1. A typical subsystem of a large scale partially observable multiagent system

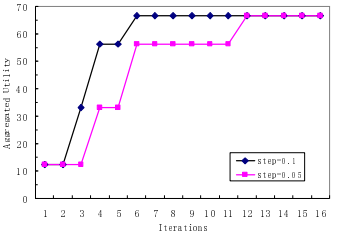


Fig. 2. Aggregated utility evolving with the iterations

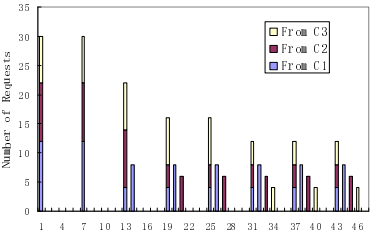


Fig. 3. Task number distribution evolving with the iterations

algorithm is developed for partially observable stochastic games. The algorithm is a synthesis of dynamic programming for partially observable Markov decision processes and iterated elimination of dominated strategies in normal form games [12].

Since the seminal work by Kelly on distributed network resource allocation using the language of network utility maximization (NUM) a decade ago [5], there have been extensive research efforts generalizing and applying NUM to model, analyze and design various network protocols and architectures. In the framework, network protocols are understood as distributed algorithms that maximize aggregate user utility under wired or wireless network resource constraints. This framework not only provides a powerful tool to reverse engineering existing protocols such as TCP [13], but also allows systematic design of new protocols. Please refer to [3] for a comprehensive review of NUM technique.

6 Conclusion

We have designed and evaluated an utility maximizing task scheduling method for a large scale partially observable MASs. This method is based on the primal dual

framework for distributed network optimization. In this distributed optimization framework, the resource competition is dealt with an implicit pricing mechanism. A novel optimal greedy local task scheduling algorithm is designed for the primal problem and the subgradient method is employed for solving the nondifferentiable dual problem. The reasonability and effectiveness of our design is evaluated numerically with a simple but essential subsystem topology.

We have considered the optimum and convergence of the distributed task allocation. However, we do not consider the fairness of our scheduling algorithm. We assume in our work that agents in the system are all cooperative. Incentive mechanism should be incorporated into new optimal scheduling method when agents are considered selfish.

Acknowledgment. This work was supported by NSF China Project under grant No. 61070170, Fundamental Research Project of Jiangsu Province under grant BK2009597, Open Research Fund from Key Laboratory of Computer Network and Information Integration in Southeast University, Ministry of Education, China, and Suzhou Key Laboratory of Converged Communication under Grant No.SZS0805.

References

1. Li, J.: On peer-to-peer (P2P) content delivery. Peer-to-Peer Netw. Appl. 1, 45–63 (2008)
2. Chong, C.Y., Kumar, S.P.: Sensor Networks: Evolution, Opportunities, and Challenges. Proc. IEEE 91(8), 1247–1256 (2003)
3. Chiang, M., Low, S.H., Calderbank, A.R., et al.: Layering as optimization decomposition: A mathematical theory of network architectures. Proc. of the IEEE 95(1), 255–312 (2007)
4. Li, D., Sun, X.: Nonlinear Integer Programming. Springer, Heidelberg (2006)
5. Kelly, F.P., Maulloo, A., Tan, D.: Rate control for communication networks: Shadow prices, proportional fairness and stability. Jour. Oper. Res. Soc. 49(3), 237–252 (1998)
6. Bertsekas, D.P.: Non-linear Programming, 2nd edn. Athena Scientific, Belmond (1999)
7. Bertsekas, D.P., Nedic, A., Ozdaglar, A.: Convex Analysis and Optimization. Athena Scientific, Belmond (2003)
8. Shehory, O., Kraus, S.: Methods for task allocation via agent coalition formation. Artificial Intelligence 101, 165–200 (1998)
9. Wellman, M.P., Walsh, W.E.: Auction Protocols for Decentralized Scheduling. Games and Economic Behavior 35, 271–303 (2001)
10. Dash, R.K., Vytelingum, P., Rogers, A., et al.: Market-Based Task Allocation Mechanisms for Limited-Capacity Suppliers. IEEE Transactions on Systems, Man and Cybernetics (Part A) 37(3), 391–405 (2007)
11. Papadimitriou, C.H., Tsitsiklis, J.N.: The complexity of Markov decision processes. Mathematics of Operations Research 12(3), 441–450 (1987)
12. Hansen, E.A., Bernstein, D.S., Zilberstein, S.: Dynamic programming for partially observable stochastic games. In: Proc. 19th National Conf. on Artificial Intelligence, San Jose, CA (2004)
13. Low, S.H., Lapsley, D.E.: Optimization flow control, I: Basic algorithm and convergence. IEEE/ACM Trans. Netw. 7(6), 861–874 (1999)

A Type-2 Fuzzy Subtractive Clustering Algorithm

Long Thanh Ngo and Binh Huy Pham

Department of Information Systems, Faculty of Information Technology,
Le Quy Don Technical University, No 100, Hoang Quoc Viet, Hanoi, Vietnam
{ngotlong,huybinhhvhc}@gmail.com

Abstract. The paper introduces a new approach to subtractive clustering algorithm (SC) with the fuzzifier parameter m which controls the clustering results in SC. And to manage the uncertainty of the parameter m, we have expanded the SC algorithm to interval type-2 fuzzy subtractive clustering algorithms (IT2-SC) using two fuzzifiers parameters m_1 and m_2 which creates a footprint of uncertainty (FOU) for the fuzzifier. The experiments are done based on image segmentation with the statistics show that the depends greatly on the parameter m of SC and stability and accuracy of our IT2-SC.

Keywords: Subtractive clustering, type-2 fuzzy subtractive clustering, type-2 fuzzy sets.

1 Introduction

Clustering is a common technique in many areas such as data mining, pattern recognition, image processing,.... Clustering is the division of data into clusters so that objects in a cluster have the biggest similarity. In particular, many clustering algorithms have been studied and applied, including k-mean [14], fuzzy c-mean [13] and mountain clustering [12]. However, in most real data exists uncertainty and vaguenesses which cannot be appropriately managed by type-1 fuzzy sets. Meanwhile, type-2 fuzzy sets allows us to obtain desirable results in designing and managing uncertainty. Therefore, type-2 fuzzy sets are studied and widely applied in many fields [7], [8], [9], especially pattern recognitions. On that basis, clustering algorithms has been extended and developed into type-2 fuzzy clustering algorithms to identify FOU of fuzzifiers, resulting to managing uncertainty is better.

Subtractive clustering algorithm, proposed in 1993 by Chiu [1],[2], is an extension of the mountain clustering methods by improving the mountain function to calculate potential of becoming a cluster center for each data point based on the location of the data point with all the other data points. The subtractive clustering algorithm only consider data points, not a grid points, which reduces the computational complexities and gives better distribution of cluster centers in comparison with the Mountain clustering algorithm and other algorithms. By 2005, Kim et al. have improved subtractive clustering algorithm by proposing

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 395–402.
springerlink.com

a kernel-induced distance instead of the conventional distance when calculating the mountain value of data point [6]. In 2008, J. Chen et al proposed a weighted mean subtractive clustering [5].

In subtractive clustering algorithms,except uncertainty have been created in the data, setting subtractive clustering's parameters are very influential to the results of clustering [3],[4]. So in this paper, we extend the subtractive clustering algorithm by putting a fuzzifier parameter m into mountain function to caculate for data points. The fuzzifier parameter m can alter the results of the mountain function so it has great influence to the results of clustering. Through fuzzifier parameter m, we can reduce the dependence of clustering results in initial values for the parameters of the algorithm. As with the adjustment parameter m, we can obtain better clustering results without interesting into setting the initial values of the parameters of the algorithm. Therefore, fuzzifier parameter m created uncertainties for the SC. To design and manage uncertainties of fuzzifier m , we extend to interval type-2 fuzzy subtractive clustering algorithm (IT2-SC) by using two fuzzifiers m_1 and m_2 which creates a footprint of uncertainty (FOU) for the fuzzifier m. Then, comparative experiments between IT2-SC and other subtractive clustering to showing the validity of our proposed method.

The remainder of this paper is organized as follows. In Section 2.1 introduces briefly type-2 fuzzy set and interval type-2 fuzzy set. Subtractive clustering is presented in section 2.2. In section 3, we discuss how to extend subtractive clustering method with fuzzifier parameter m and propose interval type-2 fuzzy subtractive clustering algorithm. In section 4, we provide several experiments showing the validity of our proposed method. Finally, Section 5 gives the summary and conclusions.

2 Preliminaries

2.1 Type-2 Fuzzy Sets

A type-2 fuzzy set in X is denoted $\tilde{A}$, and its membership grade of $x \in X$ is $\mu_{\tilde{A}}(x,u), u \in J_x \subseteq [0,1]$, which is a type-1 fuzzy set in [0, 1]. The elements of domain of $\mu_{\tilde{A}}(x,u)$ are called primary memberships of x in $\tilde{A}$ and memberships of primary memberships in $\mu_{\tilde{A}}(x,u)$ are called secondary memberships of x in $\tilde{A}$.

Definition 1. *A type $-$ 2 fuzzy set, denoted $\tilde{A}$, is characterized by a type-2 membership function $\mu_{\tilde{A}}(x,u)$ where $x \in X$ and $u \in J_x \subseteq [0,1]$, i.e.,*

$$\tilde{A} = \{((x,u), \mu_{\tilde{A}}(x,u)) | \forall x \in X, \forall u \in J_x \subseteq [0,1]\} \quad (1)$$

in which $0 \leq \mu_{\tilde{A}}(x,u) \leq 1$.

At each value of x, say $x = x'$, the 2-D plane whose axes are u and $\mu_{\tilde{A}}(x',u)$ is called a *vertical slice* of $\mu_{\tilde{A}}(x,u)$. A *secondary membership function* is a vertical slice of $\mu_{\tilde{A}}(x,u)$. It is $\mu_{\tilde{A}}(x = x',u)$ for $x \in X$ and $\forall u \in J_{x'} \subseteq [0,1]$, i.e.

$$\mu_{\tilde{A}}(x = x', u) \equiv \mu_{\tilde{A}}(x') = \int_{u \in J_{x'}} f_{x'}(u)/u, J_{x'} \subseteq [0, 1] \tag{2}$$

in which $0 \leq f_{x'}(u) \leq 1$.

Type-2 fuzzy sets are called an interval type-2 fuzzy sets if the secondary membership function $f_{x'}(u) = 1 \ \forall u \in J_x$ i.e. a type-2 fuzzy set are defined as follows:

Definition 2. *An interval type-2 fuzzy set $\tilde{A}$ is characterized by an interval type-2 membership function $\mu_{\tilde{A}}(x, u) = 1$ where $x \in X$ and $u \in J_x \subseteq [0, 1]$, i.e.,*

$$\tilde{A} = \{((x, u), 1) | \forall x \in X, \forall u \in J_x \subseteq [0, 1]\} \tag{3}$$

Uncertainty of $\tilde{A}$, denoted FOU, is union of primary functions i.e. $FOU(\tilde{A}) = \bigcup_{x \in X} J_x$. Upper/lower bounds of membership function (UMF/LMF), denoted $\overline{\mu}_{\tilde{A}}(x)$ and $\underline{\mu}_{\tilde{A}}(x)$, of $\tilde{A}$ are two type-1 membership function and bounds of FOU.

2.2 Subtractive Clustering Algorithm

Subtractive clustering finds the optimal data point to define a cluster center based on the density of surrounding data points, see [1,2] for more details. Consider a collection of n data points: $X = \{x_1, x_2, ..., x_n\}$, where, x_i is a vector in an M-dimensional space. The subtractive clustering algorithm includes the following steps:

Step 1: Initialization, r_a, η with $\eta = \frac{r_b}{r_a}$, $\overline{\varepsilon}$ and $\underline{\varepsilon}$

Step 2: Caculating density for all data points by using formula bellow:

$$P_i = \sum_{j=1}^{n} e^{-\frac{4}{r_a^2}\|x_i - x_j\|^2} \tag{4}$$

Where P_i denotes the density of ith data point, r_a is a positive constant defining a neighborhood radius and $\|.\|$ denotes the Euclidean distance.

Data point with the highest density is selected as the first cluster center.

Step 3: The density of all data points is revised by using formula bellow:

$$P_i = P_i - P_k^* e^{-\frac{4}{r_b^2}\|x_i - x_k^*\|^2}; i = 1, ..., n \tag{5}$$

Where r_b is is a positive constant and $r_b = \eta * r_a$ with a good choise is $\eta = 1.5$.

Step 4: Let x^* is a data point with its density is highest and equal P^*.

-If $P^* > \overline{\varepsilon} P^{ref}$: x^* is a new cluster center and back to Step 3.

- Else if $P^* < \underline{\varepsilon} P^{ref}$: back to Step 5.

- Else:

+ Let d_{min} is shortest of the distances between x^* and all previously found cluster centers.

+ If $\frac{d_{\min}}{r_a} + \frac{P^*}{P^{\text{ref}}} \geq 1$: x^* is a new cluster center and back to Step 3.

+ Else: $P(x^*) = 0$ and select x^* with the next highest density, $P(x^*)$, v back to step 4.

Step 5: Output the results of clustering.
When the membership degree of data point in each cluster is as follows:

$$\mu_{ik} = e^{-\frac{4}{r_a^2}\|x_i - x_k\|^2} \tag{6}$$

3 Interval Type-2 Fuzzy Subtractive Clustering

3.1 Extending Subtractive Clustering Algorithm

In the subtractive clustering algorithm, we must set four parameters: acceptance ratio $\overline{\varepsilon}$, reflection ratio $\underline{\varepsilon}$, cluster radius r_a and squash factor η (or r_b). The choice of parameters have greatly influences to results of clustering. If values of $\overline{\varepsilon}$ and $\underline{\varepsilon}$ are large, the number of cluster centers will be reduced. The influences of the four parameters to clusters have been described detail in papers of Demirli [3],[4]. Therefore, these parameters are uncertainties in the subtractive clustering algorithm.

On the other hand, subtractive clustering estimated the potential of a data point as a cluster center based on the density of surrounding data points, which is actually based on the distance between the data point with the remaining data points. Therefore, SC includes various types of uncertainty as distance measure, parameters Initialization...So we consider a fuzzifier parameter that control the distribution of data points into clusters by making the parameter m in the density function to calculate the potential of a data point as follows:

$$P_i = \sum_{j=1}^{n} e^{-\frac{4}{r_a^2}(x_j - x_i)^{\frac{2}{m-1}}} \tag{7}$$

If x_k is the k^{th} cluster position, has potential P_k^*, then the potential of each data point is revised by the following formula:

$$P_i = P_i - P_k^* e^{-\frac{4}{r_b^2}(x_i - x_k)^{\frac{2}{m-1}}}; i = 1, ..., n \tag{8}$$

Then the choice of the value of the parameter m has greatly influence to results of clustering. If m is small, the number of cluster centers will be reduced. Conversely, If m is too large, too many cluster centers will be generated. In addtion, through the adjustment of fuzzifier parameters m, it is easy to obtain the good results of clustering that is not dependent on the setting of the parameters for SC.

Fig.1 illustrates the influence of fuzzifier parameters m to initial parameters of SC. In Fig.1(a) represents the best results of clustering with the value of the initial parameters are $\overline{\varepsilon} = 0.5$, $\underline{\varepsilon} = 0.15$, $r_a = 0.25$, $\eta = 1.5$, respectively (Chiu [1,2]). In Fig.1(b) describes the results of clustering of expansion SC algorithm with the value of the initial parameters are $\overline{\varepsilon} = 0.5$, $\underline{\varepsilon} = 0.15$, $r_a = 0.4$, $\eta = 1.35$ and $m = 2.47$, respectively. We see that the result of clustering in the two cases are quite similar.

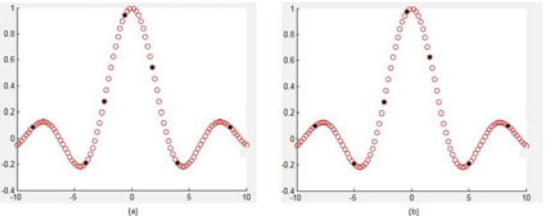


Fig. 1. The influence of fuzzifier parameters m to initial parameters of SC

3.2 Interval Type-2 Fuzzy Subtractive Clustering Algorithm

In expansion subtractive clustering algorithm, the degree of membership of a data point in k^{th} cluster center is defined as following formula:

$$\mu_{ik} = e^{-\frac{4}{r_a^2}(x_i - x_k)^{\frac{2}{m-1}}} \tag{9}$$

Where x_k is k^{th} cluster center According to formula 9, the value of membership of a data point in k^{th} cluster center depends on the position of k^{th} cluster and the fuzzifier parameter m. On the other hand, the position of the k^{th} cluster also depends on the fuzzifier parameter m.Thus, the fuzzifier parameter m is the most uncertainty element in expansion subtractive clustering algorithm. Therefore, to design and manage the uncertainty for fuzzifier parameter m , we extend a pattern set to interval type-2 fuzzy sets using two fuzzifiers m_1 and m_2 which creates a footprint of uncertainty (FOU) for the fuzzifier parameter m. Then the degree of membership of k^{th} cluster center is defined as following formula:

$$\begin{cases} \overline{\mu}_{ik} = e^{-\frac{4}{r_a^2}(x_i - x_k)^{\frac{2}{m_1-1}}} \\ \underline{\mu}_{ik} = e^{-\frac{4}{r_a^2}(x_i - x_k)^{\frac{2}{m_2-1}}} \end{cases} \tag{10}$$

Thence, we have two density functions to caculate potential of each data point as bellows:

$$\begin{cases} \overline{P}_i = \sum_{j=1}^{n} e^{-\frac{4}{r_a^2}(x_j - x_i)^{\frac{2}{m_1-1}}} \\ \underline{P}_i = \sum_{j=1}^{n} e^{-\frac{4}{r_a^2}(x_j - x_i)^{\frac{2}{m_2-1}}} \end{cases} \tag{11}$$

If the centroids are identified by the formula (11), we will have centroids v_L and v_R. Thus, we will do type reduction for centroids as bellows:

$$P_i = \frac{\overline{P}_i * m_1 + \underline{P}_i * m_2}{m_1 + m_2} \tag{12}$$

And when we identified k^{th} cluster center, the density of all data points is revised by using following formula:

$$\begin{cases} \underline{P}_i^{sub} = P_k^* \sum_{j=1}^{n} e^{-\frac{4}{r_b^2} d_{ij}^{\frac{2}{m_1-1}}} \\ \overline{P}_i^{sub} = P_k^* \sum_{j=1}^{n} e^{-\frac{4}{r_b^2} d_{ij}^{\frac{2}{m_2-1}}} \\ P_i^{sub} = \frac{\underline{P}_i^{sub} * m_1 + \overline{P}_i^{sub} * m_2}{m_1 + m_2} \\ P_i = P_i - P_i^{sub} \end{cases} \tag{13}$$

The interval type-2 fuzzy subtractive clustering algorithm includes the following steps:

Step 1: Initialization, r_a, η with $\eta = \frac{r_b}{r_a}$, $\overline{\varepsilon}$ and $\underline{\varepsilon}$, m_1 and m_2 $(1 < m_1 < m_2)$

Step 2: Calculating density for all data points with two fuzzifiers m_1 and m_2 by using formulas (11) and (12). Data point with the highest density is selected as the first cluster center: $P_k^* = \max_{i=1}^{n} P_i$ where $k = 1$ and P_k^* is the density of the first cluster center.

Step 3: The density of all data points is revised by using formula (13).

Step 4: The identification of the next cluster centers are as similar as SC.

Step 5: Output the results of clustering.

4 Application to Image Segmentation

In this section, we compare the results between subtractive clustering algorithm and interval type-2 fuzzy subtractive clustering algorithm through the tests of image segmentation. Here, multiple levels of image segmentation is a combination between FCM clustering algorithm and subtractive clustering algorithms. In particular, we use the result of SC or IT2-SC to initialize cluster centers for FCM algorithm, then through FCM given the results of simage segment and iterations of the FCM. Finally, we compare the validity between SC and our proposed IT2-SC on based the results of image segmentation and iterations of the FCM. The steps are as follows:

Step 1: Using subtractive clustering algorithm (in section 3.1) or interval type-2 fuzzy subtractive clustering algorithm (in section 3.2) initializes for the initial cluster centers matrix $V^{(0)} = [v_{ij}], V^{(0)} \in R^{d \times c}$.

Step 2: Using FCM clustering with cluster center matrix is initialized in step 1 and gives results of clustering $V^{(j)}$ and iterations of FCM j.

Step 3: Comparison of iterations to be performed and the results of clustering to rate the effectiveness of subtractive clustering algorithms.

As the result of experiments, IT2-SC makes clustering results more efficiently than SC when the results of its clustering applied into FCM to converge quickly to clustering results of FCM with better clusters and the number of iterations to perform also better than the SC.

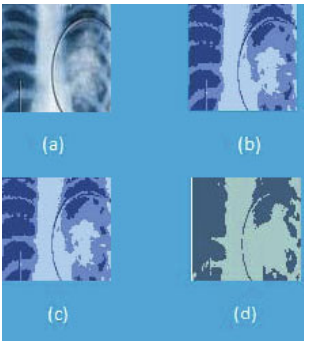


Fig. 2. Segmented results for chest trauma image for SC and IT2-SC

In this experiment, we conduct image segmentation for image of chest trauma area. In Fig.2(a), we only consider to contain two regions, namely, bone and damaged domain.

Tests was conducted with the original initialization parameters, respectively, $\overline{\varepsilon} = 0.5, \underline{\varepsilon} = 0.15, r_a = 0.25$ and $\eta = 1.25$. Fig.2(b) shows the image segmentation result of SC and FCM ($m = 2$) and Fig.2(c) shows the image segmentation result of IT2-SC and FCM with $m_1 = 1.65$ and $m_2 = 2.35$. In both of these test results, image is divided into three quite distinct regions. Fig.2(d) shows the best result of image segmentation with two distinct regions in which we easily observe the damaged area.

The statistical results of the comparison between two algorithms in Tab. 4 shows that the IT2-SC algorithm gives better results than the SC algorithm. The results show that FCM terminals in 38 iterations with FSC (m=2) whereas FCM terminals in 22 or 20 iterations with IT2-SC.

Table 1. Comparison between FSC and IT2-FSC

Algorithm	$\overline{\varepsilon} = 0.5, \underline{\varepsilon} = 0.15, r_a = 0.5, \eta = 1.25$		
	Fuzzifier parameter	Iterations of FCM	Number of clusters
FSC	m=2	38	3
IT2-FSC	$m_1 = 1.65$ $m_2 = 2.35$	22	3
	$m_1 = 1.42$ $m_2 = 2.58$	20	2

5 Conclusion

The article presents a new approach for subtractive clustering algorithm with fuzzifier parameter m which has a great influence on the clustering process. Through fuzzifier parameter m, we can reduce the uncertainty of the algorithm

in the initialization for parameters. Since then, we have expanded into interval type-2 fuzzy subtractive clustering algorithm by using two different values of fuzzifier parameter m. Through experiments on image segmentation have shown the effectiveness of our proposed IT2-SC.

For the future works, we will apply our proposed IT2-SC method to model type-2 and interval type-2 TSK fuzzy logic systems by the construction of the type-2 and interval type-2 fuzzy rules.

Acknowledgments. This paper is sponsored by Vietnam's National Foundation for Science and Technology Development (NAFOSTED).

References

1. Chiu, S.L.: Fuzzy Model Identification Based on Cluster Estimation. Journal on Intelligent Fuzzy Systems 2, 267–278 (1994)
2. Chiu, S.L.: Extracting Fuzzy Rules from Data for Function Approximation and Pattern Classification. In: Dubois, D., Prade, H., Yager, R.R. (eds.) Fuzzy Information Engineering: a Guide Tour of Applications, pp. 149–162. Wiley (1997)
3. Demirli, K., Cheng, S.X., Muthukumaran, P.: Subtractive Clustering Based Modeling of Job Sequencing with Parametric Search. Fuzzy Sets and Systems 137, 235–270 (2003)
4. Demirli, K., Muthukumaran, P.: Higher Order Fuzzy System identification Using Subtractive Clustering. J. of Intelligent and Fuzzy Systems 9, 129–158 (2000)
5. Chen, J.Y., Qin, Z., Jia, J.: A Weighted Mean Subtractive Clustering Algorithm. Information Technology Journal 7(2), 356–360 (2008)
6. Kim, D.W., Lee, K.Y., Lee, K.H.: A Kernel Based Subtractive Clustering Method. Pattern Recognition Letter 26(7), 879–891 (2005)
7. Mendel, J.M., John, R.I., Liu, F.: Interval Type-2 Fuzzy Logic Systems Made Simple. IEEE Trans. on Fuzzy Systems 14(6), 808–821 (2006)
8. Karnik, N., Mendel, J.M.: Operations on Type-2 Fuzzy Sets. Fuzzy Sets and Systems 122, 327–348 (2001)
9. Liang, Q., Mendel, J.M.: Interval Type-2 Fuzzy Logic Systems: Theory and Design. IEEE Trans. on Fuzzy Systems 8(5), 535–550 (2000)
10. Hwang, C., Rhee, F.C.H.: Uncertain Fuzzy clustering: Interval Type-2 Fuzzy Approach to C-means. IEEE Trans. on Fuzzy Systems 15, 107–120 (2007)
11. Zarandi, M.H.F., Zarinbal, M., Trksen, I.B.: Type-II Fuzzy Possibilistic C-Mean Clustering. In: IFSA/EUSFLAT Conf., pp. 30–35 (2009)
12. Yager, R.R., Filev, D.P.: Approximate clustering via the mountain method. IEEE Trans. on Systems, Man and Cybernetics 24(8), 1279–1284 (1994)
13. Bezdek, J.: Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic Publishers (1981)
14. Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y.: An Efficient k-Means Clustering Algorithm: Analysis and Implementation. IEEE Trans. Pattern Anal. Mach. Intell., 881–892 (2002)

Research on Experimental Teaching Model Innovation from the Constructionist Perspective*

Chao Chen

Department of College Computer, Bohai University,
Jinzhou 121013, Liaoning Province, China

Abstract. With the development of educational reform in Universities , basic computer laboratory teaching begins to show its drawback gradually. According to many years' teaching experience, This paper gives an analysis and research on the reform and practice of experiment teaching in basic computer courses. Based on the adoption of advanced information, particularly network based technology, we intend to strengthen student's application and arouse students' interest in learning and stimulate their innovation ability. The paper expounds the issues related to the construction of such an experimental teaching model from the constructionist perspective.

Keywords: Experimental Teaching Model, Constructivism, Innovation.

1 Introduction

Adhere to people-oriented, comprehensive quality education is the educational reform and development of the strategic themes' is pointed out in *Essentials of The National Medium and Long Term Educational Reform and Development Plan* (2010-2020), which "Focuses on investing in students' courage to explore innovation and the ability of solving problems in practice. Experimental teaching is a critical aspect of improving students' practical ability and creative ability. University computer Foundation is a public elementary courses for non - computer majors, and the basic objective of this courses is cultivating students who will have some basic knowledge of computer, will acquire the relevant hardware and software technology and will have the abilities in solving the problem in the field of professional by using computers.

Compared with other teaching links, the experimental teaching of basic computer education in university plays an irreplaceable role in training student's comprehensive quality and capacity.

From the practice of experiment teaching level of students provides validated, comprehensive and innovative environment. On the basis of known knowledge and problem verified, students can gain hands-on experience and skills that are required for innovation. Comprehensive information capacity and ability of talents should be

* This paper is subsidized by the project of the 11th five-year planning on educational science from Liaoning province, Project number: JG09DB029.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125,, pp. 403–409.
springerlink.com

developed as a starting point and end point in the perspective of the improvement of teaching quality and reform of experimental teaching of University computer Foundation.

2 The Significance of the Reformation on Experimental Teaching

2.1 The Teaching Difficulties Caused by the Different Levels of the New Students

With the development of the computerized information project in primary schools and middle schools, the course of computer in universities no longer starts from zero. But due to area differences and the varied attitudes towards computer information technology, the first-year students' qualities on basic computer skills and knowledge are quite different. And thus the teaching in this area is faced with many problems and troubles, for instance, the teaching pace is difficult to set, and we can hardly choose appropriate methods to deal with different kinds of students in the same class, let along different ways for different students and questions of the same kind.

2.2 The Impracticalness of the Teaching Contents and the Lack of the Teaching Resources

Traditionally, the computer experimental teaching is attached to the theory teaching. And we usually do the experiments to check the knowledge we have acquired. So our experiments are usually typical and our purpose is to check what we have learned. For that reason, creative and comprehensive experiments are quite rear. As we all know, computer science develops and renews quickly and new techs, new applying systems and ways come into being every day. So we should try to obtain late teaching resources and materials to keep up with the scientific development.

2.3 Backwardness of the Teaching Methods

Traditionally, students validate the content on the textbook in accordance with the experimental instruction on basic computer experiment course which is teacher-dominated, focusing on study results and neglecting the learning process. This kind of teaching model results in the missing of students' autonomous learning motivation and the lack of the spirit of exploration. In the meanwhile, college freshmen who have been used to exam-orientated education still take the obsolete learning method and habitual thinking mode to solve new problems.

2.4 Limited Time, Limited Place

The time for every experiment is limited. Facing much knowledge and a great deal of information the students don't know how to do and what to do. There are so many students in one class that it is impossible for the teacher to guide every one. So the effect and the schedule of the experiment are both affected. There tend to be the fact that the experiment is over as soon as the students just get it. Of course there are little extracurricular experiments. Further, the experiment can only be finished in the specified laboratory.

3 The Theoretical Premise of the Experimental Teaching Reform—Constructivism Theory

In recent years, constructivism, the mainstream idea in education field, is more and more widely applied in network teaching.

Constructivism is a learning theory found in psychology which explains how people might acquire knowledge and learn. It therefore has direct application to education. The theory suggests that humans construct knowledge and meaning from their experiences. Constructivism is not a specific pedagogy. Piaget's theory of Constructivist learning has had wide ranging impact on learning theories and teaching methods in education and is an underlying theme of many education reform movements. Research support for constructivist teaching techniques has been mixed, with some research supporting these techniques and other research contradicting those results.

4 The Technical Support of the Experimental Teaching Reform—an Open Platform base on Network

The development of the web-based, comprehensive-function, highly-effective and easy-to-use learning support platform, not only helps the nurturing of the students' self-learning ability, collaboration ability and innovative spirit, but also improves the quality of education, promotes the reform of education, and uplifts the level of education information. It is of crucial significance to the effective development of information education work in universities.

4.1 Teaching Resource

Comprehensive experiment teaching platform collects abundant resources related to the course and provides plentiful experiment items for every module. We develop and use the curriculum Website and "online test and exercise system ", and also make full use of public school teaching platform and FTP servers. The network supplies a wealth of teaching resources to build teaching environment with strongly support. High use of the platform as "document repository" substituting the cheapest way as the written materials should have. [1]The students would like to have all materials inside the platform instead the teachers' view of use the platform as an "updating" way to changes and addendum of the materials.

4.2 Teaching Management

On the network platform we can select the student or teacher login mode. As for student interface, students can view the announcement, class schedules information, scores information, attendance information, and fill experimental exercises and report. As for teacher interface, teachers with different authorities access different display pages. Generally, teachers have all the rights of students, but also handle the homework and experimental reports management, reports marking, student score and attendance management.

In general, the experimental teaching management system should consists of experiment reservation module, login module, announcement module, job submission module, experimental-report management module, curriculum schedule design module, attendance module, report marking and score statistics module, device management module, and experimental and laboratory arrangement module, personal information modification and role assignment module.[2]

Teachers can know students' learning process at any time and find out the problems students have in time, interact with students by different ways, such as forum, e-mail and QQ, to guide and supervise their learning. In addition, the core content is based on web lectures, tutorials, assignments and job correcting, guiding students to use the Internet to collect and process information.

5 Designing and Planning for the Experimental Teaching

Multimedia and network interactivity are leading to new forms of teaching and learning and new roles for students who act as participants and not only spectators of their own learning process. On account of the existing problems in traditional experiment course, we conduct an exploring practice in experiment teaching of basic computer courses and build a comprehensive experiment teaching platform based on the network which shown in Fig.1.

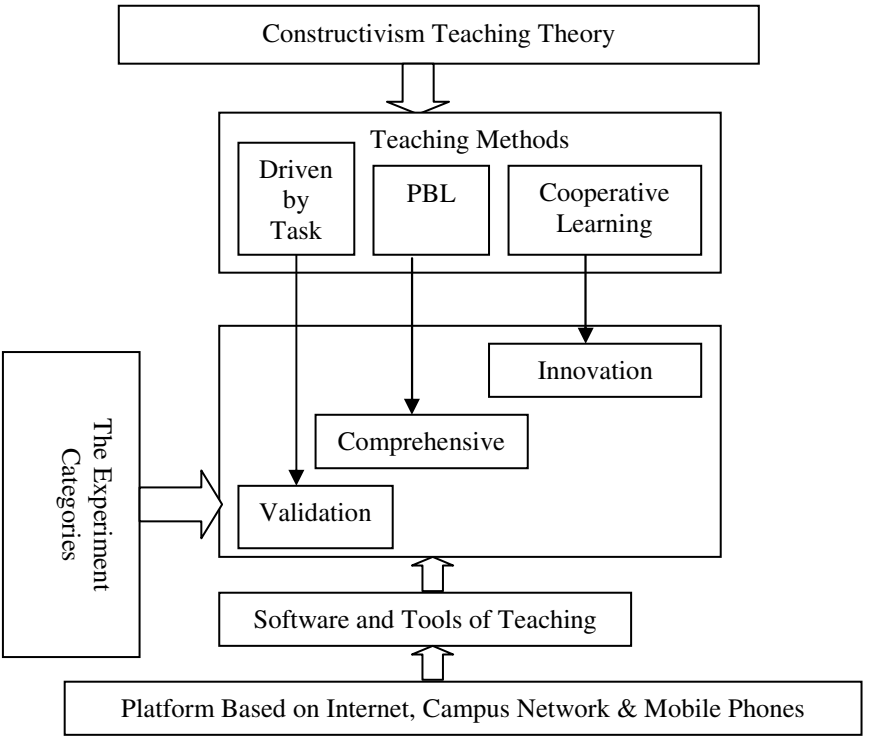


Fig. 1. The reforming scheme figure of the experiment teaching

5.1 Optimize the Content of Experiments

Importance of curriculum required teaching content must take account of foundation, application, and forward-looking and reflect the fun at the same time in order to stimulate student's interest in learning. At first, in order to stimulate students' enthusiasm we compile experiment guidebook whose content is closely related to the areas of expertise.

Then we modified the experiment instruction for each module. Around the requirements of students' capacity-building, the experiments are classified into three categories: validation, innovation and comprehensive. The validation experiments focus on basic knowledge and skills of the computer. The comprehensive experiment is a kind of Composite experiment which is mixed up with a variety of skills and method of each module. And the innovative experiment is that the title, the purpose, the procedures, the theory for the designing, and the results of the pre experiments are decided the groups themselves. All of the above are decided by the group members after their extensive reading, material finding, and selecting, reasoning and their tutors direction and instruction. Some examples are shown in table 1.

Table 1. Some typical examples of different experiment

	Cognition and Validation Experiment	Comprehensive Experiment	Design and innovation Experiment
Word	Document entry Content editing Format a document Common typographic skills	The usage of templates and styles; Mixed Graphic and Text; page setup;	Personal task using various Word skills for professional applications
Excel	editing and basics of worksheets Process data with formulas	Make charts with the datasheet Manage and use the data	Complete a file which is for some practical application with most Excel skills
PowerPoint	Build presentations; Format and beautify the PPT	Insert media, set animation and establish hyperlinks Insert and set a sound file Show slides	Personal task using all kinds of PowerPoint skills
Common Instrumental Software	The usage of compress-decompress software The usage of windows recorder	The usage of Windows Media Player; The usage of Realone Player	Complete one DV works with all kinds of multimedia software

5.2 Optimize the Method of Experiments Teaching

1. For validating experiments

Validating experiments are about basics and basic skills, the task-driven teaching is an effective teaching model. The task-driven teaching model is each one task which designs carefully through the teacher, lets the student accept the task and complete the task at the same time, masters the knowledge, the skill and the method. In order to "foster basis", each student must complete every experiment projects.

2.For comprehensive experiments

Cooperative learning is basically an instructional strategy where students form small groups or teams and work together to maximize their own and each other's learning. In this kind of model, both the instructor and students jointly construct the knowledge. Students are active learners and discoverers of own knowledge. Students are assumed to be intelligent to do construct and transform knowledge. As an instructor, he or she is responsible for developing students' competencies and talents. Due to the close interaction, relationships among students are more personal, and so is that between faculty and students.

Under the auspices of cooperative learning, students help each other learn and work in teams. Students depend on each other for goal satisfaction, rewards, resources, division of labor, roles, and so on. In cooperative process, problem characteristics are defined, responsibilities are assigned, and steps are taken to solve the problem cooperatively.

3. For innovative experiments

We encourage students to design and develop integrated experiment projects with PBL (Problem-led Learning). Problem-led learning can best be defined within the constructivist tradition of education theory that advocates learning as a process in which knowledge is actively constructed by learners interacting with their environment and with others. Educational models that reflect such a philosophy typically emphasize a learner-centered and integrative approach to teaching and learning. Different names such as enquiry-based learning, problem-based learning, enquiry and action learning, and experience-based learning, etc. have been coined for pedagogical approaches that adopt this orientation towards learning and teaching.

An engaging problem for small groups/teams to solve that is reflective of what they are/will be facing in their researches and jobs; it should be challenging, yet not difficult that it is beyond the group. It has a reasonable level of complexity so it speaks to a range of issues that must be solved in order to address the central problem. It reflects the basic computer knowledge of the learners and encourages them to contribute their expertise.

5.3 Optimize the Evaluation Methods

We emphasize the process evaluation and form a scientific and reasonable evaluation method for experiment.

1) Daily Performance Scores: Based on the platform, the amount and quality of working for students in their course can be automatically and objectively recorded in real time. The daily performance, problems solved correctly, frequency of

participation and credits combined and changed in real time to fully reflect students' dynamic learning process.[3]

2) Self-evaluation questions and answers: These are the best-used pages of the platform because students need to face objective tests that demonstrate the real degree to which they know the subject.

3) Final examinations: The final exam is no longer the primary means of examination. Because of their excellent performance, some students can get free examination.

6 Conclusion

Colleges and universities have set up basic computer-related courses to develop a comprehensive need of social development professionals. This Paper analyses the problems existing in computer experiment lesson in colleges and universities, discusses the reform of the courses and outline the model we have created and the methods we have implemented and validated. We must firmly focus on the training of applied ability to make a comprehensive, coordinated reform on the traditional teaching contents, methods, and means.

References

1. Webber, C.G., Lima, M.F.W.P., Casa, M.E., Ribeiro, A.M.: Towards Secure e-Learning Applications: a Multiagent Platform. Journal of Software 2(1), 60–69 (2007)
2. Zhang, Y., Yan, H.: Reform and Practice of Programming Teaching Based on Applied Ability Building. In: 2010 10th IEEE International Conference on Computer and Information Technology (2010)
3. Wang, A.-D., Guo, X.-H., et al.: Developing Core Competence of Computer major-A Tentative Reform on Experiment Teaching of Programming Courses. In: 2010 Third International Conference on Education Technology and Training (ETT), pp. 31–34 (2010)

Application of Coanda Effect in Robots – A Review

Elango Natarajan* and Nneka Obianuju Onubogu

Faculty of Mechanical Engineering,
Kolej Universiti Linton, Mantin, Negeri Sembilan, West Malaysia
{cad_elango_n,come2nippon2000}@yahoo.com

Abstract. The Coanda effect, invented by Henri Marie Coanda is an interesting phenomenon in fluid mechanics. The basics of the Coanda effect is based on the property of a jet flow to attach itself to a nearby surface and to remain attached even when the surface bends away from the initial jet direction. This effect has been applied in many areas since 1960. In the recent years, researches have been initiated towards the application of Coanda effect in robots. This paper is focused on the utilization of Coanda effect in the field of robotics. The research that the authors are carrying out in this area is also addressed at the end of this paper.

Keywords: coanda effect, enthalpy, robot gripper.

1 Introduction

Robots have been developed and employed in almost all fields. The development in engineering and technology leads to advanced robotics in recent years. Different kinds of robots have been in research and development. Domestic robots which can perform housework, medical robots which can perform surgery, service robots which can work in hazardous environments are a few examples. Micro robots and biological robots are milestone of this decade. Intense research has been carried out on the actuation, sensor technology and intelligence. Coanda effect is a classic phenomenon in fluid mechanics and one of the fundamental discoveries of the Romanian inventor Henri Marie Coanda (1886-1972). In 1969, he defined the phenomenon: "… the Coanda effect relies on the principle of creating a depression zone in the air, along a wall, which allows the fluid to project itself onto and take the direction of the wall where the depression was generated… which can be used in various applications…" [1]. When the fluid is sent through a nozzle, a jet of fluid entrains and mixes with its surroundings as it flows away from the nozzle. When a surface or another stream is placed close to the jet, this restricts the entrained air flow from surroundings into that region. As flow accelerates trying to equalize the transfer momentum, a loss of pressure results across the jet and the jet is deflected closer to the surface up to attaching to it.

* Corresponding author.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 411–418.
springerlink.com

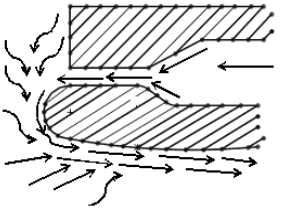


Fig. 1. Coandă patent [2]

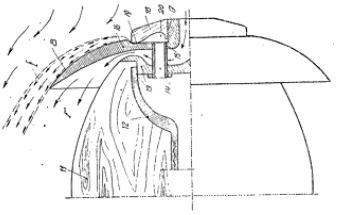


Fig. 2. Coandă patent [3]

Three different phenomena are associated with the name "Coanda":

1. The tendency of a fluid jet set tangentially on a curved surface to remain attached to that surface.
2. The ability of a free jet to attach itself to a nearby surface.
3. The tendency of jet flows over convex curved surfaces to entrain ambient fluid and increase more rapidly than that of plane wall jets [4].

The devices can be built using at least one of the phenomena of Coanda effect. Henri Coanda was the first to employ these ideas and received many patents [1] [2] [3] for devices utilizing one or more of these effects.

The following remarkable aspects were highlighted by Boscoianu et al [5].

1. The depressurized region determines:
    a) flow acceleration upstream in the slot, without increasing upstream pressure or temperature,
    b) the displacement of the local fluid.
2. Detaching and re-attaching is characterized by hysteresis (the re-attaching is formed at lesser angles than the detaching).
3. The total flow that results from the mixing between the main flow and the displaced flow is located in the depressurized region and is characterized by lower temperature.

Xu et al [6] used fluidic flip flop utilizing the Coanda effect to study the mechanical properties of human arm during unconstrained posture and movement. A new air jet design based on the Coanda effect which operates at a high frequency was presented in their research. The Coanda effect has been widely used in aeronautics, medical applications, air moving technology and other fields [7]. In recent years, it has been applied in robotics and some Coanda ejectors have been devised. This review paper focuses on the researches done on the application of Coanda effect in robots.

2 Analysis of the Mixing Process in Coanda Ejection Device

Analysis of the mixing process in Coanda ejection device has been elaborately discussed by Boscoianu et al [5].

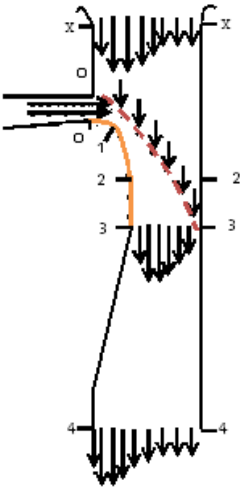


Fig. 3. Coanda ejector with: non uniform speed distribution

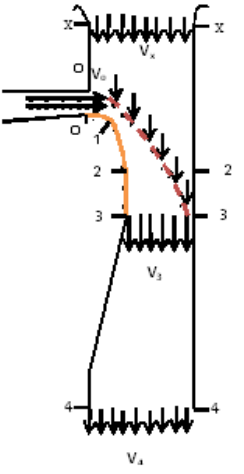


Fig. 4. Coanda ejector with: uniform speed distribution

The primary flow is introduced by compression or through absorption directly from the environment in the inlet (Section 0-0). The section x-x is the absorption section through which the resulting inflow travels into the system. It is characterized by the fact that the total enthalpy H of the flow is equal to the enthalpy of the environment H^*. The depressurization is supposed to be maximum at the area around 1. Section 2-2 shows the end of the Coanda profile (line 0-1-2). The absorption section ends at Section 3-3 and the thickness of the mixing area is equal to this section. 4-4 is the exit section from the ejection disposal and is characterized based on the fact that the static pressure P and the environment static pressure P^* are equal. The area x-0-3-2-x is considered to be the absorption area where the total enthalpy H of the flow and the environment H^* are same. Area 0-1-2-3-3-0 is considered to be the mixing area

where the entire quantity of generated flow is received through the permeable surface 3-0. Area 3-4-4-3 is the area of attaining homogeneity for aerothermogazodynamic parameters in section 3-3. It usually has a divergent form which adds to the increase of efficiency of the ejection device. The generated flow is increased due to its existence. The complete geometry of the ejection device will have to be taken into consideration while performing the research on the force increase. In order to obtain the highest force possible for an existing used energy it is better to put into motion the highest amount of fluid possible with the lowest speed possible instead of a small amount of fluid put into motion with a high speed [5].

The nature of the Coanda effect, with boundary layer separation and entrainment interaction make it more difficult in solving the flow numerically and analytically. In fact, Wille and Fernholz [8] claimed that there was no single solution to this type of flow. Hence, most recent work on the subject is based on experiments. Rask [9] and Patankar and Sridhar [10] had studied two-dimensional and three dimensional jet flows around cylindrical surfaces, looking at flow characteristics in all three dimensions (normal to surface, laterally across curved surface, and stream wise. Parks and Peterson [11] apply similarity analysis to approximate two-dimensional laminar Coanda flow, finding mass entrainment, thrust, and jet-sheet thickness formulas. The analysis is valid only for $b/a \ll 1$ and for Reynolds number less than 3×10^4. Moser [12] used numerical simulation on low Reynolds number convex and concave surfaces and found constant results with experimental data. Sawada and Asami [13] found a solution for two and three dimensional supersonic Coanda jets. Wernz and Fasel [14] studied turbulent Coanda flow using three-dimensional Navier-Stokes simulations.

3 Applications of Coanda Effect in Robots

It is found from the literature that only a limited research has been done in the field of robotics. Some have worked on ejectors as grippers for handling different objects such as textile fabrics, meat, fish, vegetables and fruits. Some have attempted to develop actuators which could work with Coanda effect.

3.1 Coanda Ejectors as Gripper

Coanda ejectors can generate moderate underpressure and large airflow locally. The following ejectors (Figures 5 and 6) were conceived and patented by a Romanian private company (AERODIN srl). It has been obtained at 4.7 bar, a flow of 2750 m3/h at a medium pressure loss of 90 mm H2O having an ejection coefficient of 60 [15].

Fig. 5. Level ejection device

Fig. 6. Cylindrical ejection device

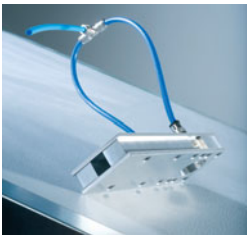


Fig. 7. The Coanda gripper: Prototype [16]

The Norwegian University of Science and Technology (NTNU) developed a Coanda gripper for automatic handling of soft and porous materials such as leather and textiles without damaging the objects. It was also later patented. University of Salford has also developed a novel non-contact end effector (Figure 7) which operates on the principle of generating a high-speed fluid flow between the nozzle(s) and product surface, thereby creating a vacuum which levitates the product. Erzincanli et al [17] has been working on non-contact robotic handling system for non-rigid materials. The advantages of non-contact type gripper over other kinds of grippers; bladder type, adhesive type, needle type and clamping gripper have been discussed in their research paper. The nozzles of these grippers can be manufactured not only from Stainless steel but also from rigid plastics and aluminium, depending on the hygiene requirements and availability of the materials. If the nozzle is made of materials of low specific mass, payload on the robot arm will get reduced. Erzincanli et al was discussing the requirement of many nozzles and impact of surface finish of objects as well. In order to create an attracting force, the clearance gap of a nozzle must be very small compared with the diameter of the central tube and the radial distance through which the fluid flows. The gap between the end-effector and the non-rigid product is to some extent self-regulating, and imposes local rigidity which is useful during manipulation. A single nozzle could be used for a rigid material, by applying a low air flow rate and keeping a small coverage area [18]. Even though the non-contact end effector has several advantages over the conventional end effectors, it has the following disadvantages.

1. The radial air outflow of a product with a relatively compliant structure cannot be gotten due to the presence of indentations and local bending on the surface of the product.

2. The behaviour of the surface of the product is inhomogeneous if the fibres of the product are loose, example beefsteak. The air flow may diffuse between fibres.

According to Bonfield [19], a vision system can provide information of the location, orientation, size and shape of the products to be handled. This information can then be used to actuate the robot and nozzle valves. The gap distance between the nozzle(s) and the product surface will be measured by sensors. The gap could be between 0.1 to 0.3mm. The ejector will be stopped according to the gap information and the corresponding air valves will then be turned on. The possibility of nozzle contact with the product to be lifted will be eradicated by these sensors.

Ameri and Dybbs [20] did a detailed flow analysis for cylindrical Coanda ejectors and presented a method of calculations for flow field of the ejectors. Their modelling reveals that the ejector can have the ratio of secondary to primary flow in the order of 10 which will be well suited for suction gripping of porous materials. An array of Coanda ejectors could also be used for picking up piles of textiles. Since the suction gripping finds difficulty in handling porous materials, uneven and rough leathers, Lien and Davis [21] has developed a novel gripper for handling limp materials based on lateral Coanda ejectors. Lien and Davis devised a planar Coanda ejector and compared the performance of it with a standard Coanda ejector. They established the following design rules: (1) Coanda suction ejectors should have a diffuser length around 8 times the diffuser width (2) Single sided planar Coanda suction ejectors should have height/width (h/b) ratio of 2 for the diffuser. It is concluded that the performance of the planar Coanda ejector in terms of underpressure as a function of primary flow is better. They conducted experiments on four different materials and settled on the following: (1) the diffuser channel should be quite narrow for best performance on leather materials (2) it must be wider for porous materials (3) the height over width ratio of 2 is the best compromise for universal use (4) The multi head gripper can also be developed.

3.2 Coanda Effect for Actuation

Mazumdar and Asada [22] developed a compact egg-shaped body for underwater inspection. A multiaxis integrated thruster mechanism using Coanda effect was used in this manoeuvrable compact vehicle, which could manoeuvre in a cluttered environment by the coordination of high band width Coanda effect valves. A Coanda jet device (CJD) with two control ports which operates on Coanda effect to switch the direction of water jet between exit 1 and exit 2 was developed. A jet actuator system for both forward motion as well as manoeuvring was also proposed.

Further research is presently being carried out by the authors of this paper on the design and development of robot for handling vegetables and fruits. The principle of Coanda effect will be used for actuation and gripping of soft materials as well.

4 Conclusion

This paper has discussed the Coanda effect principle, analysis of the mixing process in the Coanda ejection device and various applications of Coanda effect in robots.

Coanda effect has been applied in grippers for handling of soft, hard and porous materials. Design rules have been established from a comparison of a planar Coanda ejector with a standard Coanda ejector. High-band width Coanda effect valves have been applied in a compact underwater vehicle for low speed precision manoeuvring in mixed-up environments. Due to the rapid growth in engineering and technology and the need for advanced robotics in some special areas; the application of Coanda effect in robots would certainly be a great contribution. Therefore, further research is currently in progress to identify more applications of Coanda effect in robots.

References

1. Coandă, H.: Procédé de propulsion dans un fluide. Brevet d'invention France, No. 762.688 (November 23, 1932)
2. Coanda, H.: Procédé et dispositif pour faire dévier une veine de fluide pénétrant dans un autre fluide. Brevet d'invention France, No. 792.754 (October 08, 1934)
3. Coanda, H.: Perfectionnement aux propulseurs. Brevet d'invention France, no. 796.843 (January 15, 1935)
4. Circiu, I., Dinea, S.: Review of Applications on Coanda Effect. History, Theories, New Trends. In: Proceedings of International Conference on Review of the Airforce Academy, vol. 2, pp. 14–21 (2010)
5. Boscoianu, M., Prisecariu, V., Circiu, I.: Applications and Computational Aspects Regarding the Coanda Effect. Science & Military 1, 26–30 (2010)
6. Xu, Y., Hunter, W.I., Hollerbach, M.J., Bennett, J.D.: An Airjet Actuator System for Identification of the Human Arm Joint Mechanical Properties. IEEE Transactions on Biomedical Engineering 38(11), 1111–1122 (1991)
7. Joyce, J.W., Gottron, R.N.: Fluidics: Basic Components and Applications, T U.S. Army Electronics Research and Development Command, Harry Diamond Laboratories (1979)
8. Wille, R., Fernholz, H.: Report on the First European Mechanics Colloquium. Journal of Fluid Mechanics 23, 801–819 (1965)
9. Rask, R.B.: An Experimental Study of Two-Dimensional and Three-Dimensional Curved Wall Jets, Ph.D. thesis, University of Minnesota (1973)
10. Patankar, U.M., Sridhar, K.: Three-Dimensional Curved Wall Jets. Journal of Basic Engineering (changed to Journal of Engineering Materials and Technology; and The Journal of Fluids Engineering) 94(2), 339–344 (1972)
11. Parks, Peterson: AGARDograph AG-169 (1990)
12. Moser, R.D.: The Effects of Curvature in Wall-Bounded Turbulent Flows. Journal of Fluid Mechanics 175, 479–510 (1987)
13. Sawada, K., Asami, K.: Numerical Study on the Underexpanded Coanda Jet. Journal of Aircraft 34(5), 641–647 (1997)
14. Wernz, S., Fasel, H.F.: Numerical Investigation of Transition Mechanisms Influencing the Development of Turbulent Wall Jets. In: Collection of Technical Papers - 36th AIAA Fluid Dynamics Conference, vol. 4, pp. 2442–2457 (2006)
15. Dinea, S.: Contribuţii la studiul efectului Coandă, Teză de doctorat, Bucureşti (2009)
16. Gripper, C.: NTNU Technology Transfer, Patent number WO 2009/054732 (2009)
17. Erzincanli, F., Sharp, J.M., Erhal, S.: Design and Operational Considerations of a Non-Contact Robotic Handling System for Non-Rigid Materials. Journal of Tools Manufacturing 38(4), 353–361 (1998)

18. Erzincanli, F., Sharp, J.M.: Robotic Handling Using Non-contact End Effectors: The Effects of Nozzle Configuration on Rigid Materials. In: Symposium on Robotics and Cybernetics, CESA 1996 IMACS–IEEE/SMC Multiconference, Lille, France, pp. 371–376 (1996)
19. Bonfield, T.: Flexible Vision-Assisted Handling Systems for the Packaging of Fresh Meat and other Products, M.Sc. Thesis, University of Salford (1993)
20. Ameri, M., Dybbs, A.: Theoretical Modeling of Coanda Ejectors, American Society of Mechanical Engineers, Fluid Engineering Division (Publication) FED. Fluid Machinery 163, 43–48 (1993)
21. Lien, T.K., Davis, P.G.G.: A Novel Gripper for Limp Materials Based on Lateral Coanda Ejectors. Journal of Manufacturing Technology 57, 33–36 (2008)
22. Mazumdar, A., Asada, H.: A Compact Underwater Vehicle Using High-Bandwidth Coanda-Effect Valves for Low Speed Precision Manoeuvring in Cluttered Environments. In: IEEE International Conference on Robotics and Automation, China, pp. 1544–1550 (2011)

Servo Motor Position Control Based on DSP

Zhenyan Wang, Zhimei Chen, and Jinggang Zhang

School of Electronics and Information Engineering,
Taiyuan University of Science and Technology, 030024 Taiyuan, P.R. China
w9851@126.com

Abstract. The position control system for Permanent magnet DC servo motor based on DSP is presented in this paper. The hardware and software design are implemented to control the position system accurately. A Buck circuit is utilized to execute step-down chopper and to suppress the motor iron loss, and the Buck circuit component parameters are achieved by Pspice simulation. The control system consists of three loops: position loop, speed loop and current loop. Three PID controllers are designed and evaluated according to the feedback signals measured by Hall current sensor and the optical encoder. The control performance is verified experimentally for the position servo motor. Experiment results show that high-precision positioning control is achieved and prove the effectiveness of the proposed position control system.

Keywords: DSP, DC motor, Position servo control system, Position control.

1 Introduction

Permanent magnet DC servo motor has good servo performance and line quality, and is usually used in position servo control. In terms of performance, permanent magnet DC servo motor is superior to other motor, as of AC variable frequency motor and step motor. Supplied by DC power, permanent magnet DC servo motor has the advantages of convenient system power supply design, simple power supply design and small power system losses. Therefore, DC motor is still used for the large complete sets of production equipment and production line in Europe, the United States and other countries.

In recent years, the structure and control diversification of DC motor is an inevitable trend. With the development of the embed system in the control field and the development of new power electronic power components, most studies focused on the controller design using modified PID algorithm [1] or complex control algorithm[2-4], and the pulse width modulation(PWM) control using full-controlled switch power component[5]. This design concept is the basis for digital control of DC motor.

The buck converter is well known in power electronics [6-8]. As a step-down DC to DC converter, buck converter is often used to reduce torque ripple [7] or suppress iron loss [8]. In this paper, a Buck converter is introduced and developed in the control system and its parameters are obtained by Pspice simulation. As well as, the new low-power motor driver modules can reduce system power consumption and

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 419–426.
springerlink.com

improve the efficiency. With the high-speed computing performance, DSP is utilized to realize complex control algorithms and regulate on-line the motor control parameters. The study of the new permanent magnet DC servo motor control system based on DSP is essential and adapt to the current international environment need for energy saving.

2 System Architecture

2.1 Motor Control Principle

It is well konown that the motor speed n can be expressed as Eq.(1) according to DC motor equivalent circuit and mathematical models .

$$n = \frac{U_a - (I_a R_a + L_a \frac{dI_a}{dt})}{K_e \Phi} \tag{1}$$

Where U_a and I_a are armature voltage and current, and R_a and L_a are the total equivalent resistance and inductance of the armature circuit. K_e denotes the EMF constant; Φ and n denote Flux per pole and motor speed respectively.

It can be seen that the motor speed can be adjusted by changing armature voltage so as to regulate armature current. As well as, the motor rotation direction can be regulated by changing the polarity of armature voltage. Therefore, the position control can be realized by adjusting armature voltage.

2.2 System Framework

The entire closed-loop system consists of three loops: position loop, speed loop and current loop. The overall block diagram is shown in Fig.1.

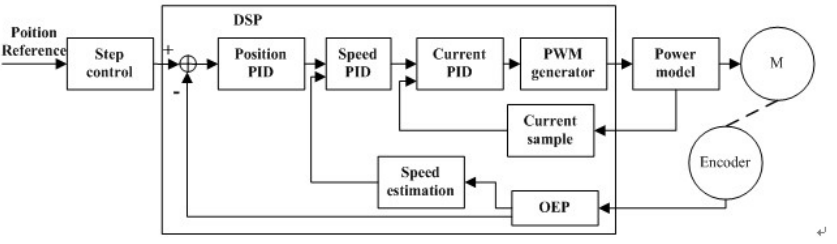


Fig. 1. Motor position servo control system framework based on DSP

The motor position can be captured by a 2000-lines incremental optical encoder. Control system have IO interface, which can receive the instruction signal of position and direction from PC, prohibit the forward rotation and reverse rotation, and export the fault signal of motor driver system.

Because the position servo motor need rotate frequently to achieve precision position, the high-frequency current transition in the motor windings and the consequential magnetic field transition will induce large eddy current loss in the stator and rotor. In the position control system, a Buck converter is designed to suppress iron loss of motor, which can provide modulated voltage source for the inverter system. It is highly efficient, simple to design mostly due to not having a transformer, puts minimal stress on the switch, and requires a relatively small output filter for low output ripple. The inductance is paralleled to the circuit to reduce the influence from the high-frequency current transition and to improve switching frequency. Therefore, the given motor voltage is modulated by the chopper circuit.

3 Hardware Circuit Design

The hardware circuit of the position control system includes the electrical isolation circuit, current detection and conditioning circuit, the position feedback circuit, step-down chopper circuit, and H-bridge power driver circuit, etc.

TMS320F28335 IC, produced by U.S. TI Company, is a floating type DSP of 2000 series. It contains the AD, PWM, SCI and CAN modules, and does not need external extension [9]. In this position servo control system, DSP TMS320F28335PGFA is utilized to run the control algorithms of the three closed loops. The logic functions including power monitoring and restricting the motor reversing, were completed by FPGA coprocessor.

In order to avoid the electromagnetic interference to control signal, the electrical isolation should be done between power amplifier and control signal, especially with the high-power situation. Photo coupler isolation circuit is designed to isolate the signal between power pulse amplifier and control pulse signal. Photo coupler output is pulled up by +5V and a Schmitt inverter is added to the photo coupler to avoid the photo coupler failure. A multi-range current sensor LA28-NP, produced by Switzerland LEM Company, is adopted for the current detection. On the other hand, a conditioning circuit is designed for the detected current signal because the DC motor armature winding current has directional.

For the position control system, the motor position is coded by a 2000-line incremental encoder, and the encoder output has the square wave differential form. It is important to note that the twisted-pair shielded cable is required for the encoder signal transmission and the encoder cable connector shell need to be connected to cable shield and to ground.

MC33887, produced by Freescale company, can be used as motor driver module. As an H-bridge power switch IC, the chip includes the control logic, charge pump, gate driver and MOSFET output circuits, etc. It has self-protection abilities including short-circuit protection, under-voltage protection, over-temperature protection and so on. Considering the limited drive capability of the single MC33887, multi-chips can be paralleled to improve driver ability. Here two chips are utilized.

In order to suppress iron loss of motor, a Buck circuit is designed to provide modulated voltage source for the inverter system in this control system. The Buck circuit topology is shown in Fig.2. Considering that the converter contains two energy storing elements, a coil and a capacitor, smooth DC output voltages and currents with

very small current ripple can be generated [10]. The operating mode of Buck circuit include continues mode and discontinues mode. The Buck converter operates in discontinuous mode when low current is drawn by the load, and in continuous mode at higher load current levels. The critical condition between the continuous mode and discontinuous mode can be described as Eq.(2).

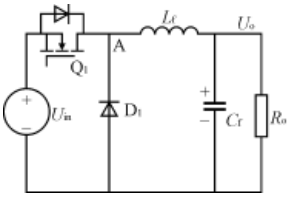


Fig. 2. The Buck circuit topology

$$f_s L / R \geq (1-D)/2 \tag{2}$$

Where D denotes the duty cycle; f_s denotes the switching frequency; R and L denote the load resistance and inductance. The Buck will work in continues mode when Eq.(3) expression is satisfied. And the filter capacitor C can be selected according to the Eq.(4).

$$L_{\min} = (1-D)R/(2f_s) \tag{3}$$

$$C \geq U_o(1-D)/(8Lf_s^2 \Delta U_o) \tag{4}$$

The Buck circuit component parameters are achieved by Pspice simulation. The simulation results show the current amplitude of 3A, the current ripple of 0.3A, the inductance of 1mH, the capacitance of 22uF and the cut-off frequency between 1kHz and 30kHz.The Buck circuit principle is displayed as Fig.3.

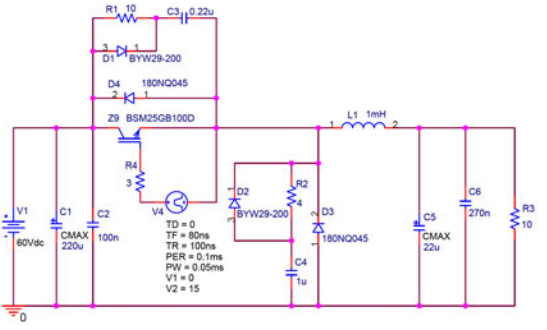


Fig. 3. The Schematic diagram of Buck circuit principle for simulation

4 Control System Software Design

4.1 Software Flow Chart

In the software design of the position control system, the PWM module, QEP module, ADC module and other module initialization are performed in the main program. The regulation programs of position loop, speed loop and current loop are completed in the interrupt program, and then external interrupt is executed by DSP. The software flow chart is shown in Fig.4.

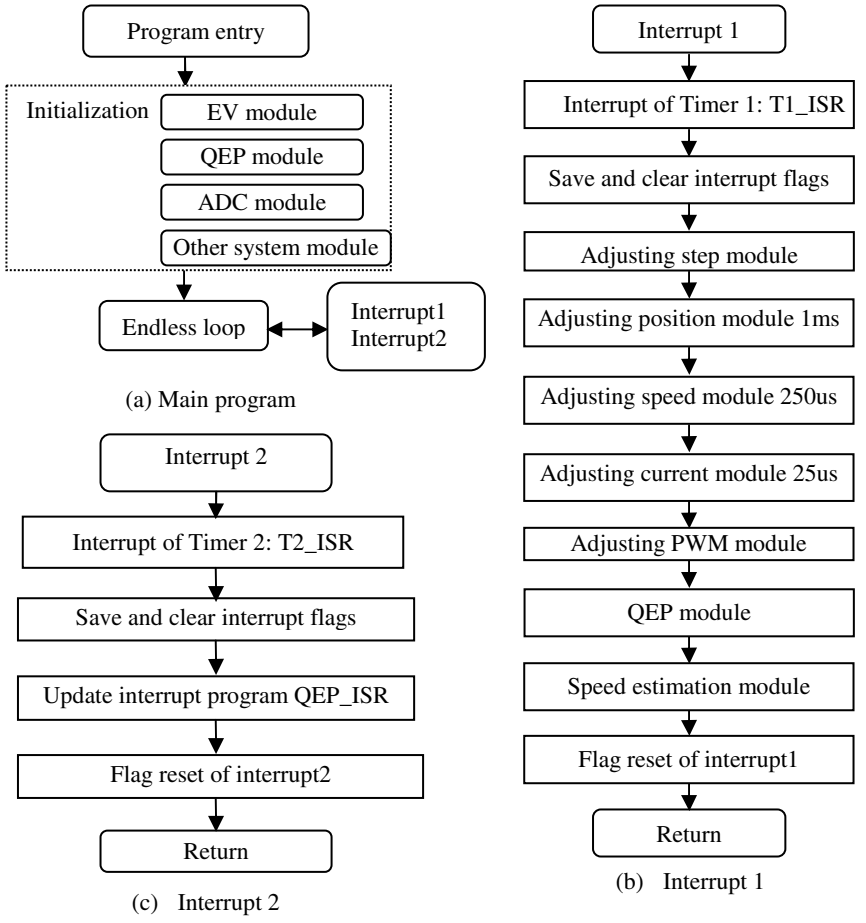


Fig. 4. Software flow chart

4.2 Control Algorithm

The position control system consists of three closed loops: position loop, speed loop and current loop. So the three controllers are required for the control system.

PID control algorithm is simple because of its good robustness and high reliability, widely used in position control areas. In engineering, the PID parameters are often tuned through error frequently [11]. Here an incremental PID control algorithm with separate-segment variable gain is proposed for the position loop control to reduce the number of possible adjustments and tunings. The incremental PID control is shown as Eq.(5).

$$\Delta U(k) = K_P(e(k) - e(k-1)) + K_I e(k) + K_D(e(k) + e(k-2) - 2e(k-1)) \tag{5}$$

Where K_P, K_I and K_D are proportional, integral and differential gains respectively. $e(k)$,$e(k\text{-}1)$ and $e(k\text{-}2)$ are error signal of the reference signal and the practical output signal at the time k, (k-1) and (k-2). The proportional and integral gains K_P and K_I for the position loop change according to the following expression.

$$K_P \in \begin{cases} \begin{bmatrix} K_P^1 & K_P^2 \end{bmatrix} & if\ |e(k)| > e_{ref} \\ \begin{bmatrix} 0 & K_P^2 \end{bmatrix} & if\ |e(k)| < e_{ref} \end{cases} \tag{6}$$

$$K_I \in \begin{cases} \begin{bmatrix} 0 & K_I^1 \end{bmatrix} & if\ |e(k)| > e_{ref} \\ \begin{bmatrix} K_I^1 & K_I^2 \end{bmatrix} & if\ |e(k)| \le e_{ref} \end{cases} \tag{7}$$

Where $0 < K_P^1 < K_P^2$ and $0 < K_I^1 < K_I^2$.

The control algorithm for the speed loop and current loop is design according to the error signal and the given reference signal. The reference V_{ref} and I_{ref} of the speed loop and current loop also mean the output signal of the position loop and speed loop. It can be seen form Fig.1 that the feedback signal V_{fdb} of speed loop can be calculated with T-method by the checked position from encoder module. And the feedback signal I_{fdb} of current loop can be obtained by ADC measured by current sensor LA28-NP and conditioning circuit. The speed loop and current loop controller is presented as Eq.(8-9).

$$U^V(k) = K_P^V \times [K_D^V \times (V(k-1)_{fdb} - V(k)_{fdb}) + V(k)_{ref}] + K_I^V \times U^V(k-1) \tag{8}$$

$$U^I(k) = K_P^C \times [K_D^C \times (I(k-1)_{fdb} - I(k)_{fdb}) + I(k)_{ref}] + K_I^C \times U^I(k-1) \tag{9}$$

Following the engineering design principle, the PID parameters tuning of the inner loop controller are given priority over the outer loop. The current loop controller is regulated firstly. Then, the current loop can be seen as a part of the overall speed regulating system, and the speed regulator is designed. Similarly, the regulator of position loop can be realized.

5 Experiment Result

To verify the high-precision positioning performance of the control system proposed in this paper, an experiment system based on DSP is performed with a brush DC servo motor control system. Motor parameters of experimental system are shown as Table 1. The motor load is hystereis brake.

Table 1. Motor parameters

Rated torque (Nm)	Stall torque (Nm)	Rated current (A)	Rated voltage (V)	Rated speed (rpm)
1.5	1.5	2.5	60	1000
Peal current (A max)	Power (W)	Maximum speed (rpm)	Back EMF constant (V/Krpm)	Torque constant (Nm/A)
22	200	1200	49	0.4

Experiment proved that test motor can operate smoothly in low speed, and the minimum speed can be run on 1rpm. For the position loop, the proportional gains k_P^1=50 and k_P^2=250, and the integral gains k_I^1=50 and k_I^2=250. As well as, the proportional gains of speed loop and current loop are K_P^v=10 and K_P^C=1; the proportional gains of speed loop and current loop are K_I^v=100 and K_I^C=10.

Using the Code Composer Studio (CCS), the given motor position signal is shown as the top waveform and the practical position signal are displayed as the button waveform in Fig.7. The abscissa denotes the time and the vertical axis express the position. It can be seen that the given signal is almost equivalent to the position feedback signal, and the phase difference is almost equal to zero. Experiment results show that the control system can drive motor with good tracking effect.

When the motor rotation commutated, the current waveform is shown in Fig.8. Obviously, the convergence speed is fast as the smaller overshoot of current.

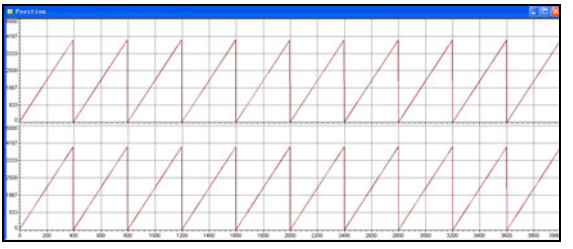


Fig. 5. Given position and practical position signal

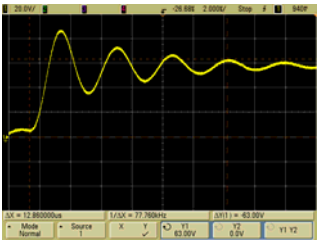


Fig. 6. The current waveform at motor commutation

6 Conclusions

The position servo control system based DSP is presented in this paper. The hardware architecture and software process are completed. The Buck circuit and its parameters are designed, and the experimental data are achieved. The control algorithm and parameters in three closed loop subsystems are described. Experiment result show that the position tracking control system for permanent magnet servo motor can achieve high-precision tracking and can be used when the motor rotates frequently.

Acknowledgement. This work is supported by the Shanxi Province National Natural Science Foundation of China (2011011011-2) and the Doctoral Fund of Colleges and Universities(Positioning and Anti-sway Control for Crane).

References

1. El-Gammal, A.A.A., El-Samahy, A.A.A.: Modified Design of PID Controller for DC Motor Drives Using Particle Swarm Optimization PSO. In: International Conference on Power Engineering, pp. 419–424. IEEE Press, Lisbon (2009)
2. Lin, L., Tang, H.W., Tang, J.: Adaptive Backstepping Control for an Interior Permanent Magnet Synchronous Motor Servo System. Advanced Materials Research 317-319, 120–123 (2011)
3. Zhou, R., Yuan, W., Zhangfei: Application of NN-PI Controller in Direct Current Motor Servo System. In: International Conference on Electronic Measurement and Instruments, pp. 4468–4471. IEEE Press, Xi'an (2007)
4. Mazumder, S.K., Nayfeh, A.H., Borojevic, A.: Robust Control of Parallel DC–DC Buck Converters by Combining Integral-Variable-Structure and Multiple-Sliding-Surface Control Schemes. IEEE Transactions on Power Electronics 17, 428–437 (2002)
5. Wang, J., Xu, J.: A Novel PWM Control Method for Switching DC-DC Converters with Improved Dynamic Response Performance. In: International Symposium on Power Electronics for Distributed Generation Systems, pp. 85–88 (2010)
6. Raja Ismail, R.M.T., Ahmad, M.A., Ramli, M.S.: Speed Control of Buck-converter Driven Dc Motor Based on Smooth Trajectory Tracking. In: International Conference on Modelling and Simulation, pp. 97–101. IEEE Press, Hefei (2009)
7. Zhang, X., Lu, Z.: A New BLDC Motor Drives Method Based on BUCK Converter for Torque Ripple Reduction. In: International Power Electronics and Motion Control Conference, vol. 3, pp. 1631–1635 (2007)
8. Linares-Flores, J., Reger, J., Sira-Ramirez, H.: Speed-sensorless tracking control of a DC-motor via a double Buck-converter. In: Proceedings of the IEEE Conference on Decision and Control, pp. 6229–6234. IEEE Press, San Diego (2006)
9. TexsInstruments. Literature Number: SPRS439A (2007)
10. Guo, L.: Implementation of Digital PID Controllers for Dc-Dc Converters Using Digital Signal Processors. In: IEEE International Conference on Electro/Information Technology, pp. 306–311. IEEE Press, Chicago (2007)
11. Geethanjali, P., Vijaya Priya, P., Kowsalya, M., Raju, J.: Design and Simulation of Digital PID Controller for Open loop and Closed Loop Control of Buck Converter. IJCST 1(2), 202–206 (2010)

Based on the 3G Oilfield in Remote Monitoring Data Transmission Systems Research

Gong Chang Ren, Qing Ye, Bo Chen, and Yong Fei Wang

College of Mechanical & Electrical Engineering,
Shannxi university of Science & Technology, Xi'an, China, 710021
rengc@sust.edu.cn, 419727812@qq.com,
chen_bo_16@163.com, yongfeio@126.com

Abstract. Focusing on the disadvantages of the present domestic oilfield in security and measuring distance monitoring technology, formulated the overall scheme of the data transmission system. Concretely analyze the wireless transmission system based on the 3G network, study the connection between 3G network with data terminal equipment and upper computer, as well as to analyze the reliability of 3G modular during the data transfer.

Keywords: 3G, Data transmission, reliability.

1 Introduction

With the development of the telecommunication, the computer, the network technology and the fusion, 3G[1] technology (the abbreviation of the third generation mobile communication technology) have been matured, which is able to handle many media forms ,such as music, video, images and so on. Most of the oil fields in our country are distributed , that is, the number of oil Wells is big, but their distribution is scattered and the distance between the drills is far away. Using wireless data transmission system to realize the data acquisition, run the well site condition monitoring, and measure the project establishment has become an inevitable tendency in distributed oilfield. Using 3G networks of wireless data transmission can be more comprehensive, rapid, stable and reliable and it also can reduce the cost and improve the operation efficiency.

1.1 System Solutions

The transmission system is made up of the data terminal part, the network part and the upper computer. The Data transmission system occupies a very important role in the whole remote monitoring system. The rationality of the design will directly affect the actual data to feed back to the upper computer promptly, accurately and rapidly. This can timely control the drill dynamic .The design of the whole system structure as shown in figure 1 below.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 427–432.
springerlink.com

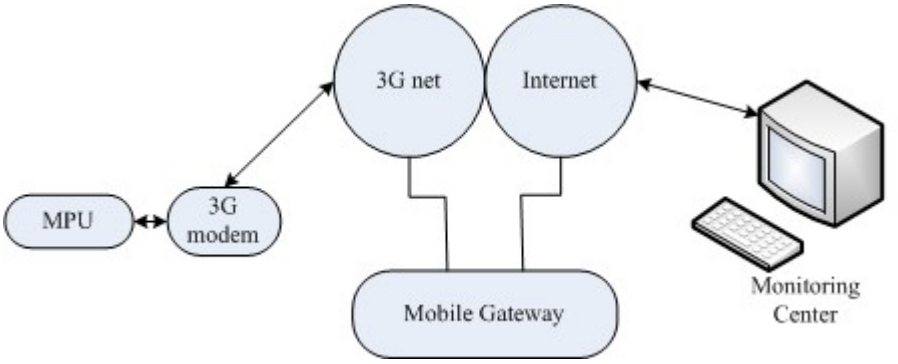


Fig. 1. The whole system structure

2 The Choice for 3G Wireless Transmission

2.1 Reality of Choosing 3G

Owing to the oil fields are distributed in very special environment place, the primary problem to realize the rapid , reliable communications is to overcome the geographical position of limitations from Oil field itself. To establish a communication system is a process which will waste a lot of time and high cost, and after construction it brings a lot of difficulties for system maintenance. While 3G network can provide greater system capacity and higher data transfer rate to support wireless access to the Internet and wireless business. It makes personal terminal realize high quality information of mobile communication and transmission in context of the global with any way no matter any time, any place and anyone. Compared With GPRS wireless data communication, it has incomparable advantages[2]. Therefore, using 3G wireless data transmission scheme not only solved the problem occurred at the beginning of the project but also has higher reliability, security, stability and rapidity.

2.2 The Advantage of the 3G Wireless Transmission

The reason for choosing 3G standard-TD-SCDMA developed from mainland China is that it overcomes the breathing effect and distant effect. It can't increase the interference and affect the cover along with the increase of the terminal equipment, and it can also adjust power with terminal equipment distance. The biggest characteristics is that it has the smart antenna. Base station system through the digital signal processing technology and adaptive algorithms make the smart antenna covered in space dynamically, it still forms the directional equipment beam to users, it also make full use of downside signal energy to suppress the interference signals. Base station can be in the location of the mobile terminal in tracking through the smart antenna. So the signal-to-noise ratio of the terminal has been greatly improved, and the efficiency of the transmission quality has been improved too.

3 The Realization of 3G Wireless Data Transmission

3.1 Data Terminal Realize 3G Transmission

Data terminal includes data acquisition part combined by single-chip microcomputer and sensor and wireless terminals (3G Modem) communication modem part. Its main task is to complete the data of the acquisition and to send the processing and packaging data.

Data acquisition system is used in a series of single AVR chip. It can prefetch instructions, break fast response, speed the data processing. It's also cheap with high reliability and powerful I/O mouth functions. Specific SCM dispose data collected by Sensors as shown in figure 2 shows:

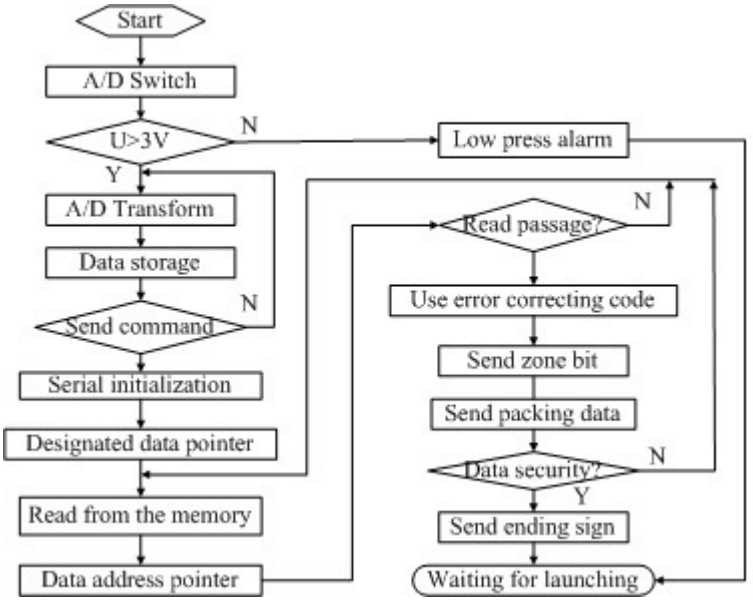


Fig. 2. The data acquisition software flow chart

3G Modem support to send and receive the Chinese, English, digital text messages. It provides VB/VC/c # / DEPHI DLL and test programs, and it also provides a convenience for the user in the development and integration. Its system design, industrial anti-jamming ability and characteristics of the stability of the system are suitable for use in the industrial control area. At the same time it has excellent low and high temperature performance. It is convenient, flexible and reliable to use .Using single-chip microcomputer and 3G Modem, it solves well the original faults of not effectively support image and remote monitoring. Single-chip microcomputer and 3G Modem are connected through a standard serial port.

Data acquisition system will collect the well data timefixed or recyclable. When receiving a PC instruction, it will process the collected data according to the communication protocol and send to 3G Modem .The Workflow diagram is shown below between the Data terminal and 3G Modem.

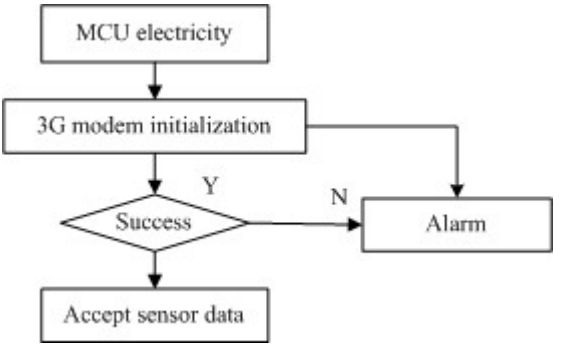


Fig. 3. Data terminal 3G transmission flow chart of realization

3.2 PC Realizes 3G Transmission

In the upper machine, it needs a Wireless network card to realize the data transmission. Its function likes a cable modem, with wireless signal coverage in any place, it can be connected to the internet using USIM card. Figure 4 is the flow chart of its job.

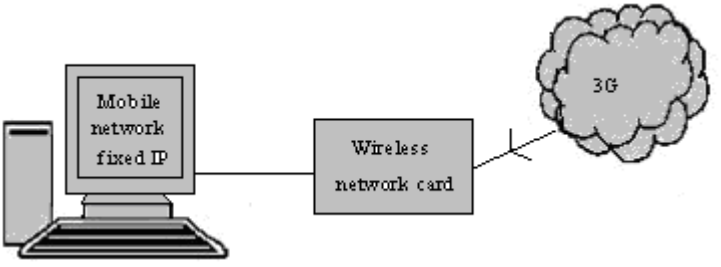


Fig. 4. PC accept data schemes

After the installation and dial-up connection of the Wireless network card, upper computer can search signal for data transmission with 3G Modem, and it will send the data to the database after receiving.

3.3 AT Instructions

The communication agreement between the 3G wireless data transmission modem and the data acquisition modem is through the AT instructions to communicate, following the "AT command set for GSM Mobile Equipment (ME) (GSM 07.07 version 6.4.0 Release 1997)" agreement standard. In this case, it need to set their communication AT

instructions among them to ensure the normal wireless data transmission for the whole system, so it can provide stable, reliable safeguard for the whole system. AT instructions are send from terminal equipment (TE) or data terminal equipment (DTE) to data terminal equipment. The basic interaction diagram between the detailed diagram for the terminal and the control mobile is shown in figure 5.

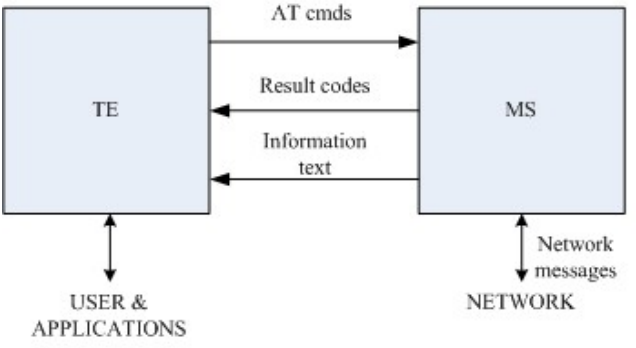


Fig. 5. TE and MS interaction diagram

3.4 The Realization of 3G Communications

(1) initial: oilfield terminal of the microprocessor must first detect 3G modem state in connecting to the server. If without initialization, it should be initialized.
(2) dial-up connection: the monitoring center has been in protecting reliably against detective state. After the data acquisition modem and 3G modem being in electrify, AT instructions will connect data center server. When the communication pathway is unimpeded, the monitoring center computer will send the predetermined results yards.
(3) mission: after the modem and data center connecting successfully, modem begans to sent the data gained by The data acquisition modem to the data center server in the specified data form. Data center server has been in the interception state, which can receive, store and handle these data.
(4) disconnected: when the task has been executed, it will be automatically disconnect.

4 The Reliability Analysis of 3G Wireless Data Transmission Modems

4.1 The Analysis of Transmission Capacity

The popular wireless network is GSM and GPRS. GSM's transmission rate is generally in 9.6kbps, and the theoretical maximum transfer rate of GPRS is in 171.2kbps. In fact, due to the network and terminal realistic conditions, the actual rate is not up so high. For real-time remoting monitoring of images and running the sound is impossible. While the transmission rate of 3G can be up to 144 KBPS~2 MBPS, it can fully meet the demand of user.

4.2 The Analysis of Safety

Using 3G network to communicate with the position machine must solve the problem that is the choice of Communication network and Internet data transmission chiefly. The system uses the TCP protocol, which is the reliable data transfer protocol with high data security, and it provide reliable ,orderly, end to end data transmission service , in the meantime it can provide the data encryption technology and signature technology in transmission.

4.3 The Analysis of Instantaneity

At present the time delay of data transmission is 80ms-200ms. The data packet sent from 3G terminal will reach the data center such a short period of time delay. Also the data transmission will also need 80ms-200ms from the data center to the 3G terminal. It is enough to meet the time need of the occasion drill.

5 Conclusion

In view of the popular 3G network technology at present, this paper mainly Analysis and research the data transmission system based on 3G wireless. The system can realize the wireless data transmission in reliable and stable condition. It has a great guidance to remote data transmission system in the oilfield and granary, and will improve the operating efficiency and reduce the production costs.

References

1. Huang, J.Z.: 3G Technology and Its Application in the Correspondence. China's High Technology and New Technology Enterprise 16(2), 117–118 (2009)
2. Zheng, L.: The Overview of 3G Mobile Communication Technology. Communications Market 26(1), 55–57 (2003)
3. Hu, Y.H., Jia, J., et al.: Ship Remote Monitoring System Based on 3G Technology. Science, Ship Technology 31, 101–103 (2009)
4. Zhang, P.Z.: Wireless Data Technology Industry and Its Implementation Scheme Application. China Telecom Construction 21(8), 11–13 (2003)
5. Zhang, J.J.: Telecom 3G Business Realize Family Video Monitor. Science and Technology Information Development and Economic 38(1), 167–170 (2009)
6. Zhai, Y.F., Zhang, T.P.: The Intelligent Home Control System Design Based on AVR Single-Chip Micro Computer. Microcomputer and Its Applications 30(11), 98–104 (2011)

An Improved Linear Target Detection Method Based on Probabilistic Hough Transform in a Remote Sensing Image

Kun Gao, Yan Wang, and Guo-qiang Ni

Key Laboratory of Photoelectronic Imaging Technology and System,
Ministry of Education of China
National Key Laboratory of Science and Technology on Low Light Level Night Vision
School of Optoelectronics, Beijing Institute of Technology
Beijing, China, 100081

Abstract. Regular linear targets detection in a remote sensing image depends on how to extract the long straight edges of the target from the background efficiently. Hough Transform is a classic method for line detection. But Standard Hough Transform (SHT) may induce too many short lines in the complex background to distinguish the true long linear targets. In this paper, an improved method is proposed, which uses probabilistic Hough Transform (PHT), a faster version of SHT, to detect short lines and then cluster short lines into groups to avoid most of the interfering lines. By this means, the reliability of target detection is improved. Experiments show that our method can detect the linear target efficiently when the background is large and complex.

Keywords: Hough Transform, linear target, remote sensing.

1 Introduction

In remote sensing image, many artificial objects, such as airports and bridges, own the distinct features of one pair long and straight parallel edges in the complex background. In most condition, to detect objects, one effective way is to find out their edges firstly, and then refine the results, according to the interested targets' other features, such as shape, texture, intense and so on.

Line detection is the fundamental problem in computer vision. Lots of researches have been tried to resolve this problem in the past decades and various methods have been proposed. Among them, Standard Hough Transform (SHT) [1] is a powerful method of line extraction, which was proposed by Hough firstly. Since this method utilizes the local feature to vote for all the possible lines, HT is a robust and powerful method to extract lines in very noisy images, and even when the line is discontinuous.

In SHT, each of the M edge pixels (x_i, y_i) is mapped to a sinusoidal voting pattern

$$\rho = x_i \cos\theta + y_i \sin\theta \tag{1}$$

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 433–440.
springerlink.com

in an $N_\theta \times N_\rho$ accumulator that represents the (θ, ρ) normal parameters plane, representing all possible lines that could pass through that pixel, where ρ is the distance to original point, θ is the line's direction. The sinusoids corresponding to collinear data points are ideally intersected at a single point in the parameter space, which leads to a peak bin in the accumulator array.

However one disadvantage of standard Hough Transform (SHT) is its computing complex, which increases linearly with resolution and its high cost of memory storage. Thus a serial of improved HT algorithms have been proposed to reduce the computational complexity, such as adaptive Hough Transform (AHT) [2], fast Hough Transform (FHT)[3], multi-resolution Hough Transform (MHT) [4], and probabilistic Hough Transform (PHT) [5].

Most remote sensing images with high resolution have very large sizes and complex background, so an interested linear target may occupy very small part in the whole image. Since the Hough Transform have only one accumulator for the whole image, Hough Transform is only efficient if the line is long enough to have a high number of votes fall in the right bin of the accumulator to make it easy to be detected. In a large scale remote sensing image with complex background a large target's edge is still not long enough to right bin to be detected in the background noise. Another disadvantage for HT is that the background noise is easy to be linked as lines, which may make the detection unreliable.

2 Algorithm

In this paper, an improved method (fig.1) based on PHT is proposed to detect linear targets in remote sensing image. The image is firstly divided into many small sub-images, and then PHT is applied to each of the sub-images, so we get many short straight lines. Then we group up the line segments with similar direction and position. Then after a projection of each group to the orthogonal direction, the most distinct parallel lines will be found. At last, such pairs of parallel lines are examined further to judge if their features match the interested targets.

Fig. 1. Procedures of our method

2.1 Dividing Image

The remote sensing image usually is very large and complex. In such image, it is unrealistic to extract line using Hough Transform directly, because background has much more voting right than the interested targets. If we prepare to detect the lines in the image, the image is firstly divided into many small sub-images with the width smaller than the interested target's length. One benefit of dividing the image is that, the thresholds of edge detection can vary in different areas. This means that the edges

whose gray gradient is high in local area yet relative low in the whole image can also be detected. Applying Hough Transform to small parts instead of the whole image can reduce the size of accumulator. Therefore the cost of memory can be saved considerably.

While dividing the image, lines are also divided. Sometimes, the line can be divided into very short segments and each segment is not long enough to be recognized as an integrated line, especially when the line crosses sub-image's corner. A continuous line in the original image may become short and discontinuous. To make the method more reliable, the neighboring sub-images should overlap each other. But the more area overlapped the more parts needed processing. Generally, every sub-image should overlap no less than 1/3 size with its neighbor parts. In this way, we can get an array of sub-images named $P_{i,j}(i \in [0, s-1],\ j \in [0, t-1])$, where *s*, *t* is the number of rows and columns of the sub-image array.

2.2 Probabilistic Hough Transform

Probabilistic Hough Transform proposed by Kiryati *et al* [5] is similar to SHT. The difference is that rather than using all *M* edge points, only a subset *m* is used. As *m*<*M*, the complexity of the voting stage is reduced from $O(M, N_\theta)$ to $O(m, N_\theta)$. Intuitively, this works because a random subset of *M* will fairly represent all features and noise based on the area they occupy in the image. Choosing a smaller value for *m* will lead to a faster algorithm but clearly it must not be so small that features can no longer be detected.

Before applying PHT, we must detect the edges in each part via Canny edge detector [7]. The Canny detector needs two thresholds, we can determinate the thresholds for one part by using the accumulated histogram $H_{i,j}(g)$ of gray gradient of $P_{i,j}$. Assuming that the edge points make up $p\%$ of pixels in every part, the high threshold for Canny detector can be set g_t satisfying:

$$\begin{cases} H_{i,j}(g_t) \le N \cdot (1-p\%) \\ H_{i,j}(g_t+1) > N \cdot (1-p\%) \end{cases} . \tag{2}$$

Where N is the number of pixels in $P_{i,j}$. And the lower threshold can be a fixed percentage of the high threshold. These two percentages can be both determined by experiments.

Then Probabilistic Hough Transforming is applied to each sub-image, we can get the number of $C_{i,j}$ line segments named $l_{i,j,k}(k \in [0, C_{i,j}])$ in sub-image $P_{i,j}$.

2.3 Detecting of Parallel Lines

A probability method in [7] is used to cluster points into lines. But in our method, instead of points, the short line segments are clustered into long parallel lines. Comparing to method used in [7], our method is less complex and can utilize the

position and direction information of the line segments. Our method is a growing method: every group starts from a line segments, and absorbs nearby collinear or parallel line segments, till no such line segment can be found in neighbor sub-images.

First we must set criteria for the combination. One can be $A_{\max}$, the maximum included angle between lines, and the other is $D_{\max}$, the maximum distance between line segments. Because $A_{\max}$ is small, the definition of distance between parallel lines can be used here to determine the distance ($D(l_1, l_2)$) between two line segments in a group.

Then, a graph G(V,E) whose vertices is all the line segments is obtained as follows:

$$V = \bigcup_{i,j} \{ l_{i,j,k} \mid k \in [0, C_{i,j}] \} \quad . \tag{3}$$

And the edges of G is the line segments in neighboring or the same sub-images that fulfill the criteria of $D_{\max}$ and $A_{\max}$. So the edges can be express as:

$$\begin{aligned} E = \{ (l_1, l_2) \mid & A(l_1, l_2) < A_{\max}, D(l_1, l_2) < D_{\max}, \\ & |i1 - i2| < R, |j1 - j2| < R, l_1 \in P_{i1,j1}, l_2 \in P_{i2,j2} \} \end{aligned} \tag{4}$$

Where R is distance of the furthest sub-images needed to be searched and can be determined by experiments.

Once the graph G is created, it is easy to traverse G to separate the line segments into unconnected sub-graphs $G_i (i = 0 \ldots N_g)$. There is no link between two sub-graphs and within any sub-graph all vertices are connected. In this way the lines parallel and close to each other are put into one group.

For the objects like airport and highway, the two long edges always have opposite direction of gray gradient. Then the line segments in each group G_i can be divided into two groups G_i^+ and G_i^-, according to the main gray gradient direction of points on line segment, and the two edges of the target should be find in two groups respectively.

After separating line segments into groups, we will search for parallel line pairs within proper distance and length range ($[L_{\min}, L_{\max}]$ and $[D_{\min}, D_{\max}]$) in each group pair G_i^+ and G_i^-. Firstly, a pair of one-dimension histograms $h_i^+(\rho)$ and $h_i^-(\rho)$ of the line segments' ρ (distance to original point) will be computed respectively for G_i^+ and G_i^-. The ρ can be achieved by putting the coordinate of line segment's midpoint and G_i's mean direction θ_i into eq. (1), and the voting power of each segments will be the line segment's length.

If there is any pair of peaks $h_i^+(\rho_{peak}^+)$, $h_i^-(\rho_{peak}^-)$ appearing at ρ_{peak}^+ and ρ_{peak}^- respectively fulfill the criteria:

$$\begin{cases} h_i^+(\rho_{peak}^+) > L_{\min} \\ h_i^-(\rho_{peak}^-) > L_{\min} \\ \left|\rho_{peak}^+ - \rho_{peak}^-\right| \in [D_{\min}, D_{\max}] \end{cases}, \tag{5}$$

the line segments in G_i^+ whose distances to line L^+:

$$\rho = x\cos\theta_i + y\sin\theta_i \tag{6}$$

are smaller than certain threshold should be linked as candidate edges of the detected target. The distance between line segment AB and L^+ can be defined as the average of the distances of all points on the line segment to the straight line:

$$d\left(AB, L^+\right) = \begin{cases} \left(|d_1| + |d_2|\right)/2, d_1 \cdot d_2 \geq 0 \\ \left(d_1^2 + d_2^2\right)/2(|d_1| + |d_2|), d_1 \cdot d_2 < 0 \end{cases}, \tag{7}$$

where

$$\begin{cases} d_1 = x_1\cos\theta_i + y_1\sin\theta_i - \rho_{peak}^+ \\ d_2 = x_2\cos\theta_i + y_2\sin\theta_i - \rho_{peak}^+ \end{cases}, \tag{8}$$

(x_1, y_1), (x_2, y_2) are AB's end points. And the line segments in G_i^- are linked similarly.

2.4 Refine the Results

After above steps, usually the target can be located, but there are some other features can be utilized to make the detection more accurate.

Taking runway for example, the features such as brighter, smoother surface than background, and high aspect ratio can be used to improve the detection. So the parallel lines can be accepted as runway edges, if an area between the line pairs satisfies the flowing criteria:

1. brighter than its neighbor area;
2. length is proper for a runway;
3. aspect ratio is high enough.

3 Experiments

In our experiment, we use a satellite remote sensing image (showed in fig.2(a)) containing the runway of an airport with a resolution of 2000*2000 (only the part containing runway is showed). As the airport is located in suburbs of Beijing city, thus there lots of high ways, rivers and farmland, which all contain complex and sharp

edges. What is more, the airport only takes up a relative small area in the whole image, while the complex background occupies most of the space in the image. These all add to the difficulties of detection.

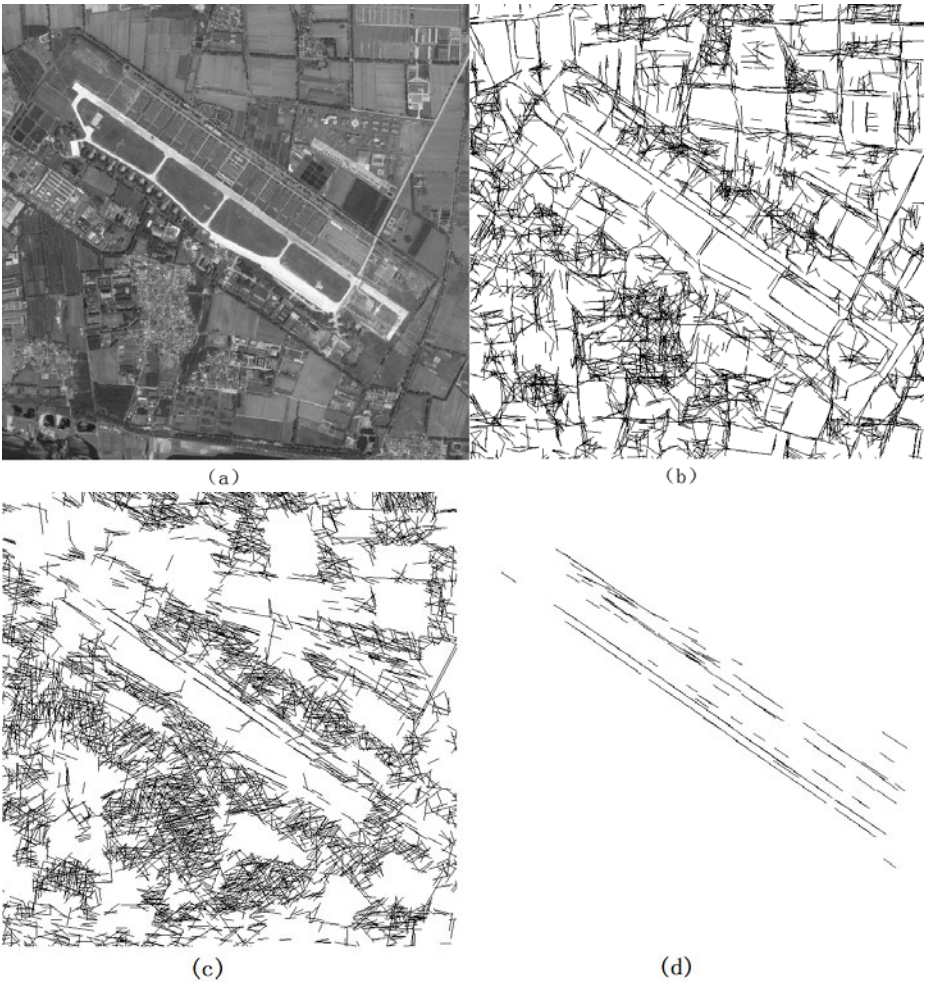


Fig. 2. (a) The original gray scale image. (b) shows the lines segments found by applying PHT after the image is divided into parts and (c) shows the result of PHT applied to the whole image. (d) A group of line segments containing runway.

Fig.2(b),(c) make a comparison of result of PHT applied in the whole image and in the sub-images. Obviously, when applying Hough Transform to the whole image (fig.2 (c)), only part of the long edges of the runway extracted and they are discontinuous, most of the short edges cannot be detected, and there are large number of meaningless short lines. While in fig.2 (b), when the image is divided in to sub-images, before applying PHT, the line segments mesh the edges well, and the lines are

continuous. What is more, most of the short edges also can be found. Although in some place, there are still some useless line segments, because we set very low threshold to find as many edges as possible. And after clustering, most the useless line can be most eliminated as showed in fig.2(d). Most line segments belonging to the background cannot find many neighbors with similar direction, so such group can be ignored directly. In this example, 22280 groups were found, but most groups only contain one or two line segments, only 124 groups contain more than 15 line segments. The group in this image contains the runway has 168 segments. That means most line segments can be eliminated in this step, and the left procedure will be very fast.

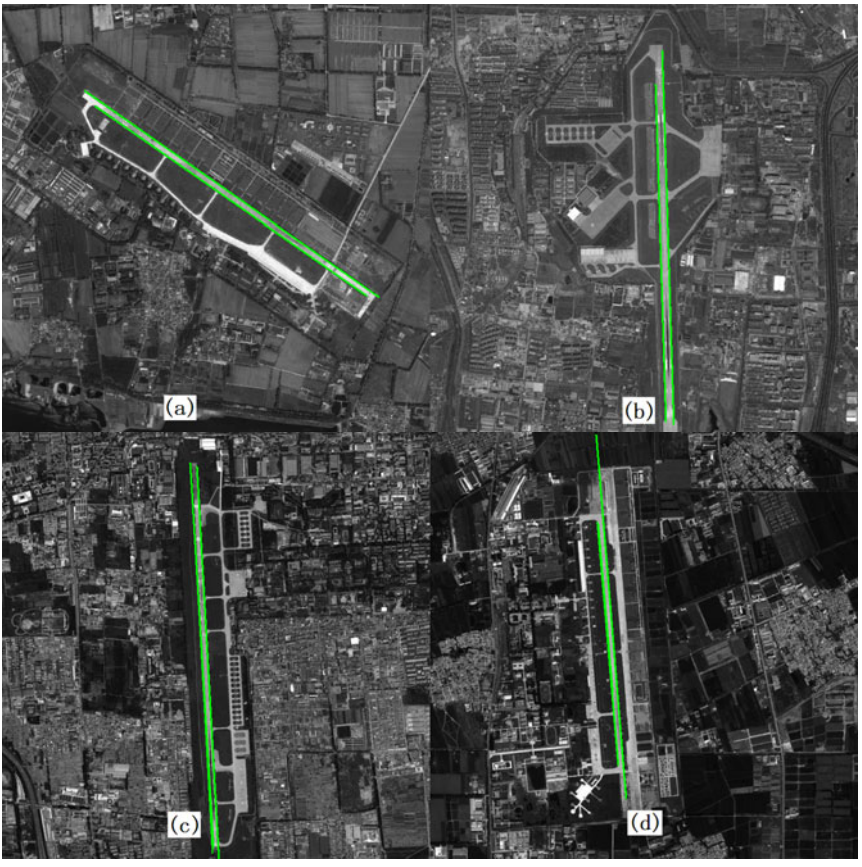


Fig. 3. The result of our method in four images

And after use the criteria for runway, which is mentioned above, we get the correct final result (fig.3 (a)). When this program is applied to three other images it also works fine (fig.3 (b), (c) (d)), and no false alarm is found, while all of the four images' resolutions are 2000*2000(only the part containing runway is showed).

4 Conclusion and Future Work

The experiments show that our method is effective and reliable for large linear target detection in remote sensing image. This method retains the robustness of Hough Transform, and by separating the image to small sub-images, a smaller accumulator is used, the memory usage and computing complex is reduced greatly. Also the incorporating of the large linear target's unique features, the reliability of the method is improved. What is more, this method is flexible, can be easily adjusted to be applied to various kinds of targets with straight edges.

Acknowledgments. This work is supported by National Natural Science Foundation of China (NSFC Grant No. 60702017) and Science and Technology Foundation of National Key Laboratory of science and technology on low light level night vision (STFNKL Grant No. J20110502).

References

1. Duda, R.O., Hart, P.E.: Use of The Hough Transformation to Detect Lines and Curves in Pictures. Commun. ACM 15, 11–15 (1972)
2. Illingworth, J., Kittler, J.: The Adaptive Hough Transform. IEEE Trans. Pattern Anal. Mach. Intell. 9, 690–698 (1987)
3. Li, H., Lavin, M.A., Master, R.J.L.: Fast Hough Transform: A Hierarchical Approach. Comput. Vision Graph. Image Process. 36, 139–161 (1986)
4. Atiquzzaman, M.: Multiresolution Hough Transform-An Efficient Method of Detecting Patterns in Images. IEEE Trans. Pattern Anal. Mach. Intell. 14, 1090–1095 (1992)
5. Kiryati, N., Eldar, Y., Bruckstein, A.M.: A Probabilistic Hough Transform. Pattern Recogn. 24, 303–316 (1991)
6. Barni, M., Cappellini, V., Mecocci, A., Paoli, A.: Unsupervised Detection of Straight Lines Through Possibilistic Clustering. In: ICIP 1996, IEEE International Conference on Image Processing, pp. 963–966. IEEE Press, Lausanne (1996)
7. Canny, J.F.: A Computational Approach to Edge Detection. IEEE Transactions on Patern Analysis and Machine Intelligence 8, 679–698 (1986)

Towards a Dedicated ASIP for AES Implementation

Yibin Li, Zhiping Jia, and Renhai Chen

Computer Science and Technology Department,
Shandong University,
Shunhua Road No.1500, 250101,
Jinan, P.R.China
{liyibing,jiazp,chenrh}@sdu.edu.cn

Abstract. Advanced Encryption Standard (AES) is a symmetric-key encryption standard. It is widely adapted to encryption/decryption sensitive information. As the speed of network transmission keeps growing and embedded applications such as Wireless Sensor Network (WSN) have been mature in recent years, to facilitate higher throughout and less cost, dedicated accelerator is demanded with less power consumption. Meanwhile, in EDA field, Electronic System level (ESL) method has become a practical development method even for commercial product. In this paper, a commercial ESL tool based on Lisa language is used. By implementing this development flow, several specific instructions are identified and implemented in a AES processor. For AES algorithm, with 40% memory saving, it is shown 50% performance improvement over ARM ISA.

1 Introduction

Many applications, the necessary protection of sensitive information is critical. Therefore, encryption and decryption have been widely implemented to protect sensitive data [1]. Though many algorithms have been adopted and tested, AES is one of the most used symmetric-key algorithm. To raise the performance of AES execution, the hardware realization of AES becomes an interested research topic. Among past related works, hardwired implementation of AES has been focused [2,3]. To carry more flexibility in AES engine, the application specific instruction processor (ASIP) also has been studied [4]. In recent years, Electronic System Level (ESL) has emerged as a main stream design approach. Therefore, it become author's interest to explore ASIP solution for AES using ESL based method.

2 AES Background

For the initial profiling of AES application, Simplescalar is used to identify the specific instructions. Through experiments, it is realized that the S-box occupied the major dynamic complexity (Fig.1). Therefore, a dedicated instruction set for S-box calculation is the key to accelerate AES. Such dedicated instruction set also can be used to raise the performance of other encryption algorithms as indicated in Fig.1.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 441–444.
springerlink.com

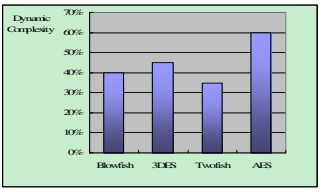


Fig. 1. The dynamic complexity for S-Box operation in different encryption algorithms

3 AES Specific Instructions and Processor Architecture

To explore the application specific instruction, a commercial ESL based ASIP tool is adopted. Firstly, the identified instructions are modeled using LISA language. Through simulation, the cycle count information are collected and analyzed. Then, the number of opcode and their bits number are also found. Finally, the format of each instructions are identified as Table.1.

Table 1. The four AES specific instructions

Instruction	Format											
ifand	opcode		src1		src2		xor_src1		xor_src2		xxxxxx	
	31	26	25	21	20	16	15	11	10	6	5	0
getbit	opcode		dest		src2		bitpos		xxxxxxxxxx			
	31	26	25	21	20	16	15	11	10			0
xor_5	opcode		src1		src2		src3		src4		src5		dest		x
	31	26	25	22	21	18	17	14	13	10	9	6	5	1	0
matrixpos	opcode		dest		src1		src2		src3		src4		x
	31	26	25	21	20	16	15	11	10	6	5	1	0

A. Ifand instruction

In MixColumn operation, Multiplication in GF(28) costs large part of computation. The worst case is the execution of 256 time iterations. Ifand instruction is defined to calculate multiplication in GF(28).

B. Getbit Instruction

In affine transformation, bit level manipulation is needed frequently. Therefore, a specific instruction named getbit is defined in the ISA.

C. Xor_5 Instruction

After affine transformation, the xor operation takes place. It is defined as in below formula

$$y[i] = x[i] \text{ xor } x[(i+4)\text{mod}8] \text{ xor } x[(i+5)\text{mod}8] \text{ xor } x[(i+6)\text{mod}8] \text{ xor } x[(i+7)\text{mod}8]$$

For accelerating this operation, one custom instruction named xor_5 is designed. As described in [], xor_5 instruction has five operands. Each of them has only four bits, as a result, only register 0 to 15 can be addressed as operand in this instruction.

D. Matrixpos Instruction

In mix-column, sub-byte operation is dominant. The c code as

x=sbox[i][j]

Using proposed instruction, the functionality can be implemented by matrixpos instruction as

move r1, sbox // sbox represent the base address of sbox array matrixpos r1 = r1, i, j,n // n represent the size of sbox

To implement these AES specific instructions, a 32 bits processor is used to accommodate the identified AES specific instructions. As described in Fig. 2, the processor contains three stages for pipelining. Mainly, it contains 32K Data RAM, 32K Prog_RAM and a simple 3 stages pipeline with 34 registers. Both Data_RAM and Prog_RAM can be addressed from 0x0000 to 0x7FFF. Apart from 32 general purpose registers, one Fetch Program Register (FPR) and Stack Pointer Register (SPR) are also used. The whole instruction set contains 25 instructions including four AES specific instructions which are described in Table 1.

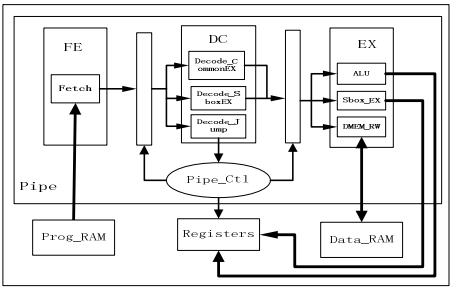


Fig. 2. The architecture of AES processor

4 Implementation Methodology

The whole design flow can be illustrated in Fig. 3. In this ESL based design flow, the Verilog description of AES processor and associated C-compiler are generated automatically. The Verilog model of processor is validated using Modelsim. After that, the processor design is synthesized using Xilinx and related data is collected. From that, the design space of processor can be explored.

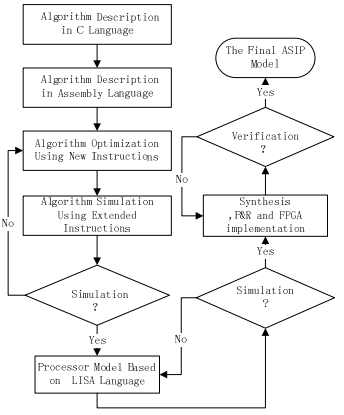


Fig. 3. ESL based design flow

5 Result and Discussion

Through recursive simulations, compared with widely used ARM ISA, it can be found that 50% cycle count (fig.8) can be saved with 40% less program memory utilization. When the AES processor is synthesizes, targeted to Virtex4, it is shown that the LUT count only increase from 17035 to 19004 which represent 11.5% more resource utilization for additional AES specific instructions. To further optimize our design, a Data Flow Graph (DFG) method is considered in identifying new AES specific instructions. Also, power and hot issues are needed to be considered.

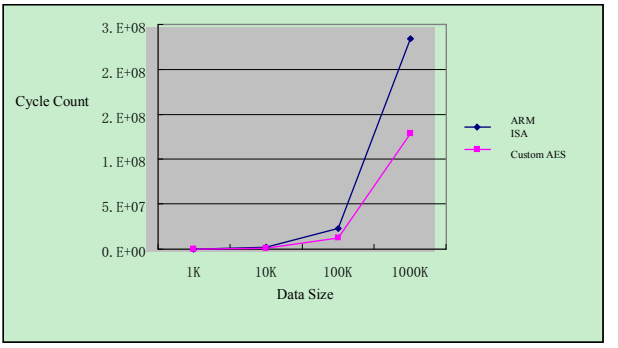


Fig. 4. Performance Improvement

Acknowledgments. This research is supported by Shandong Province Natural Science Foundation (Y2008G13), Shandong Province Research Program (2010GcG20101), Independent Innovation Foundation of Shandong University (IIFSDU 2010TB020), and IT Development Fund of Shandong Province (2010R1501).

References

1. Aaraj, N., Raghunathan, A., Jha, N.K.: Analysis and design of a hardware/software trusted platform module for embedded systems. ACM Transactions on Embedded Computing Systems (TECS) 8, 1–31 (2008), doi:10.1145/1457246.1457254
2. Li, Y., Sakiyama, K., Batina, L., Nakatsu, D., Ohta, K.: Power Variance Analysis breaks a masked ASIC implementation of AES. In: Design, Automation & Test in Europe Conference & Exhibition (DATE 2010), pp. 1059–1064 (March 2010)
3. Rachh, R.R., Anami, B.S., Ananda Mohan, P.V.: Efficient implementations of S-box and inverse S-box for AES algorithm. In: TENCON 2009 - 2009 IEEE Region 10 Conference, pp. 1–6 (January 2009), doi:10.1109/TENCON.2009.5395837
4. Scharwaechter, H., Kammler, D., Wieferink, A., Hohenauer, M., Karuri, K., Ceng, J., Leupers, R., Ascheid, G., Meyr, H.: ASIP architecture exploration for efficient IPSec encryption: A case study. ACM Transactions on Embedded Computing Systems (TECS) 6, 12-es (2007), doi:10.1145/1234675.1234679

Compressive Sensing of Image Reconstruction Based on Shearlet Transform

Fangyi Wang[1], Shengqian Wang[2], Xin Hu[1], and Chengzhi Deng[2]

[1] Jiangxi Science & Technology Normal University,
Nanchang, China
[2] Department of Information Engineering,
Nanchang Institute of Technology,
Nanchang, China
hmily1024@126.com

Abstract. Compressive sensing (CS) is a radical new way of sampling signal at a sub-Nyquist rate, acquires a signal of interest indirectly by correcting a very limited number of its "projections", and the signal can be exactly reconstructed from these "projections", this new signal acquisition paradigm has revolutionized the way digital data are traditionally acquired. Generally, researchers always use orthogonal wavelet as sparse basis, but it fails to provide an optimal sparse representation for images that contain texture details. In this paper, we use shearlet which is a new directional multiresolution transform, can efficiently represent the directional information of images, meanwhile, using the RecPF of [4] as the reconstruction algorithm. The experimental results indicate that the quality of reconstructed image is improved and obtain better performance of PSNR.

Keywords: Compressive sensing, wavelet, shearlet, RecPF, PSNR.

1 Introduction

The Shannon/Nyquist sampling theorem states that a signal can be reconstructed perfectly from its samples, only if it was sampled at a rate at least twice the highest frequency of the signal (Nyquist 1928; Shannon 1949). This rate is known as the Shannon/Nyquist rate of that signal, but for many signals, such as images, the Nyquist rate can be very high. This may result in acquiring a very large number of samples, which must be compressed in order to store or transmit them, as well as confronting with a high requirement on the sampling equipment needed to sample the signal. In recent years, the new theory of compressive sensing is introduced to handle these problems independently by Candès [1] and Donoho [2]. Compressive sensing is a radical new way of sampling signals at a sub-Nyquist rate, capturing a small number of random linear projections than evenly Shannon/Nyquist sampling, but still effect perfect reconstruction for a large class of signals from their random projections via appropriate reconstruction algorithm.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 445–451.
springerlink.com

1.1 Signal Sparse Representation

Compressive sensing in general is based upon the fact that signals are sparse or compressible in the sense that they have sparse or approximately sparse representations when expressed in an appropriate basis or dictionary.

Let $u \in R^N$ be an unknown signal which itself may or may not be sparse in the canonical basis, but it is very likely that it will look quite sparse in another domain with a proper basis, without loss of generality, two-dimensional images or other higher dimensional data can be vectorized into one-dimensional vectors. The signal is sparse or compressible means that the number of salient features hidden in it is much less than its length (for two-dimensional images, is resolution), and the signal sparse representation is:

$$u = \sum_{i=1}^{N} \psi_i x_i = \Psi x \quad . \tag{1}$$

where $\Psi = [\psi_1, \psi_2, \ldots\ldots \psi_N] \in R^{N \times N}$ be an orthogonal basis of R^N, $x \in R^N$ is the vector of inner products $x_i = \langle u, \psi_i \rangle$. Typically, we say that u is in the spatial domain if it is a spatially dependent signal, such as an image and we say that x is in the Ψ domain. If the $\|x\|_0 = k$, the number of nonzeros in x, we say that u is K -sparse under Ψ, and we are interested in $k << N$.

1.2 Compressive Sensing

The core of compressive sensing is the linear measurement. We measure the signal u by sampling it with respect to a measurement matrix Φ with $M \times N$, and the M satisfies $K < M << N$, the measurement vector $b \in R^M$ as M linear measurements of u by:

$$b = \Phi u \, . \tag{2}$$

substituting $u = \Psi x$ into equation (2), we obtain:

$$b = \Phi u = Ax, \quad A = \Phi \Psi \in R^{M \times N} \quad . \tag{3}$$

our goal is to reconstruct the signal u from the measurements $b = Ax$, because $M \ll N$ which means that the number of equations less than the unknowns', it is seems to be an ill-posed inverse problem. Compressive sensing theory has proved that x can be reconstructed (and thus u by $u = \Psi x$) exactly from b as long as x is sparse or compressible and measurement matrix Φ satisfies restricted isometry property (RIP) condition. While it is difficult to verifying the RIP condition, it has been proved that this property holds with high probability for some matrices, such as independent and identical distributed (iid) Gaussian, Bernoulli, and partial Fourier matrices, we will be focus on partial Fourier matrix in this paper.

With the prior knowledge of sparsity in x, we can recover x by using optimization to search for the sparsest signal that agrees with the M linear measurements in b:

$$\min_x \left\{ \|x\|_0 : Ax = b \right\}. \tag{4}$$

Unfortunately, solving the l_0 minimisation programme is an NP-complete problem and impractical for nearly all real applications. A common way to solve this problem is using the l_1 norm as a substitute in the l_0 minimisation:

$$\min_x \left\{ \|x\|_1 : Ax = b \right\}. \tag{5}$$

It has been shown that l_0 norm and l_1 norm yield an equivalent solution with very high probability under some desirable conditions (see [6]), so model (4) is changed into the tractable model (5). When Φ is Gaussian random or partial Fourier, the number of measurements sufficient for exact recovery x from b is $M = O(K \log(N/K))$ and $M = O(K \log N)$, respectively ([3]). The number of measurements M is larger than K but still much smaller than N. Since Ψ is a known basis, once x is recovered, then u by $u = \Psi x$.

In consideration of the existence of noise, we relax model (5) to be:

$$\min_x \left\{ \|x\|_1 : \|Ax - b\|_2 \le \sigma \right\}. \tag{6}$$

or Lagrangian version:

$$\min_x \|x\|_1 + \lambda \|Ax - b\|_2^2. \tag{7}$$

where $\sigma > 0$ and $\lambda > 0$ are parameters. From optimization theory, (6) and (7) share common solutions.

2 Reconstruction Algorithms

In [4], Yang *et al.* introduce the RecPF algorithm which jointly minimizes the total variation, l_1 norm , and a least squares measure , is able to solve the CS reconstruction problem in the special case where Φ is a partial Fourier matrix, and $\bar{u} \in R^{n \times n}$ represents a two-dimensional grayscale digital image with N pixels($N = n \times n$). First, obtaining partial frequency measurements of $\bar{u}$:

$$f_p = F_p \bar{u} + \omega. \tag{8}$$

where $F_p \in C^{M\times N}$ represents a partial discrete Fourier transform, serving as measurement matrix Φ in CS, $\omega \in C^M$ denotes random noise. The RecPF algorithm reconstructs $\bar{u}$ from f_p via solving the problem as follows:

$$\min_u \sum_i \left\|D_i u\right\|_2 + \tau \sum_i \left|\psi_i^{\mathrm{T}} u\right| + (\lambda/2)\cdot \left\|F_p u - f_p\right\|_2^2 . \tag{9}$$

where $\sum_i$ is taken over all pixels, $\sum_i \left\|D_i u\right\|_2$ represents a discretization of the total variation (TV) of u, $\sum_i \left|\psi_i^{\mathrm{T}} u\right|$ is the l_1 norm of the representation of u under a wavelet basis Ψ, $\tau, \lambda > 0$ are balance regularization parameters.

The key idea of RecPF is to recast (8) into the following minimization problem:

$$\begin{aligned} &\min_{w,z,u} \sum_i \phi_2(w_i, D_i u) + \tau \sum_i \phi_1(z_i, \psi_i^T u) + \lambda\theta\left(u, f_p\right) \\ &\qquad s.t \quad w_i = D_i u,\ z_i = \psi_i^T u,\ \forall i \end{aligned} \tag{10}$$

where $\phi_1(s,t) = |s| + (\beta/2)\cdot|s-t|^2$; $\phi_2(s,t) = \|s\| + (\beta/2)\cdot\|s-t\|^2$ $s, t \in R^2$, $\theta\left(u, f_p\right) = (1/2)\cdot\left\|F_p u - f_p\right\|_2^2$. $D_i u$ represent the horizontal and vertical local finite differences of u at pixel i, $D^{(1)}, D^{(2)} \in R^{N\times N}$ denotes the horizontal and vertical global finite difference operators, respectively. Then, using an alternating minimization algorithm to solve model (10).

We define the relative error as :

$$\text{Rel.Err} = \frac{\left\|u - \bar{u}\right\|_2}{\left\|\bar{u}\right\|_2} . \tag{11}$$

where $\bar{u}$ and u are original and recovered images, respectively.

3 Numerical Experiments

In this paper, we select 512*512 standard grayscale image lena for testing, shearlet transform replaces wavelet transform of RecPF, using radial sampling in the Fourier domain and the number of radial lines is in direct proportion to the number of measurements M. The sampling ratio r is defined to be $r = M/N$.

Figure 1 shows 24 radial lines in the Fourier domain:

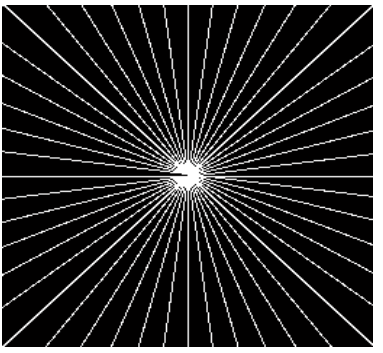


Fig. 1. Fourier domain sampling positions (highlighted in white) with 24 radial lines

The original and reconstructed images are shown in **Figure 2** and **Figure 3** with different sampling ratio, using wavelet transform and shearlet transform, respectively.

(a) (b) (c) (d)

(a) original image; (b) $r = 8.79\%$; (c) $r = 12.92\%$; (d) $r = 16.94\%$

Fig. 2. Recovered images with different sampling ratio and wavelet transform

(e) (f) (g) (h)

(e) original image; (f) $r = 8.79\%$; (g) $r = 12.92\%$; (h) $r = 16.94\%$

Fig. 3. Recovered images with different sampling ratio and shearlet transform

As can be seen from **Figures 2** and **3,** image quality of recovered images by using shearlet transform is better than using wavelet transform.

We record relative error and peak signal to noise ratio (PSNR) at the different sampling ratios in **Table 1**.

Table 1. Numerical results at different sampling ratios

Sapling ratio r	Wavelet transform		Shearlet transform	
	Rel.Err	PSNR(dB)	Rel.Err	PSNR(dB)
8.79%	6.99%	28.8	6.65%	29.2
12.92%	5.06%	31.6	4.64%	32.3
16.94%	4.04%	33.5	3.61%	34.5

From the comparison in **Table 1**, it can be seen that the quality of image reconstruction by using shearlet transform, are better than those using wavelet transform at each of the same sampling ratio.

4 Conclusions

In this paper, we use shearlet transform to replace wavelet transform, and recovering images by using both of them for compressive sensing, at different sampling ratios. From the results of experiments, because of shearlet can efficiently represent the directional information of images which is the weakness of wavelet, the visual effect and PSNR by using shearlet transform are better than wavelet transform, but both transformations give faithful images at the sampling ratio of 16.94%. The shearlet has more superiorities than wavelet in terms of signal sparse representation, so it's can be extended to other compressive sensing reconstruction algorithm.

References

1. Candès, E., Romberg, J., Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory 52(2), 489–509 (2006)
2. Donoho, D.: Compressed sensing. IEEE Transactions on Information Theory 52(4), 1289–1306 (2006)
3. Candès, E., Tao, T.: Near optimal signal recovery fromran-dom projections: universal encoding strategies. IEEE Transactions on Information Theory 52(1), 5406–5425 (2006)
4. Yang, J., Zhang, Y., Yin, W.: A fast alternating direction method for TVL1-L2 signal reconstruction from partial fourier data. IEEE Journal of Selected Topics in Signal Processing, Special Issue on Compressed Sensing 4(2), 288–297 (2010)

5. Wang, Y., Yang, J., Yin, W., Zhang, Y.: A new alternating minimization algorithm for total variation image reconstruction. SIAM Journal on Imaging Sciences 1(3), 248–272 (2008)
6. Guo, K., Labate, D.: Optimally sparse multidimensional representation using shearlets. SIAM J. Math. Anal. 39(1), 298–318 (2007)
7. Easley, G., Labate, D., Lim, W.Q.: Sparse directional image representations using the discrete shearlet transform. Appl. Comput. Harmon. Anal. 25(1), 25–46 (2008)
8. Ma, S., Yin, W., Zhang, Y., Chakraborty, A.: An efficient algorithm for compressed MR imaging using total variation and wavelets. In: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2008), pp. 1–8 (2008)
9. Cai, J.-F., Osher, S., Shen, Z.-W.: Linearized Bregman iterations for compressed sensing. Mathematics of Computation (78, 267), 1515–1536 (2009)
10. Geman, D., Yang, C.: Nonlinear image recovery with half-quadratic regularization. IEEE Transactions on Image Processing 4, 932–946 (1995)
11. Lysaker, O.M., Tai, X.-C.: Iterative image restoration combining total variation minimization and a second-order functional. International Journal of Computer Vision 66, 5–18 (2006)
12. ShearLab-1.1, `http://www.shearlab.org/index_software.html`

Simulation and Optimization of Logistics Collaborative Operation Based on Flexsim

Xu Linwei* and Zhong Xing Li

CISDI Chongqing Iron & Steelmaking Plant Integration Co., Ltd.
CAE Technology Center, Chongqing, 400013
linweixu168@163.com

Abstract. This paper based on analyzing process of logistics Collaborative operation, with coordinated task sequences of Flexsim and optimization tools OptQuest of Flexsim, to provide a new way to solve allocation of resources of collaborative operation. A simulation model is established for a typical storage outbound collaborative operation. By simulation testing, we can know Utilization of all kinds of resources; we also can use optimization tools to optimize the allocation of resources of collaborative operation, the optimal allocation of resources can provide decision for the manager.

Keyword: logistics cooperation, Optimization, Simulation, Flexsim.

1 Introduction

System simulation as an effective tool for process re-engineering and resources optimization in foreign countries has been widely used in business management. Using the simulation software to simulate and analyze the flow of service system for decision-making becomes one of the research hot spots. Flexsim is a new commercial discrete event system simulation software developed by Flexsim Software Production company in the United States, which combines three-dimensional computer image processing technology, discrete system simulation, artificial intelligence, data processing technology as a whole[1-2].

Most of the actual process of logistics operations is a collaboration of all kind of resources. Such as container ports: quayside container cranes, yard crane, container trailer, and the cooperative scheduling of the operators and so on. Completing Collaborative task requires the operation of complex coordination among two or more resources. A typical example of storage collaboration is that: a team of three operators sharing two forklifts, one operation need a forklift and an operator. Forklift operator goes to the location of the forklift, and drives forklift to the loading location, and then loads the goods to the unloading location, finally the operator drives the forklift to the parking location and get down. This typical job scheduling optimization process is very difficult because it handles two or more different resources scheduling in a coordinated way. At present paper about Collaborative process of logistics simulation and optimization is rare, but modern logistics operations of loading and unloading

* His main research area is simulation and optimization of logistics systems.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 453–457.
springerlink.com

will be completed by the coordination of various resources, and the material handling operation is open, complex , Random, so the traditional methods is hard to described the rule. Simulation and optimization technology in this area shows great potential. Flexsim with powerful ability to control the task sequence can simulate all kinds of logistics operations processes [3-4]. Based on the analysis of the typical warehouse outbound process of collaborative handling operations, this paper uses Flexsim to establish the simulation model, and uses optimization tools OptQuest to optimize the allocation of resources, and then obtains the best allocation of resources of the operating system.

2 The Analysis and Modeling of Storage Collaborative Operation Process

2.1 The Analysis of Storage Collaborative Operation

A typical enterprise storage assignment is used in this paper to do the research. The outbound process of this storage is shown in the Figure 1.

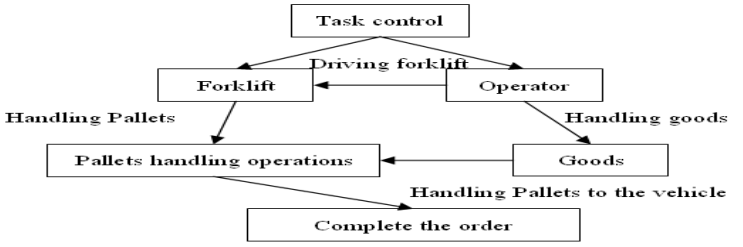


Fig. 1. Outbound process of the storage

The outbound process of this storage is not very complicated, but because the Order flow and the Operation time is random, and the collaborative work between operator and forklift. So optimization the allocation of resources is a difficult problem.

2.2 The Simulation of Storage Collaborative Operation

The main process of forklift and operator collaborative work is illustrated in Figure 2.

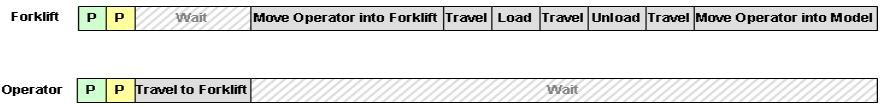


Fig. 2. Sequence of collaborative operation in Flexsim

Then we can establish model according to layout and the operation procedure, input the statistical data .for example, the order flow obey the index distribution with mean 60. And set logical codes of each Flexsim modules, and create global arrays to obtain statistical information. After model checking (Mainly inspecting logic errors of the model) and model testing (examining whether simulation model is coincident with the fact), finally, we can establish the simulation model as follows in Figure 3. The simulation time is set to 4800 time units (20 time units equate to one minute in reality).

Fig. 3. Simulation model of collaborative operation Flexsim

3 Sensitivity Analysis and Optimization of Resource Allocation

3.1 Sensitivity Analysis of Resource Allocation

Flexsim software built-in experiment analyzer and the characteristic of tree structure can be used to carry out sensitivity analysis on any variable. According to the actual situation of the enterprise, the maximal number of operators is 8, and the maximal number of forklift is 3. On this basis, we can carry out the sensitivity analysis separately on number of operator and forklift; we put order quantities completed per time unit, order fill rate and the task cycle to finish one order as objective function.

When analyzing the sensitivity of forklift, the operator number is set to the maximum number.

Based on the previous settings, we can do the sensitivity analysis and get the result. Influence of the forklift number on task cycle, order fill rates and order quantities is illustrated in Figure4.

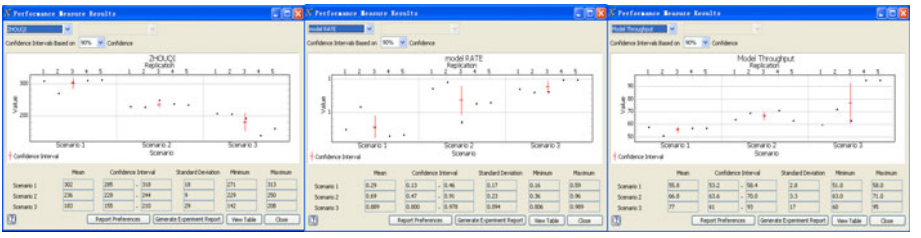


Fig. 4. Influence of the forklift number on task cycle, order fill rates, order quantities

From the comparative analysis on figure 4, we can draw a conclusion that the number of forklifts has great effect on the three index parameters, in other words, With the increase of the forklifts number, the task cycle will decrease, order fill rates will improve and the order quantity will increase. Number changing from 1 to 2 has a greater effect on the three index parameters than the changing from 2 to 3. In the consideration of economical efficiency, the best answer is 2 forklifts.

Setting the forklift number as 2, we can do sensitivity analysis on operator number.

Based on the previous Settings, we can get the result of sensitivity analysis. Influence of the operator number on task cycle, order fill rates and order quantities is illustrated in Figure5.

Fig. 5. Influence of the operator number on task cycle, order fill rates, order quantities

From the comparative analysis on figure 5,we can draw a conclusion that the influence on the three index parameters change slightly as the operator number change when the forklift number was set as 2. When the operators is changed to 3 or 4, which has little influence on the three index parameters. But as a practical matter, it's very hard for the operators to drive the forklift as well as deliver goods. So, the operator number can be set as 4.

Flexsim can be used to do sensitivity analysis on two or combination of more other variables. However, the situation of resource combination is very complicated. In order to get the best results, we can use the Flexsim optimization function.

3.2 Optimization of Resource Allocation

Optimization module of the Flexsim is based on the software OptQuest. OptQuest use the Meta heuristic approach (A set of optimization method: Genetic algorithm, the simulation processing, Tuba search, Scatter search and other scattered hybrid method) to structure the optimization scheme. In fact, OptQuest can choose one method to get the best answer as soon as possible according to the character of the model. In order to perform optimization, maximizing or minimizing target variables need to be defined, the target function here refers to maximum number of orders and order fill rates. The task cycle order, waiting time of order, the number of forklifts and operators are combined as a punishment function.

Running the model, we can get optimization results as the Figure 6. The number of forklifts in the optimal project is 2 and the number of operators is 3 which are identical with the answer of the previous sensitivity analysis. By comparing to the real production data, the result of the optimization is in accordance with actual production requirements.

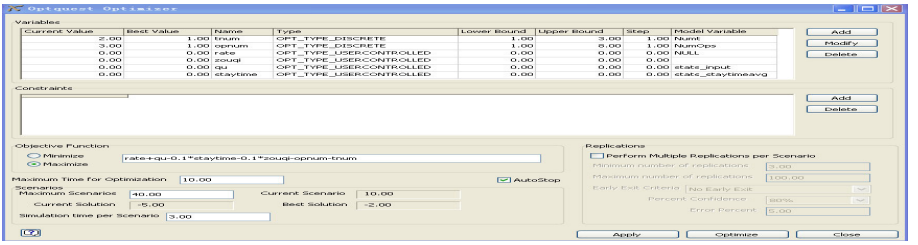


Fig. 6. Optimization results

4 Conclusions

This paper indicates that by using discrete system simulation and Flexsim simulation software, we can build a simulation model of logistics collaborative work. The simulation result can be got easily by carrying out a sensitivity analysis. And the best Resource allocation can be got by using the optimization function of the simulation software. The research also shows that three-dimensional animation simulation software is helpful for the manager to analysis process and test models.

References

1. Sun, H., Xu, L.: Optimization of Scheduling Problem for Auto Mixed Model Assembly Line. In: Proceedings of the First International Workshop on Intelligent Systems and Applications, vol. 3, pp. 2015–2017 (May 2009)
2. Sun, H., Xu, L.: Simulation and Optimization for Noshery Service System. In: Proceedings of the First International Symposium on Information Engineering and Electronic Commerce, pp. 721–723 (May 2009)
3. Hopp, W.J., Spearman, M.L.: Factory Physics—Foundations of Manufacturing Management. Tsinghua University Press, Beijing (2002)
4. Laguna, M., Marklund, J.: Business Processing Modeling, Simulation and Design. Prentice Hall, New Jersey (2004)

The Research on Inland River Ship Rescue Methods

Zongxiang Yuan[1], Zichao Huang[2], Xiumin Chu[1], and Deqi Huang[1]

[1] Engineering Research Center for Transportation Safety (Ministry of Education), Wuhan University of Technology, 1040# HePing Avenue, Wuchang District, Wuhan City, Hubei Province, 430063
[2] School of Logistics Engineering, Wuhan University of Technology, 1040# HePing Avenue, Wuchang District, Wuhan City, Hubei Province, 430063
sunxing@whut.edu.cn, {hzc_dragon,dqhuang}@126.com, chuxm@mail.whut.edu.cn

Abstract. The Usage of computer systems and communication technology to improve the efficiency of the decision and command is the important research in the current maritime emergency rescue. By classifying inland river ship rescue methods and application conditions, a unified rescue methodological framework is put up, thus the reasonable data structure of ship rescue method basic elements is designed, which includes the searching strength, rescue skills, rescue applicable condition and support model and so on, the ship rescue method is organized and electronized based on the unified framework, the ship rescue method base is built. , put forward the digital represented method of ship rescue is put forward combined with the ontological method.

Keywords: Maritime traffic safety, emergent rescue, digital, decision support.

1 Introduction

As the rapid development of computer and communication technology, the information technology for emergency treatment is becoming a research focus in accident emergency rescue field. The emergency communication system, monitoring and early warning system, emergency response system and emergency information management system are developed and constructed, and some of them has been used in the practical application [1-3]. For the purpose of timely and accurate emergency information interaction between emergency management and emergency response agencies, and aiming at the information exchange and sharing, some countries began to research on system development, also established a number of emergency information exchange standards, such as CAP(Common Alerting Protocol) [4], EDXL(Emergency Data Exchange Language) [5], IEEE1512[6] and so on. Therefore, in foreign countries, water safety supervision and accident disposal are inclined to informatization and decision-making intellectualization. The current foreign research on emergency response system is primarily focus on sea accident, while inland river accident emergency system technology research is relatively few [7]. According to both foreign advanced experience and the related technology of China's maritime accident emergency system about inland river, and based on the inland waterway traffic

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 459–468.
springerlink.com

characteristics and accident emergency conditions of inland river, the rescue method digital research is developed. which includes ship rescue method and its application conditions research, the built(n) of the unity of the rescue method framework•the organized and electronized of ship rescue method based on a unified framework•the building of ship rescue method base, the design of reasonable data structure, and the representation method of ship rescue.

2 Emergent Rescue Demand

After the Ship accident occurred, the drowning people must be rescued in time, besides, the effectively rescue for the ship in distress is needed, in order to minimize property damage [8]. According to the definition of accident type and cause in Statistical Measure of Water Transport Accidents (5th Order Issued by the Ministry of Communications P.R.C.in2002), the water transport accidents rescue mainly involves 16 kinds of incident ship rescue, they are collision, grounding, Striking on a Rock, fire, explosion, hull damage, hull flooding, wind hazard, drowning himself, host/actuator failure, power supply faults, engine room accident, Oil Spilling from Ship, movement of cargo, self rescue parabolic cargo accident•hazardous sudden accident etc.

Because of the diversities in accident cause, the ship type, ship size, hull structure, site of the accident hydrological and meteorological conditions, traffic condition of the accident scene, the ship self-help capabilities, crew emergency experience, the consequences of maritime accidents (or potential consequences) will be quite different, the demand for ship rescue will be different too. However, the ship assistance demand can be roughly summarized as follows:

(1) Carrying plugging material, and go to the scene for plugging;
(2) Carrying drainage equipment, and go to the scene for drainage;
(3) Dispatching high-powered vessels for towing, or dispatching a vessel to follow until the ship in distress berths;
(4) Replacing the damaged parts to ensure that the ship runs normally;
(5) Carrying stretchers, blankets, etc. to the scene to evacuate the wounded;
(6) Carrying bags to the scene to evacuate the corpses;
(7) Dispatching a passenger ship or emergency ship to evacuate persons;
(8) Dispatching the emergency ship carrying survival suit, life raft, etc. to search and rescue lives;
(9) Assisting the crew to shift load, and dispatching the cargo ship to shift load, and dispatching floating crane and floating balloon to float;
(10) If the ship grounding incidents take place in the reservoir area, the ship will refloat by raising or lowering the water level (when the ship is damaged and the direction of the ship is perpendicular to the channel, to raise the water level will crash the hull and cause the ship to break off.);
(11) Carrying fire extinguishers and other equipment to put out the fire.

These necessary supplies and equipments of salvation measures are shown in Table 1.

Table 1. Rescue supplies and equipment needs of the ship

Serial number	Types of salvation	Necessary supplies and equipments
1	Carry plugging material, and go to the scene for plugging	Plugging sandbags, blanket plugging, plugging board, plugging box, plugging screw, plugging column, plugging cork, sand and cement and so on
2	Carry drainage equipment, and go to the scene for drainage	Plugging sandbags, blanket plugging, plugging board, plugging box, plugging screw, plugging column, plugging cork, sand and cement and so on
3	Carry drainage equipment, and go to the scene for drainage	Spare parts of engines, diesel engines, oil purifier, turbocharger, air compressors, steering gear, pumps, communication and navigation equipment and marine deck machinery and so on
4	Carry materials of casualty evacuation to the scene to evacuate the wounded	Stretcher, blankets, breathing apparatus, bandages and so on
5	Carry materials of corpse evacuation to the scene to evacuate the corpse	Corpse bags, stretchers, chemical tape and other materials
6	Carry life-saving supplies to rescue lives	Life suits, life rafts, buoys and other life-saving equipment
7	Use floating crane and floating balloon to float the ship	Floating crane, floating balloon and so on
8	Carry fire-fighting equipment to put out the fire	foam extinguisher, dry ice extinguisher, gas extinguisher, and so on
9	Salvage wrecked ships	(Carried by the salvage company)
10	Lighting and communication	navigation lights, searchlights, yellow xenon lamp, gold eyes, VHF telephone, shortwave radio, marine satellite telephone

3 The Basic Elements of Ship Rescue Methods

Ship rescue methods are the collections of rescue measures for ships in distress, and the materialization of emergency response measures in contingency plan.

Due to the otherness of the needs of rescue ship and the rescue conditions, it is very difficult to achieve standardization, refinement, structured for ship rescue methods. But from the aforesaid rescue needs, ship rescue methods of water transportation accidents

involves five basic elements that are rescue needs, they are rescue efforts (institutional, marine, materials, etc.), rescue measures, conditions of acceptability, standards of behavior. It is shown in Fig. 1.

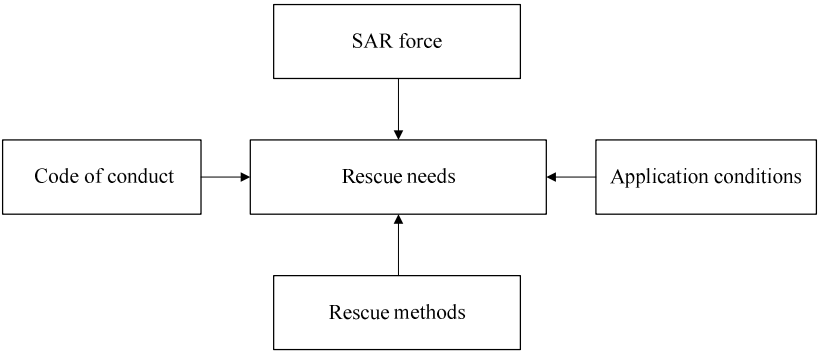


Fig. 1. Basic factors of ship rescue methods

4 The Description of the Digital Structure of Ship Rescue Methods

The digitalization of ship rescue methods mainly refers to store the rescue methods, the applicable condition of rescue methods and the support models of ship rescue in the computer system on digitized method. Its digital structure is mainly composed of ship rescue methods base, rescue methods applicable condition base, ship rescue support model base, SAR database and inland-river electronic chart. The description of the digital structure of ship rescue methods is shown in Fig. 2.

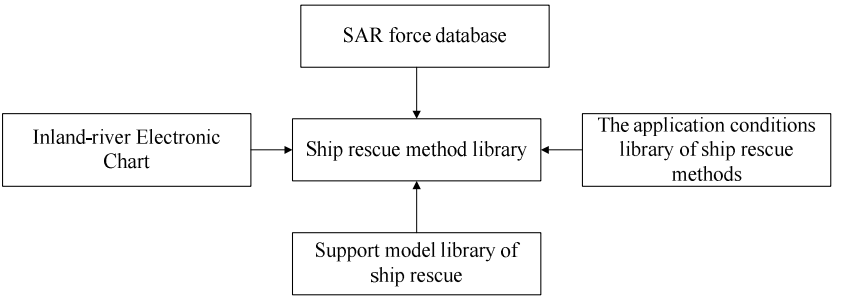


Fig. 2. The description of the digital structure of ship rescue methods

Inland-river Electronic Chart is the SAR situation made up of GIS platform, the integrated display and plotting of distressed ships and rescue strength, and provides comprehensive functions of calculation, measurement and information search tools.

4.1 The Design of the SAR (Search and Rescue) Force Database

To make a scientific and rational decision of SAR, Firstly, we need to accurately master all available information of SAR force, including information of vessels and survivors in distress, professional rescue ships, mariners, SAR aircraft, overbank SAR site, hydrometeorology, health units, social forces, military forces, SAR experts, and so on. The SAR force database Contains eight basic data table: SAR ship statistics, SAR aircraft statistics, SAR force units communication table, the SAR center communication table, health units communication table, meteorological unit communication table, telecom unit communication table, coast radio communication table. To build the subsystem, the key technology is the standardization technique of constructing a static SAR forces base.

Table 2. Statistical tables of SAR (search and rescue) Ships

Field name	Type	Length	Null	Content	Remarks
ID	int	8	N	Identification code of SAR Ships	primary key
Units name	char	30	N	Name of the unit who owns the ship	Chinese characters
Office Tel	char	30	N	Office Tel of the unit who owns the ship	Chinese characters or digital
Ships name	char	20	N	Names of SAR ships	Chinese characters or digital
Navigation area	char	20	Y	Classification of Navigation area	Chinese characters
Type of the ship	char	10	Y	Type of SAR ships	Chinese characters
Ship length	float	single	Y	Ship length	M(meter)
Total tonnage	float	single	Y	Total tonnage of a ship	Ton
Horsepower	char	20	Y	Horsepower index of a ship	Maybe more than one engine
Navigational speed	float	single	Y	Navigational speed, in knots	Maybe character
Cruise duration	float	single	Y	Cruise duration, in hours	Maybe character
Wind resistance	char	10	Y	Wind resistance of the ship	
Firefighting capability	char	10	Y	Point out Whether it has the Firefighting capability	Maybe other types
Communication equipment	char	20	Y	Communication equipment of the ship and its type	Note changes
Mooring point	char	20	Y	Name of the dock	Chinese characters
Remarks	char	50	Y	Remarks	Chinese characters
Mooring position	char	20	Y	Save the position information of the mooring to display it	Expansion field

These database tables are provisional data model. There will be changes according to the actual situation. The real-time information of the SAR units used in the monitoring of search and rescue process is stored in the ship-database of the electronic chart platform.

4.2 Design of Ship Rescue Method Base

Ship rescue method base composed by the unstructured data stored in unstructured database. Ship rescue method base includes a variety of specific methods and the corresponding implementation.

Table 3. Ship rescue method base

Field name	Type	Length	Null	Content	Remarks
ID	int	8	N	Identification code of SAR Ships	primary key
event	char	50	N	Inland river traffic accident type	Chinese characters
State	Object	variable	N	The describe to inland river traffic accident	Object
Rescue Method	Object	variable	N	Corresponding rescue implementation	Object
Remarks	char	50	Y	Remarks	Chinese characters

This database tables are provisional data model, it will make changes based on the actual situation. As a rescue ship method base which formed by the unstructured data, stored in the "Ship Rescue Knowledge Base" of electronic chart platform.

4.3 Design of Ship Rescue Applicable Condition Base

Ship rescue applicable condition base composed by the structured data stored in a structured database, which mainly includes the applicable conditions of various rescue methods. For each ship rescue method, number of records form a variety of application conditions, the fuzzy query is needed in the rescue method query and determine to show methods of rescue ships applicable conditions.

Table 4. Ship Rescue Applicable Condition Base

Field name	Type	Length	Null	Content	Remarks
ID	int	8	N	Identification code of SAR Ships	primary key
Event	char	50	N	Inland river traffic accident type	Chinese characters
Applicable condition	Char	200	N	Describe to ship rescue applicable condition, include the corresponding support model	Chinese characters
Remarks	char	50	Y	Remarks	Chinese characters

This database tables are provisional data model, it will make changes based on the actual situation. As a rescue ship method base which formed by the structured data, stored in the "Ship Base" of electronic chart platform.

4.4 Design of Support Model Base

Support model base is a series of calculation program packages which composed by supported model, the package is called by the system. While the support model is needed, the system will call the matched calculation program of support model.

Table 5. Support model base

Field name	Type	Length	Null	Content	Remarks
ID	int	8	N	Identification code of SAR Ships	primary key
Event	char	50	N	Inland river traffic accident type	Chinese characters
Calculation package path	char	256	N	Corresponding support model calculation package path	English characters
Remarks	char	50	Y	Remarks	Chinese characters

This database tables are provisional data model, it will make changes based on the actual situation. As a rescue ship method base which formed by the structured data, stored in the "Ship Base" of electronic chart platform.

5 The Formal Representation of the Ship Rescue Methods

As the ship rescue methods is part of emergency response plan, ship rescue methods are described in the form of unstructured text in most cases. Similar to the response plan, ship rescue methods can be formalized expressed by Ontology approach.

5.1 The Overall Structure

As shown in Formula 1, the rescue methods can be formally represented by the following triple,

$$\text{RescueMeth} = (\text{ID}, \text{MngInfo}, \text{MethCnt}) . \quad (1)$$

Of which: ID stands for the unique identification code of the rescue methods; MngInfo represents the manage information of the rescue methods, and it is also the program records of making, amending or enacting the methods. MethCnt stands for the core content of the rescue methods which is used to guide the ship's emergency rescue.

Furthermore, MethCnt can be generalized as the following Quintuple:

$$\text{MethCnt} = (\text{Demands}, \text{Acts}, \text{Res}, \text{Rules}, \text{Constrs}). \quad (2)$$

Of which: Demands is the set of the needs to rescue the vessels in distress, it general determine the type and content of the information of the rescue methods. Acts is the set of the related disposal measures of the rescue approaches. Res is the set of the various rescue source and resources. Rules is the set of the rules of the relief operations, which represents the behavior standard of the various subject in the process of the ship rescuing. Constrs is the collection of all constraint conditions.

5.2 The Definition of the Main Elements of the Collection

The demands for rescue can be defined as the Quintuple:

$$\text{Demands} = (\text{RescueType}, \text{Place}, \text{DangLevel}, \text{TimeDem}, \text{Resource}) . \quad (3)$$

Of which: RescueType is the type of rescue; Place represents the place where rescue happened; DangLevel represents the level of the Danger; TimeDem is the time demand for rescue; Resource represents the needs of source and resource to rescue.

The demands for rescue can not only be supposed by the ships (rescue object) in distress, but also be formed by assistant decisions using professional mathematical models, such as Hydrodynamic Model, Ship Refloating Tension Model and Oil Weathering Model. These professional models can be stored as the package of program-to apply and be related to rescue methods through the accessing path of program package.

$$\text{Act} = (\text{ID}, \text{Agents}, \text{ActType}, \text{Obj}, \text{Res}). \quad (4)$$

Of which: ID is the identity of methods; Agents represents the executive subject (can be an agency, a post, an Automatic Device and many others); ActType represents the type of rescue actions; Obj represents the operation object; Res stands for the available resource.

The rules of rescue actions (Rule) can be described by ECA, Which is equation:

$$\text{Rule} = (\text{ID}, \text{Event}, \text{Cond}, \text{Acts}). \quad (5)$$

Of which: The rule “ID” rules that if accord with provided conditions, the workflow “Acts” should be executed when the accident “Event” occurred, and there are many methods named “Acts”. Thus, rules represent the mapping relation of accident, conditions and workflow. “Event” and “Cond” is represented by predicate logic.

5.3 The Formalized Structure of Rescue Methods

According to above definition, the formalized structure of rescue methods (digital) is shown in Figure 3.

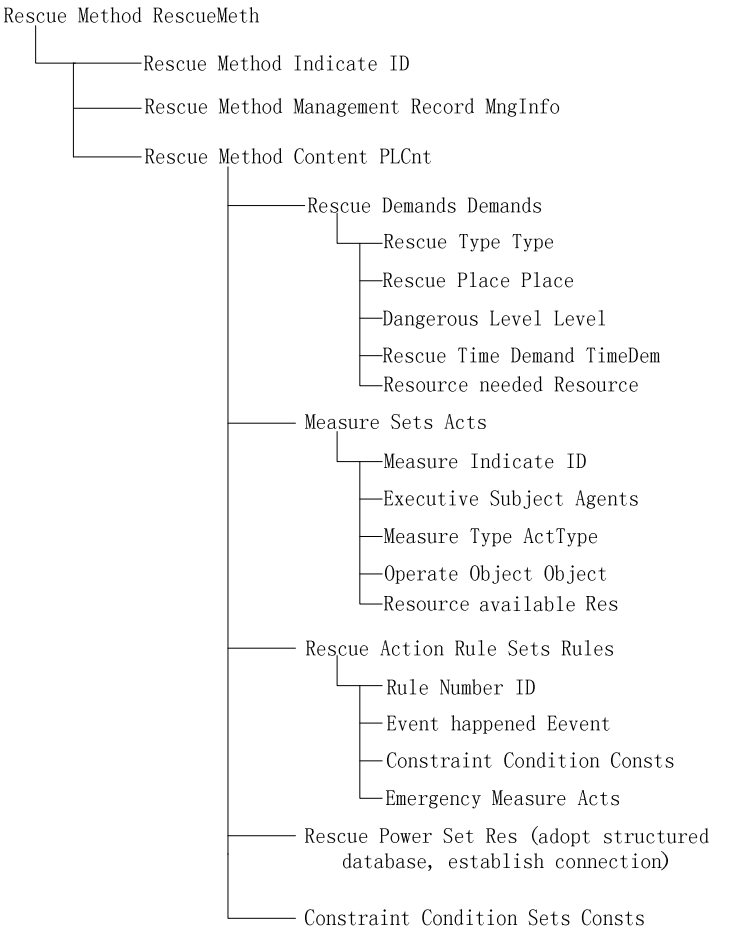


Fig. 3. Formalized Structure of Rescue Methods

6 Conclusions

According to the analysis of traffic accident emergent rescue demand of different ships and the relationship among basic elements of ship rescue method, unified ship rescue method framework can be established. Based on analyzing basic elements of ship rescue method, it can be designed to establish ship rescue method base including rescue power, rescue method, rescue applicable condition and data structure of basic elements of ship rescue method such as bolster model, in order to realize traffic accident storage of ship rescue method and add relevant attributes for various digital rescue methods which depends on the application demand of rescue method, such as method number, category, applicable condition, bolster model, etc. This paper used ontology to construct formal structure of ship rescue method, in order to realize digital management and application of ship rescue method and applicable condition, offer information

support for ship rescue emergent decision and command, and provide fast and effective reference rescue methods for grounding, touch damages, submergence, out of control, dragging, fire and thus cause ship influent and oil spill.

This paper analyzed emergent rescue command of inland ship accidents, researched digital technology and program of ship rescue methods, thereby laid the foundation for the ship traffic accident emergent rescue assistant decision system. Meanwhile, it put forward technical scheme of ship digital rescue methods, provided theoretical basis and technical project for emergency response system of China water ship accidents and maritime accident emergent rescue assistant decision system. It offered technical support for China water transportation security administration.

Foundation Project

Traffic Construction in Western China (grant number: 2009328811064).

References

1. Zhang, C., Pei, Y., Qiu, H.: Current status and development trend of digital emergency plan at home and abroad. Journal of Safety Science and Technology (5), 154–158 (2010)
2. Liu, X., Xue, A.: Review on the Development of Emergency Response Plan Supporting System. China Safety Science Journal (9), 87–91 (2007)
3. Hu, H., Liu, X.-M., Yang, X.-K.: Study on building up scientific and rational emergency management mechanism in China, 24–25 (2007)
4. Common Alerting Protocol (V1.2), http://docs.oasis-open.org/emergency/cap/v1.2/CAP-v1.2-os.pdf
5. Emergency Data Exchange Language (EDXL), http://www.safecomprogram.gov/NR/rdonlyres/71F415AC-8002-4A01-A2A7-F5E5C74B1FC0/0/FINAL_EDXL_OnePager.pdf
6. IEEE 1512 (IEEE Standard for Common Incident Management Message Sets for Use by Emergency Management Centers), http://www.ncef.org/pubs/emergency_response.pdf
7. Fu, Y., Zhu, Y., Wu, Z.: Water Safety Supervision and Management. Dalian Maritime University Press (2001)
8. Du, J., Xu, B., Tian, B.: Dangerous condition and ability of rescue for damaged ship. Journal of Dalian Maritime University (11), 25–27 (2003)
9. Worlds' First IMO-Compliant VDR, Safety Technology Section. The Naval Architects Journal, (October 2006)

Design and Implementation of Low-Power Time Difference Ultrasonic Flow-Meter

Liu Qiang, Wang Rang-ding, and Chen Chang-gen

School of Information Science and Engineering,
Ningbo University, Ningbo 315211, China

Abstract. Development of low-power ultrasonic flow meter is a very important process, directly determines the uninterrupted service life of ultrasonic flow-meter. Based on the chip MSP430F449, a more in-depth study for low-power issues of the time difference ultrasonic flow-meter, can guarantee 5+1 years in the case of power battery of 19Ah, greatly reducing the ultrasonic flow total system power consumption.

Keywords: Chip MSP430F449, low power, frequency control, mode selection.

1 Introduction

Currently, the use of ultrasonic flow-meters becomes more widely, while its application in the water, oil, natural gas and other fluid measurement becomes more popular. With the increase of accuracy, and the reduction of power consumption and cost, ultrasonic flow meter is bound to gradually replace the traditional Flow meter, and become the pillar of fluid measurement.

The research for low-power of ultrasonic flow-meter has always been the key to the researchers want to break. There are two power supplies for ultrasonic flow-meter: AC and DC. But in the case we need to monitor the water, oil or natural gas pipelines in the field, AC is clearly not convenient; contrarily, the use of battery-powered for ultrasonic flow-meter is shown the outstanding advantage. However, the premise to use battery-powered systems is low power consumption. In this paper, we will explore the connection between the various modules from system foundation of the ultrasonic flow-meter, give reasonable control to reduce some system power consumption according to the different characteristics of each module, and through online simulation test, reduce system power consumption significantly, while ensuring the flow-meter accuracy.

2 Hardware Structure of the Secondary Meter

Build secondary instrument reasonably, not only can guarantee the measurement accuracy, but can also reserve more affluent space for reducing system power. In the design process of time difference ultrasonic flow-meter, we use a low power microcontroller MSP430F449[1] as the system MCU, the chip TDC-GP2 as a core component of timer module. In addition, the system also includes power supply module, key modules, liquid crystal display module, external memory module, serial communication interface and signal processing modules, shown in Fig. 1.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 469–475.
springerlink.com

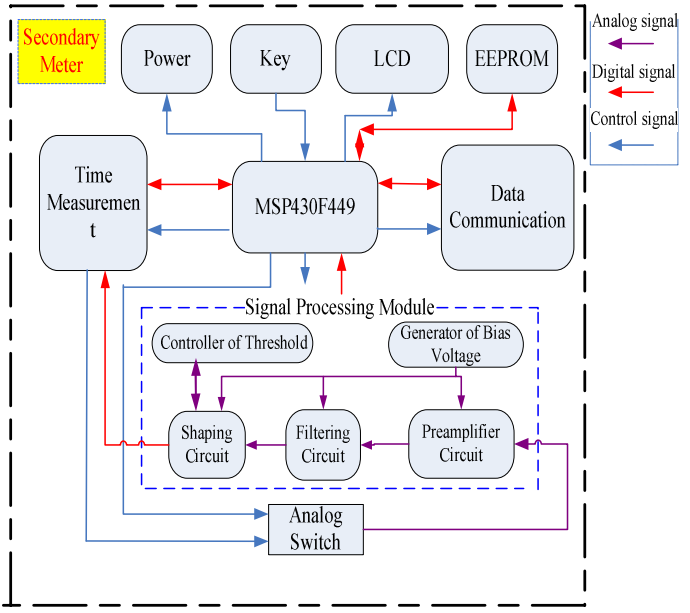


Fig. 1. Hardware structure of the secondary meter

Through systematic analysis to the hardware structure, by branching design of the power module, monitoring and control of the voltage, selecting a reasonable way of detection to key modules, software control to liquid crystal display, selecting the appropriate timing mode and system clock can effectively reduce the system power consumption[2] and improve battery efficiency, extend the ability to work uninterrupted, which can enhanced usability of the ultrasonic flow-meter.

2 Methods of Reducing System Power Consumption

2.1 Switch Control of Power Module

Branch Design of the Power. The ultrasonic flow-meter system includes a digital part and analog part. when designing the PCB board ,we usually separate the two parts and copper, then connect the grounds of two parts by the 0 ohm resistor, so that get the smallest interaction between the parts of digital part and analog. For the time difference ultrasonic flow-meter, its digital part includes two important modules, namely the microcontroller MSP430F449 and time measurement module. In the actual design process for the circuit, power can be divided into three ways: a way for the all Analog part; a way for the MCU alone, because in general, the MCU always be in a working state, even if it works on the ultra-low power Mode; and the last way is designed to provide power for the time measurement module, this part is the core module to the measuring system of time difference ultrasonic flow-meter, that because only when we

ensure the timing accuracy, we will ultimately be able to ensure the accuracy of flow meter. To the general timing chip, it usually requires a stable power supply, so after providing a separate power supply, we should add a voltage regulator module to ensure the accuracy of its measurement, shown in Fig. 2.

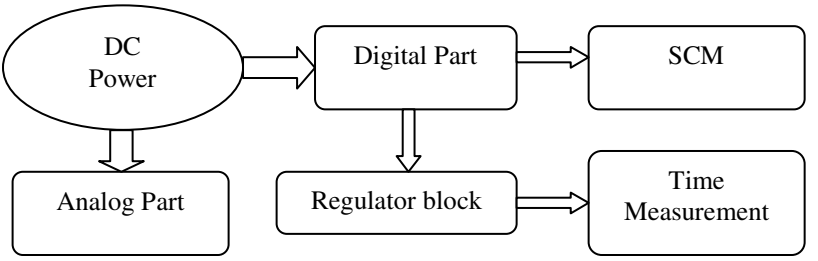


Fig. 2. Branching design of power supply

The benefits of branching design to the power are: according to the actual measurement process of ultrasonic flow meter, we can realize the independent on / off control to the analog part , the SCM and the time measurement module. In the actual measurement, the microcontroller must keep working , but its is very small; the power consumption of time measurement module is larger, but we need ensure the flow meter's accuracy, so we usually keep its power being left open; and for the analog portion which have the maximum power, the role of the branching power reflects more clearly. When the analog part is in working condition, the current will reach up to 18mA, but the ultrasonic flow-meter of low power requires the mean system current should be around 0.3mA. So through the branching design, we can turn off the power in the period the analog part is not needed to reduce the current consumption.

Through the simple design of branching power and the switch control on software, the system power consumption can be reduced most, which is very effective to the whole flow system.

Control of Supply Voltage Supervisor (SVS). SVS can operate in two modes: supports the voltage supervision (the system will automatically reset) and the voltage monitoring of power supply (devices can not automatically reset). We can use the function of the voltage monitoring to set the minimum supply voltage by software; both choose different modes of SVS. When the voltage drops below the minimum supply voltage, it can remind the SCM to do some related operations, such as save the important measurement parameters, cumulative flow, dates and so on. The minimum supply voltage can be selected by software, there are a total of 16 different values as shown in Table 1. To the ultrasonic flow-meter of battery-powered, a good use of voltage monitoring of SVS module, can not only give the measurement data protective effects, but also show whether the system power is insufficient or we should change the battery in order to avoid the flow meter stopping working, which will cause a major mistake to the whole system.

Table 1.

M0	M1	M2	M3	M4	M5	M6	M7
OFF	**1.9V**	**2.1V**	**2.2V**	**2.3V**	**2.4V**	**2.5V**	**2.65V**
M8	M9	M10	M11	M12	M13	M14	M15
2.8V	**2.9V**	**3.05V**	**3.2V**	**3.35V**	**3.5V**	**3.7V**	**1.2V**

2.2 The Way of Key Detection

In the process of research and development of ultrasonic flow-meter, there will be key requirements. To the ultrasonic flow-meter based on MSP430F449, three options can be used as its key detection: scanning, interrupts and the application of on-chip AD. For these three ways, if we use the way of on-chip AD, finally we will find that it is almost the same as the way of scanning. The difference between two ways is: the power consumption of the on-chip AD will be relatively larger, and if we use this method, the microcontroller should keep cyclic scan, as a result, it will take up the program space. So these two methods will cause a power loss to some extent. But the way of interrupt can meet the different low-power operating modes of MSP430F449 and be able to reduce the system power consumption, at the same time, it can response interrupts, and return to the low-power mode before after the interrupt service. This way can not only realize the function of key detection, but also minimize the power consumption of the system. State transitions of the interrupt mode are shown in Fig. 3.

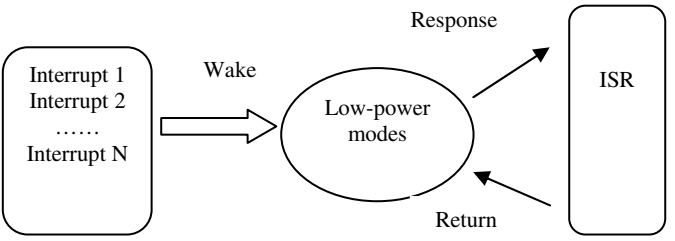


Fig. 3. State transition of the interruption

2.3 Control of Liquid Crystal Display (LCD)

The biggest role of LCD should be to help read the data, but the LCD display also consumes power. To the ultrasonic flow-meters used for outdoor, we need the LCD can display only when the maintenance personnel read data, it is not necessary to always show. Therefore, in the stage of research and development of the flow meter, the LCD can be displayed in the way of interrupt, that is when we need to read the data, such as the cumulative flow, instantaneous flow and so on, what we should do is just to click some key to start the interrupt, it will display the current measurement data, with another click, we can then achieve the normal function of the LCD module. In addition, without affecting the LCD display, we can also set the clock frequency to the minimum level in order to reduce system power consumption.

The LCD module will produce an operating current about 3uA or so in the scanning frequency of 128Hz. If a battery of 19Ah can be used for 5 +1 years, it can be calculated that only the power savings of the LCD module will keep the entire system working for about 20 days, which is also a major contribution to the low-power design of ultrasonic flow-meter.

2.4 Low-Power Control in the Process of Time Measurement

Module of time measurement is the core of time difference ultrasonic flow-meter system, the measurement accuracy of this module directly affects the final accuracy of the flow-meter. Generally speaking, we do not make too much power control in this part, so that it can work in the best condition, and achieve the highest measurement accuracy at last. However, microcontroller MSP430F449 used in the system can meet the module of time measurement because of its different low power modes, the function of fast wake-up and the advantage of timely response to interrupt. This can achieve lower power consumption without affecting the measurement accuracy.

For the module of time measurement, the stability of the power is highly required. Although we have made a branching design for the power when make the PCB, and provide a separate branch of power supply for timing module, we do not control its switch. Otherwise it will affect the measurement accuracy, and will not meet the national standards of ultrasonic flow-meter. In contrast, we must add a voltage regulator in this branch in order to ensure the stability of its supply voltage. We can do some crystal control of the time measurement module instead its switch control. This just needs a reasonable delay in the crystal start-up phase.

By low-power control, the operating current of time measurement module can be reduced from the original about 470uA to 250uA or so. Compared with the LCD module, it can be seen that the power consumption of this part has a relatively significant decrease.

2.5 Control of the Systemic Clock

Selection of the systemic frequency in varying degrees will give different effects on power consumption, of course, when try to control the systemic frequency, we should also take into account the program execution time, the speed of data processing, the stability of the whole system and so on.

In the process of writing code, six clock frequencies are set, those are: 1MHz, 4MHz, 8MHz, 10MHz, 12MHz, 16MHz. We can use the clock frequency of 1MHz for the program of LCD, Judgments and assignments; 4MHz or 8MHz for the module of time measurement; 8MHz, 10MHz or even higher for the code of digital filtering, data modification and the flow calculation. Complexity and runtime of the program together determine the clock frequencies which we actually need. To the part the code is easy or there is no data processing and requirements of speed, we can choose the low-frequency clock, so that the power consumption could be reduced. And to the part the microcontroller is required to process fast, then we need a high-frequency clock, but the frequency should not be too high, because as the frequency doubling, the power consumption will become larger, of course, the frequency can not be too low, or data processing of the system will be too slow, it will also produce relatively large power, so

when set the systemic clock, we must consider the overall situation, minimize system power consumption while making fast data processing. After the on-line simulation of the program, we can get at the clock frequency of 1MHz, it is about 250uA or so, while the systemic current will increase to about 560uA when the frequency upgrade to 8MHz.

3 Analysis of Power Consumption

After a reasonable low-power control to each module of the time difference ultrasonic flow-meter, the systemic power consumption can be effectively reduced. To use only a battery of 19Ah and 3.6V of the voltage, the flow meter can be able to keep on working for about 5+1 years. The current consumption of the ultrasonic flow-meter in various stages showed in Table 2.

Table 2.

System Status	System Current	Duration (s)	Time multiplied by the current (uA*s)
Time sampling	16.157mA	0.001831055	29.58435564
LCD	247uA	0.00213623	0.52764881
8MHz of the systemic frequency	561uA	0.001586914	0.890258754
Calculation of the time difference	3.968mA	0.125274658	497.0898429
Low-power modes	60uA	1.869171143	112.1502686

Table 2 shows the power consumption of each stage in one cycle, the time for each cycle is 2S which can be set by the basic timer of MSP430F449. By calculating, the average value of time multiplied by the current is 320.12uA * S, so the service life is:

$$T = \frac{19A \times 1000000}{320.12uA * S \times 24 \times 365} = 6.775 \quad \text{Year}$$

4 Conclusions

Since the development of the time difference ultrasonic flow-meter goes along with the requirements of low-power and high accuracy, there is a mutual restraint between the two, but one can not. Therefore, we have done more in-depth study to each module of the time difference ultrasonic flow-meter based on MSP430F449, at the same time guarantee the systemic accuracy, and then give reasonable low-power control to each module. As a result we have significantly reduced the systemic power consumption, so that the whole system can keep on working normally for 5+1 years with the power supply of one battery, which has solved the problems that the inconvenience of using

AC power supply in outdoor and the high power consumption of using DC power supply. This study has much stimulating effect to the development of the low-power ultrasonic flow-meter.

Acknowledgement. This work is supported by the National Natural Science Foundation of China (NSFC: 60873220, 61170137), Doctoral Fund of Ministry of Education of China(20103305110002), Zhejiang Natural Science Foundation of China (Y108022, Z1090622, Y1090285), Zhejiang Science & Technology Preferred Projects of China (2010C11025), Zhejiang Province Education Department Key Project of China (ZD2009012), Ningbo Science & Technology Preferred Projects of China (2009B10003), Ningbo Key Service Professional Education Project of China (2010A610115), Ningbo Natural Science Foundation (2009A610085) and Ningbo University Foundation (XYL10002, XK1087). In addition, our programs are supported by High School Special Fields Construction of Computer Science and Technology (TS10860), Zhejiang Province Excellent Course Project (2007), Zhejiang Province Key Teaching Material Construction Project (ZJB2009074).

References

1. Shen, J., Yang, Y., Zhai, X.: Principle and application of the 16-bit ultra low power microcontroller of MSP430. Tsinghua University Press (2005)
2. Bu, A., Li, J., Wang, C.: Low-power design policy for mixed-task system. Journal of Southeast University (Natural Science Edition) (2010)

Research and Application of Integrated Scheduling System of Steelmaking-Continuous Casting-Hot Rolling in Shagang Group

Liu Lijun, Wen Zhi, Su Fuyong, Dou Ruifeng, Liu Xunliang, and Lou Guofeng

Thermal Eenergy Engineering Department,
University of Science and Technology Beijing, Beijing 100083, China
beris-mcc@163.com, wenzhi@me.ustb.edu.cn

Abstract. In this paper, a mathematical model of scheduling system is developed based on the analytical investigation in Steelmaking-Continuous Casting-Hot Rolling process, additionally, according to which an integrated scheduling system is established and being put into practice in Shagang Group. The application of the system achieves that more than 90% hot delivery and hot charging ratio is conserved in the process, more than 5% oxidization loss is reduced in the process of Continuous Casting (conticast) slabs and consumption of reheating furnace is ≤0.95GJ/t in per kilo product.

Keywords: Conticast slabs, hot delivery and hot charging, mathematical model, online control.

1 Introduction

A 1700mm hot strip rolling line was put into production in Shagang Group in May of 2005,so far the annual productivity has reached 4,000,000t. The main equipment in the workshop includes two 1×2 slab CCMs, three reheating furnace, a primary descaler, a roughing mill, a crop shear, a secondary descaler, a finishing mill, a laminar cooling device, an underground coiler and etc.. Fig.1 shows the conveying flow of slabs from the CCMs to the reheating furnaces.

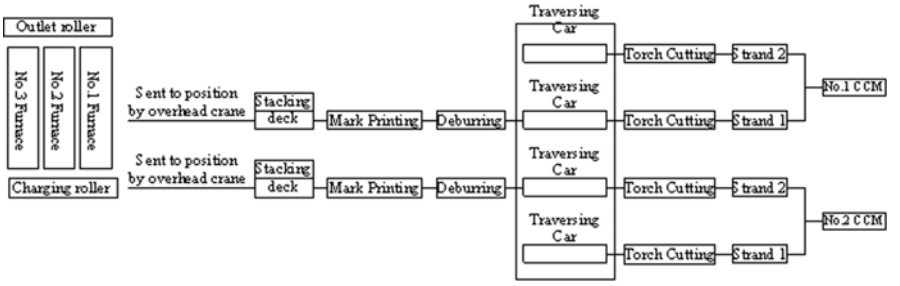


Fig. 1. Process Flow Chart of Conticast Slab Charging into Furnaces before Revamping

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 477–483.
springerlink.com

The conticast slabs produced by the CCMs is sent into warehouses by overhead cranes after torch cutting, deburring and mark printing. Upon receiving reheating furnace charging command, the slabs are delivered to the lifting roller tables by a crane beside the charging roller tables which are in front of the reheating furnaces and then be pushed onto the charging roller tables, where the slabs can be loaded into the reheating furnaces.

Because of the insufficient coordination and cooperation between steel-making and steel rolling the two major processes which has two separate production schedules, the high-temperature slabs produced by the CCMs are stored in warehouses failed to be delivered into the heating furnaces immediately leading to a huge loss of physical sensible heat. Consequently, not only the heat is wasted, but also, in the following process, low temperature slabs extends working periods in furnaces resulting in production declined, energy consumption increased, oxygenation loss added and cost raised.

Besides, due to the limitation of building arrangement, the length of any of the three reheating furnaces is restricted in 35m, and the cold slab reheating capacity of each furnace can only be 250t/h. With the current production technology, i.e. the present hot charging standards, it would be difficult to meet the requirement of 4,800,000t annular output in the steel rolling plant.

Therefore, Shagang Group decided to carry out hot delivery and hot charging technical innovation on the 1700mm production line, by adopting high-level hot delivery and charging technology and partly direct charging processes[1][2][3], to achieve 4,800,000t annual productivity and more energy conservation.

2 Implementation Solution

The hot delivery and hot charging revamp of the 1700mm production line in Shagang Group involves both hardware and software aspects. The hardware work includes building a new connecting roller table from the exit of No. 2 CCM slab stacking deck to the inlet of reheating furnace, so as to open a smooth channel to deliver conticast slabs, constructing 5 new heat holding pits to store the spare offline slabs to regulate productivities between steel making and steel rolling processes which creates a necessary condition for conticast slab hot charging processes. The software work contains an integrated production management system concerning steel making-CCM production-steel rolling and a Level 2 computer optimization control system for reheating furnaces, which is designed for order management, production scheduling, dispatching management, reheating furnace optimization control and etc. to assure a high ratio of direct charging and hot charging and high productivity, low oxidization and low energy consumption.

2.1 Hot Connecting Roller Table

72 new roller tables are added from the exit of No. 2 CCM slab stacking deck to the inlet of reheating furnaces, which are divided into 12 groups with one hot metal detector in each group. Every roller table is controlled by a singly operated drive and every motor has 5.5kW power and 30m/min roller speed. A centralized grease lubrication system is adopted.

2.2 Additional Heat Holding Pits

Five new heat holding pits are built beside the hot connecting roller table. The maximum storage capacity of each pit is 30 conticast slabs. The pit covers have high-and-low push-and-draw mode, the advantages of which is rational structure, easy to manufacture, better performance and etc. Thermal insulation of the pits is guaranteed by multilayer refractory bricks compound masonry at bottom of the pits and fully-fiber module structures used on the covers and inner walls of the pits. Each pit has its own high-and –low cover with electric controls which can be operated on the spots, in ICS operating tables or be remotely controlled from crane.

The revamped process flow chart is fig.2:

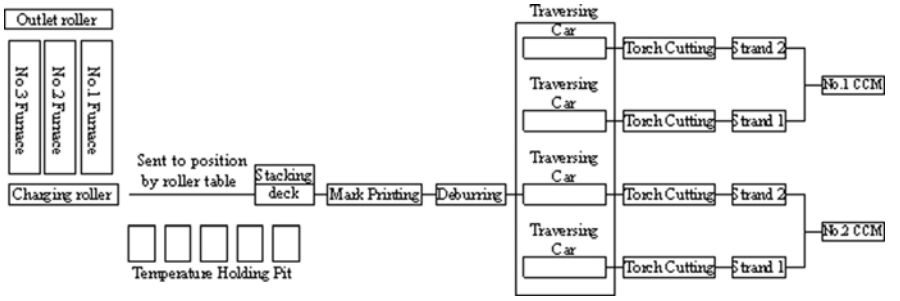


Fig. 2. Process Flow Chart after Revamping from CCMs to Reheating Furnaces

2.3 Building Mathematical Models for Slabs Hot Delivery and hot Charging

The mathematical models for slabs hot delivery and hot charging include: mathematical models for the heat process of CCM system; mathematical models for the heat process of slab temperature holding; mathematical models for the heat process of slab heating; mathematical models for slab rolling.

Control equation is listed for slab section:

$$\rho c \frac{\partial T}{\partial t} = \frac{\partial}{\partial x}(\lambda \frac{\partial T}{\partial x}) + \frac{\partial}{\partial y}(\lambda \frac{\partial T}{\partial y}) + q_v \tag{1}$$

Where ρ - slab density, Kg/m^3

c - slab specific heat, $J/kg.°C$

λ - slab thermal conductivity, $W/m.°C$

q_v - internal heat source, i.e. latent heat of solidification, J/kg

T - slab temperature, $°C$

Equations should be punctuated in the same way as ordinary text but with a small space before the end punctuation mark.

According to the differences of initial conditions and boundary conditions in the process of Continuous Casting, heat preservation, heating and rolling, a full set of

mathematical models have been built for the slab hot delivery and hot charging, then solutions of models and its experimental verification have been carried out. For details about modeling mechanism, control equation and solutions, please refer to Reference [4]. Furthermore, the calculation accuracy of mathematical models can fully meet operating requirements and the mathematical models provide a theoretical basis for building Level 2 control system of reheating furnaces.

By building hot process models for slab hot delivery and hot charging, the technical production qualifications in two different operation process conditions (i.e. direct slab charging and hot charging) , with 850℃ surface temperature, have been figured out. For details, please refer to Table 1 and Table 2.

Table 1. Production Qualifications under different Direct Charging Conditions

Slab Delivery Time (min)	Surface Temperature at Furnace Inlet (℃)	Maximum Productivity of Furnaces (t/h)	Unit Consumption of Furnace (GJ/t)	Oxidization Burning Loss (%)
10	805	484	0.95	0.90
15	782	470	0.96	0.92
20	762	456	0.98	0.95
25	744	445	1.0	0.97
30	726	440	1.02	0.99

Table2. Production Qualifications under different Hot Charging Conditions

Pit Holding Time (h) *	Surface Temperature at Furnace Inlet (℃)	Maximum Productivity of Furnaces (t/h)	Unit Consumption of Furnace (GJ/t)	Oxidization Burning Loss (%)
1	872	460	0.99	0.96
2	864	445	1.0	0.97
3	850	430	1.05	1.0
4	832	415	1.07	1.03
5	812	400	1.1	1.05

2.4 Optimization of the Integrated Production Management and Dispatching System

Functions of the integrated production management and dispatching system for Steelmaking-Continuous Casting- Steel rolling include generating the production schedules for every steps, arranging production automatically, real-time dispatching management and etc, which can assure an optimized working flow in the production line. In addition, with the real-time monitoring and exception handling policy, the optimum direct charging ratio, hot charging ratio and the highest charging temperature can be achieved, as a result, the productivity is increased, the energy consumption is limited, the oxidation loss is controlled and the production cycle is shortened. The integrated production management and dispatching system has following functions.

2.4.1 Generating Production Plan

During direct hot charging rolling of conticast slabs, steel making, CCMs and steel rolling is a united process following one general production plan. The steel making and steel rolling cannot be scheduled separately. There is a correspondence relationship between the slabs in the rolling unit and the charging & casting which has to be clearly specified in the rolling schedule.

According to individual order or several orders, the system generates frequency plans for steelmaking system, rolling and casting schedules via order processing modules. Taking volume of a furnace as the basic unit, the rolling sequence will be organized to meet the requirement of mill roll cycle, and then the frequency of casting will be scheduled under constraint conditions of group casting, by which the frequency of charging, rolling and casting achieves synchronous optimization. The production rhythm of steelmaking and CCM and the cycle of mill rolls will have to be taken into consideration to make charging, casting and rolling plans, furthermore, optimizing the production capacity and quality also is the part of plan in the production system, where an integrated working schedule has been made.

2.4.2 Production Dispatching Management

The system monitors the production line in real time in order to make immediate response to any emergency situations, adjusting production plans to ensure the production. According to various abnormal conditions of the steel making and rolling line, such as CCM failure, unscheduled production, mill roll cycle change and etc., dispatching strategies are made to stabilize and guide operations. Through tracking and monitoring, the operation efficiency of the whole line is coordinated and optimized, order status is real-time monitored, and the production condition of orders is confirmed.

2.5 Level 2 Optimization Control System of Reheating Furnaces

There are three walking beam reheating furnaces in the 1700mm steel rolling workshop, in which No. 1 and 2 furnaces are conventional ones with air preheating, and No. 3 is a regenerative reheating furnace with double heat regeneration of air and gas.

The primary mission of the level 2 optimization control system of the reheating furnaces is to provide rational heating for the slabs inside the furnaces based on production process requirement, to accurately control the slab discharge temperature and uniformity, to reduce oxidization loss, saving energy and assuring heating quality and productivity, also, to track and control the slabs in reheating furnace area.

The major control functions of level 2 system include:

2.5.1 Real-Time Calculation of Slab Temperature

In case of direct slab charging, the inlet temperature of every single conticast ingot will be calculated by the CCM process computation module. During hot charging, the inlet temperature of every single conticast ingot will be calculated by the CCM process computation module and heat preservation computation module.

Heat transfer calculation for every slab inside the furnace is carried out by heating process calculation module, according to the charging temperature, slab position

inside the furnace and the furnace gas temperature above and under slab surface calculated by the furnace temperature detected by the thermocouples inside the furnace. Therefore, the temperature of every slab at any moment inside the furnace has been figured out. In order to assure accurate temperature calculation, rolling process calculation module calculate the actual discharge temperature of each slab based on the surface temperature of each rolling piece after roughing mill, so as to update the calculation accuracy of heating process calculation module.

2.5.2 Furnace Temperature Dynamic Optimization Setting

After the zone deviation of each furnace section is calculated on the basis of deviation between the current temperatures of each slab calculated by heating process calculation module and theoretically optimized heating temperature, new furnace temperature settings are received by PID regulator. Please refer to Fig. 3.

2.5.3 Standby Time of Rolling

The main function of modules of standby time control and recovery process is to carry out optimization solutions for reheating furnaces under abnormal production conditions. If rolling mill, walking mechanism or slab extractor is breakdown, in the light of the provided failure grade, the operation of temperature reduction or heat preservation will be activated to minimize fuel consumptions and oxidation loss of reheating furnaces when the production is resumed and the productivities and heating quality is guaranteed.

The key of rolling standby time control is to quickly reduce the temperature of furnaces at the beginning of standby and to timely raise the temperature of furnaces at the end of the standby, however, temperature control is restricted by technological requirements of furnaces. The standby situation is divided into automatic rolling standby and rolling standby on request in standby time control policy. To be more specific, automatic rolling standby indicates the reheating furnaces stop walking operations without requests from rolling process. Rolling standby on request means the reheating furnaces terminate walking operations on requests from the process.

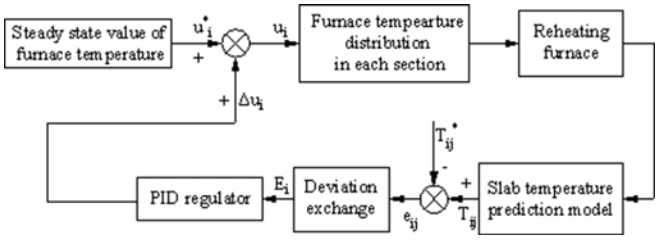


Fig. 3. Online Control Strategy Scheme of Dynamic Optimization Settings of Reheating Furnace Temperature

3 Practical Effect

Ever since putting the revamped 1700mm hot delivery and hot charging line into operation in June, 2010, satisfactory energy saving result and economic benefit have

been obtained. The statistics of practical production reports demonstrate that there are more than 90% the hot delivery and hot charging ratio achieved after the revamping, and no less than 5% oxidization loss has been reduced and consumption is ≤0.95GJ/t in per kilo product, where the expected effects of the revamped project is achieved.

4 Conclusion

The successful application of in the 1700mm hot delivery and hot charging renovation project breaks down the walls that limit corporate development, for instance, improving the operational level, increasing the production scale, reducing the consumptions, minimizing the oxidization burning loss and etc. Also, the management level and technical level of the plant is higher, the product cost is lower, the waste drainage/emission is reduced, and the competitiveness is stronger in the market, which brings both remarkable economic benefits and significant social benefits.

References

1. Kumar, A., Dutta, P.: Modeling of transport phenomena in continuous casting of non-dendritic billets. International Journal of Heat and Mass Transfer 48, 3674–3688 (2005)
2. Kim, J.G., Huh, K.Y.: Prediction of transient slab temperature distribution in the re-heating furnace of a walking-beam type for rolling of steel slabs. ISIJ International 40(11), 1115–1123 (2000)
3. Wen, Z., Gao, Z.: Research and development of hot conticast slab delivery and charging and direct rolling process. Metallurgical Energy 15(1), 22–26 (1995)
4. Hao, X.: Composite BP Networks Model for Thermal Process of Continuous Casting System and On-Line Application. Doctor Essays of University of Science & Technology, Beijing (2007)

Optimization of Producing Enzyme Conditions for Facultative Anaerobic Cellulose Degrading Bacteria in Aerobic and Anaerobic Environments

Ting-ting Fan[1], Si-ying Liu[1], Cheng-wen Wang[2,*], and Tao Long[1,3]

[1] Wuxi FengLu Environmental Protection Technology Co., Ltd., Wuxi 214200, P.R. China
fttcool@126.com
[2] School of Environment, Tsinghua University, Beijing 100084, P.R. China
wangcw@mail.tsinghua.edu.cn
[3] Nanjing Institute of Environmental Sciences, 210042, P.R. China

Abstract. A facultative anaerobic cellulose degrading bacteria was isolated from bovine stomach. Its enzyme production conditions under aerobic and anaerobic conditions were compared with orthogonal optimization. Results showed that under CMC-Na 0.6%, NH_4NO_3 0.20%, Tw-80 concentration of 0.025% and ascorbic acid concentration of 0.015%, the optimal yield CMC-enzyme activity can be achieved at 3874.34μmo/ (g·h); optimal incubation times were 96h (aerobic conditions)or 120h (anaerobic conditions), and the most favorable temperature is 37 ℃. The best initial medium should have pH = 7, with inoculation of 4% (aerobic conditions) or 8% (anaerobic conditions). Under the optimal conditions, the CMC activity in aerobic environment reached 4210.32μmol/ (g·h), while in anaerobic environment 3415.76μmol/ (g·h) was achieved. The enzyme activity under aerobic condition is 23.26% higher than under anaerobic condition.

Keywords: Facultative anaerobic cellulose degrading bacteria, sugar, CMC-enzyme activity, Optimization.

1 Introduction

Fiber material is the main hydrolysis substrate in the process of anaerobic digestion stage among agricultural waste, decomposition of cellulose provide carbon for methanogens as the growth and reproduction, but as its insoluble and heterogeneity makes it difficult to be, thereby become rate-limiting step of the fermentation process, and the anaerobic cellulose plays a critical role degraded microbial in the cellulose-rich waste anaerobic digestion, methane fermentation and food digestion, thus the research of anaerobic fibers degrade anaerobic microbial in environmental microorganisms becomes one of the hot factors[1]. Strains purified from natural is relatively low in producing cellulose enzyme activity, not suitable for direct

* Corresponding author.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 485–492.
springerlink.com

investment in production, need to strain breeding which is the most effective means to improve the production capacity of cellulose enzyme, and microbial in different media and culture conditions the growth and metabolic are different, by optimizing the medium and fermentation conditions for cellulose strains is also an effective way [2, 3]. In this paper, an anaerobic degradation of cellulose was screened from bovine stomach, based on their respective conditions in aerobic and anaerobic optimized its fermentation conditions for enzyme production, which is never been done before the attempt, and for degradation of Straw to biomass [4], which means that the use of resources in the new strain has a high application value for development.

2 Materials and Methods

2.1 Materials

Source strain: Strain screened from bovine rumen contents.Seed medium[5]: peptone 1.0g; yeast extract 1.0g; sodium carboxymethyl cellulose 5.0g; ammonium sulfate 2.0g; potassium dihydrogen phosphate 2.0g; heptahydrate magnesium sulfate 1.0g; sulfate heptahydrate 10.0mg; chloride manganese 5mg; sodium chloride 3.0g, distilled water 1L, adjust pH 6.7.

Fermentation medium: peptone 3.0g; yeast extract 3.0g; sodium carboxymethyl cellulose 5.0g; ammonium sulfate 2.0g; potassium dihydrogen phosphate 2.0g; heptahydrate magnesium sulfate 1.0g; sulfate heptahydrate 10.0mg, manganese chloride 5mg, sodium chloride 3.0g, distilled water 1L, adjust pH 6.7.

Experimental device: Aerobic incubation with 250ml of shakes flask cultures; Anaerobic sealed with rubber stoppers Hungate anaerobic tube liquid shake flask cultures [6].

2.2 Methods

1) Optimization the fermentation medium for facultative anaerobic cellulose degrading bacteria

a) Effect of different carbon and nitrogen on cellulose enzyme production

Change the basal medium carbon and nitrogen sources, with different carbon sources and nitrogen sources at 37℃, 200r/min cultured 96 h, determinates its CMC-enzymatic activity.

b) Orthogonal

Choose above-optimal carbon and nitrogen sources, added surfactant (Tw-80) and enzyme production promoter ascorbic acid to enhance the concentration of enzyme production in the culture, as carbon, nitrogen, surface active agent (Tw-80) and ascorbic acid four factors, each of the three levels of different addition, determined the composition ratio of fermentation optimum medium by four factors and three levels orthogonal experiment [7].

2) Optimization the conditions of fermentation for the strain in aerobic and anaerobic

a) The effect of different fermentation temperature and culture time on cellulose enzyme production

Take the seeds liquid 4% inoculated in 50ml fermentation medium, set different temperature, cultured in aerobic and anaerobic conditions, respectively, samples measured biomass and enzyme activity every 12h.

b) The effect of different initial pH of fermentation medium on cellulose enzyme production

Medium with 4% inoculation, initial pH adjustment in pH 6-8 range, cultured at 37 ℃, 200r/min in aerobic and anaerobic conditions, respectively, measured CMC-enzyme activities after 96h.

c) The effect of different inoculation amount on cellulose enzyme production

Set different inoculation; take the seeds of spores inoculated in 50mL medium at 37 ℃, 200r/min in aerobic and anaerobic conditions, respectively, measured CMC-enzyme activities after 96h.

3) Analytical Method

a) Determination of biomass

Measured the absorbance at 600nm of bacterial culture fluid, if the OD_{600} values obtained between 0.6-0.8, indicated that the bacteria was in the strong growth of the logarithmic phase, OD_{600}> 3 showed that the bacteria was already saturated and so on.

b) Determination of reducing sugar by DNS method.

c) Determination of CMC-enzyme activity and the activity definition

Determined the cellulose-enzyme activity by 3,5-Dinitrosalicylic acid DNS method [8]. As anhydrous glucose gradient solution to draw standard curve, test OD_{540} of enzyme samples to determine the reducing sugar content and calculated carboxymethyl cellulose (CMCase) enzyme activity. To 1.5 ml 1.0% (w / v) CMC as the substrate solution, add 0.5 mL diluted enzyme solution, at 50℃ for 30 min, added 0.5 ml DNS reagent and boiling water for 5 min, cooled and constant volume to 5ml, for the substrate without enzyme as blank ,determined photometric density (OD) at 540 nm in the spectrophotometer[9].

Enzyme activity unit is defined as:1.5 ml1.0% (w / v) CMC and 0.5 ml enzyme solution at 50℃, pH4.8 conditions, mixed decomposition of 1% CMC-Na and produce 1μg glucose in 1h,then defined as a cellulose activity unit (μmol / (g • h)).

3 Results and Discussions

3.1 Different Carbon and Nitrogen Sources on Cellulose Enzyme Production

Removal the carbon source from the basal medium, add different carbon sources: Glucose, Maltose, CMC-Na, Starch, Sucrose, content 0.3%,resepectively; Similarly, remove the nitrogen from the basic medium, addition of different nitrogen sources: NH_4Cl, Urea, $(NH_4)_2SO_4$, $NaNO_3$, NH_4NO_3,content0.2%,respectively,cultured at 37℃, 200r/min, measured activity after 96 h as shown in Table 1: found that the activity of CMC-Na as the carbon source was higher than other macromolecular carbohydrates which can promote the enzyme production; However, when NH_4Cl as nitrogen source

the activity is 2081.64μmol / (g·h), while NH_4NO_3 as nitrogen source the activity to 2177.34μmol / (g·h), so in the training process NH_4NO_3 can require to make up the bacteria nutrients timely than others.

Table 1. Effects of different carbon and nitrogen on cellulose activity

Carbon	Activity (μmol/ (g·h))	Nitrogen	Activity (μmol/ (g·h))
Glucose	1258.62	NH_4Cl	2132.68
Maltose	1124.64	Urea	1450.02
CMC-Na	2081.64	$(NH_4)_2SO_4$	1769.02
Starch	671.66	$NaNO_3$	1807.3
Sucrose	907.72	NH_4NO_3	2177.34

3.2 L9 (34) Orthogonal Experiment

According four factors and three levels to design the orthogonal experiment. Four factors contain carbon source (CMC-Na), nitrogen (NH_4NO_3), surface active agent (Tw-80) and ascorbic acid, each factor selected three different addition levels. Different levels of the design in Table 2.

Table 2. The factors, levels of orthogonal design table L_9 (3^4)

Levels	CMC-Na (A)	NH_4NO_3 (B)	Tw-80 (C)	Ascorbic acid (D)
1	0.4%	0.15%	0.015%	0.025%
2	0.5%	0.20%	0.020%	0.020%
3	0.6%	0.25%	0.025%	0.015%

According to results as shown in Table 3 by variance analysis, CMC-Na is the most significant on the enzyme production, followed by Tw-80, nitrogen NH_4NO_3, and ascorbic acid least. Final test results from the orthogonal optimization scheme are

Table 3. L_9 (3^4) orthogonal test analysis

Experiment No.	A	B	C	D	Activity (μmol/ (g·h))
1	1	1	1	1	2088.02
2	1	2	2	2	2324.08
3	1	3	3	3	2394.26
4	2	1	2	3	3364.02
5	2	2	3	1	3466.10
6	2	3	1	2	2821.72
7	3	1	3	2	3587.32
8	3	2	1	3	3631.98
9	3	3	2	1	3485.24
K1	2268.787	3013.120	2847.240	3013.120	
K2	3217.280	3140.720	3057.780	2911.040	
K3	3568.180	2900.407	3149.227	3130.087	
Range	1299.393	240.313	301.987	219.047	
Optimum scheme	A3	B2	C3	D3	

$A_3B_2C_3D_3$, which the test factors on the sizes of the enzyme production were: CMC-Na> Tw-80> NH_4NO_3> Ascorbic acid. Test that when the inoculation CMC-Na0.6%, NH_4NO_3 0.20%, concentration of (Tw-80)0.025%, concentration of ascorbic acid 0.015%, the cellulose enzyme production was the best. The experimental activity measured up to 3874.34μmol / (g·h).

3.3 Different Culture Time on Biomass and Cellulose Activity in Aerobic and Anaerobic

Aerobic incubation in 250ml Erlenmeyer flask with 50ml liquid fermentation medium for enzyme production; Anaerobic sealed with rubber stoppers bottle by Hungate anaerobic techniques with 50ml anaerobic liquid fermentation medium, inoculum seeds liquid at the conditions of 37℃, 200r/min for cultivation, measured the biomass and the enzyme activity every 12h, the results shown in Fig 1.

Aerobic culture: large quantity of mycelium growing in the early aerobic culture, the enzyme activity relatively low, but rapidly increased after 72h, when at 96h reached the peak of 2336.84μmol /(g·h) and the biomass also reached saturation at this time, then into the stable phase of fermentation, and declined with the culture time prolonged after 96h. However, anaerobic culture in the training before 108h, the biomass and enzyme activity continued rise, biomass saturation at 108h, activity reached its peak 2011.46μmol /(g·h) at 120h, further culture found the biomass and enzyme activity significantly decreased, compared the effects of enzyme production of the strain in aerobic and anaerobic conditions at different time found that under aerobic conditions the start-up time of enzyme production is earlier than under anaerobic conditions, cell growth saturation stage is also ahead of schedule.

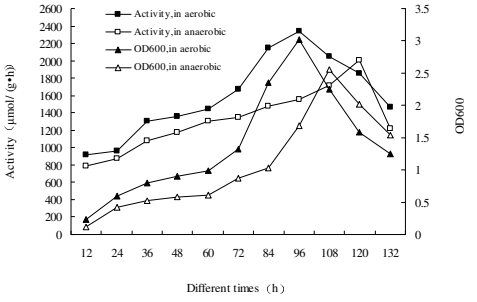


Fig. 1. Effects of different culture time on biomass and cellulose activity in aerobic and anaerobic

3.4 The Effect of Culture Temperature on Cellulose Activity in Aerobic and Anaerobic

Among many environmental factors, the most significant influence on microbial growth is temperature, choice different culture temperature 25 ℃, 28 ℃, 32 ℃, 37 ℃, 40 ℃ and 45 ℃ for enzyme assay in aerobic and anaerobic conditions, the results shown in Figure 2. Found the optimum temperature of the aerobic bacteria for activity

was 37 ℃, the enzyme activities at 40 ℃, 45 ℃ was relatively low, possibly because high-temperature made the important part of microbial protein, nucleic acid suffer irreversible damages, had a negative impact on the body and inhibit cell growth.

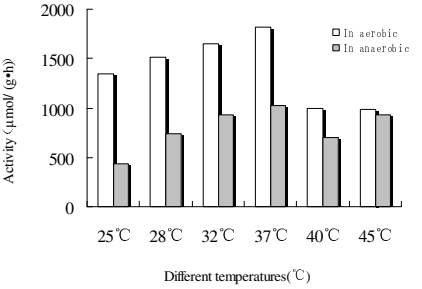


Fig. 2. Effects of different temperature on cellulose activity in aerobic and anaerobic

3.5 The Effect of Different Culture pH on Cellulose Activity in Aerobic and Anaerobic

Choice different initial culture pH6.0,7.0,8.0 at 37 ℃, 200r/min in aerobic and anaerobic conditions, measured activity after 96h, the results shown in Figure 3: the activity is the highest when initial pH=7, in aerobic conditions to 1992.32μmol/(g·h) and anaerobic conditions can achieve 1360.70μmol/(g·h), can be seen the cellulose activity production of strain has a strong stability in the neutral condition, indicates the strain has more appropriate growth in the moderate environment, can promote the production of enzyme activity.

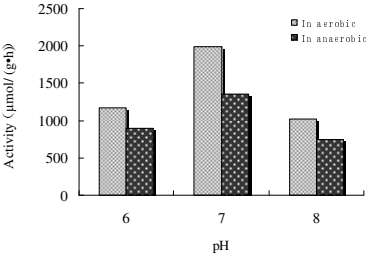


Fig. 3. Effects of different pH on cellulose activity in aerobic and anaerobic

3.6 The Effect of Different Inoculation on Cellulose Activity in Aerobic and Anaerobic

Aerobic conditions, when the inoculation volume 4% the cellulose production is the highest ,while the inoculation at 6%-10%, even if the inoculums increased cellulose production has not increased significantly, probably because when the inoculation too large, mycelium lack of nutrition leads the enzyme production decreased; However, the

highest enzyme activity in anaerobic inoculation volume is 8%, indicates that increased inoculation conducive to spore germination, mycelium growth, the rational use of nutrient medium and to form a higher yield activity under anaerobic conditions.

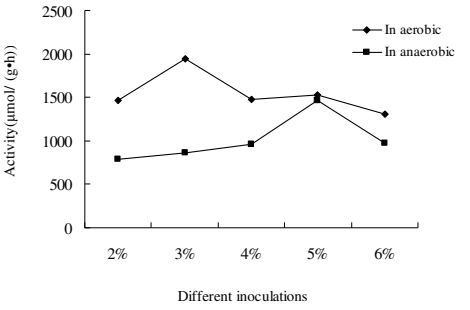


Fig. 4. Effects of different inoculums' size on cellulose activity in aerobic and anaerobic

4 Conclusions

A facultative anaerobic cellulose degradation bacteria screened from bovine rumen, as the strain-based by the single factor experiment, initially identified its optimum carbon source was CMC-Na, nitrogen source was NH_4NO_3.The orthogonal analysis, the inoculation CMC-Na0.6%, NH_4NO_3 0.20%, concentration of (Tw-80)0.025%, concentration of ascorbic acid 0.015%, the production of cellulose enzyme activity can be best achieved 3874.34μmol/(g·h). And compared the conditions of fermentation process in aerobic and anaerobic conditions, found that the optimal incubation time 96h (aerobic conditions), 120h (anaerobic conditions),the optimum culture temperature is 37 ℃, the optimum initial medium pH is 7, the optimum inoculation of 4% (aerobic conditions), 8% (anaerobic conditions). Based on the above optimum conditions, CMC activity under aerobic conditions to 4210.32μmol/(g·h), CMC activity under anaerobic conditions can be achieved 3415.76μmol/(g·h).

Previously reported is mainly about the producing enzyme activity of cellulose degrading bacteria in aerobic conditions, while above-obtained facultative anaerobic cellulose degradation bacteria under anaerobic conditions can also produce high enzyme activity, No others the same reported, the current activity of the strain is not very high compared to high enzyme activity cellulose has been reported by the Breeding at home and abroad, through mutation or genetic engineering methods can improve their enzyme activity and with potential applications, this study is in progress.

Acknowledgment. This work was supported by School of Environment, Tsinghua University. Wuxi FengLu Environmental Protection Technology Co., Ltd. and The Wuxi "530" plan.

References

[1] Garg, U., Kaur, M., Garg, V., Sud, D.: Removal of hexavalent chromium from aqueous solution by agricultural waste biomass. Journal of Hazardous Materials 140, 60–68 (2007)
[2] Lin, C., Hung, W.: Enhancement of fermentative hydrogen/ethanol production from cellulose using mixed anaerobic cultures. International Journal of Hydrogen Energy 33, 3660–3667 (2008)
[3] Zhang, D., Lax, A., Raina, A., Bland, J.: Differential cellulolytic activity of native-form and C-terminal tagged-form cellulase derived from Coptotermes formosanus and expressed in E. coli. Insect Biochemistry and Molecular Biology 39, 516–522 (2009)
[4] Chong, M., Sabaratnam, V., Shirai, Y., Hassan, M.: Biohydrogen production from biomass and industrial wastes by dark fermentation. International Journal of Hydrogen Energy 34, 3277–3287 (2009)
[5] Martins, L., Kolling, D., Camassola, M., Dillon, A., Ramos, L.: Comparison of Penicillium echinulatum and Trichoderma reesei cellulases in relation to their activity against various cellulosic substrates. Bioresource Technology 99, 1417–1424 (2008)
[6] Stieglmeier, M., Wirth, R., Kminek, G., Moissl-Eichinger, C.: Cultivation of anaerobic and facultatively anaerobic bacteria from spacecraft-associated clean rooms. Applied and Environmental Microbiology 75, 3484 (2009)
[7] Shipovskaya, A., Shmakov, S., Kazmicheva, O., Shchyogolev, S.: Optical activity of the anisotropic phases of cellulose acetates. Journal of Polymer Science Part B: Polymer Physics 47, 1605–1615 (2009)
[8] Li, Y., Irwin, D., Wilson, D.: Processivity, substrate binding, and mechanism of cellulose hydrolysis by Thermobifida fusca Cel9A. Applied and Environmental Microbiology 73, 3165 (2007)
[9] Jung, H., Yoon, H., Park, W., Choi, C., Wilson, D., Shin, D., Kim, Y.: Effect of sodium hydroxide treatment of bacterial cellulose on cellulase activity. Cellulose 15, 465–471 (2008)

Research and Application of Furnace Exception Forecasting by Expert System Based on Fuzzy Reasoning

Jianyong Li[1], Yanqiu Yang[2], Rui Li[1], Li Tian[1], and Zhenyi Wu[1]

[1] College of Chongqing Communication, 400035 Chongqing, China
[2] Chongqing Public Security Bureau, 401147 Chongqing, China
ljy1970000@126.com

Abstract. Based on the blast furnace expert knowledge and data analysis, the knowledge base can increase and delete online is established to determine the heat, the cool, the pipe trip, the hanging and the collapse materials, achieving furnace conditions prediction in the blast furnace. according to the iron expert knowledge and fuzzy theory, the blast furnace heat status, the mechanism of abnormal furnace conditions,the knowledge acquisition and the design of reasoning machine are in-depth study. With the Visual C + + and SQL Server2000 as the development tool, the blast abnormal furnace condition forecasting expert system software is developed. On-site assessment indicate that the system easy to use, fast, high hit rate,and strong practical.

Keywords: Furnace, furnace abnormal, fuzzy, expert system, Visual C++.

1 Introduction

The blast furnace expert system is a class of complex computer software system, which according to expert systems in computer science theory, relying on excellent experience and knowledge of the knowledge provided from experts ,and with logical inference reasoning and judgments of the system, To simulate blast furnace smelting process and solve the problems arising from it [1].Blast furnace smelting process is conducted in a closed container following extremely complex physical and chemical phenomenon, even the furnace conditions can not be directly observed. Reduce the energy consumption of blast furnace process, achieve stability and optimization of iron production, require timely and accurate judgments and forecast blast furnace conditions, take a variety of adjustment measures timely and appropriately.

2 System Components

Expert system is mainly composed of data acquisition module, data smoothing module, parameter feature extraction module, knowledge base management module, inference engine module, interface module and other function modules in this paper. The basic framework of the system shown in Figure: Expert system development tool is divided into two types of skeleton and language.In contrast, Language of expert systems [2] developed a wide range of tools for structural changes, which performance style is

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 493–498.
springerlink.com

flexible and can be adapted to more areas. Many of these linguistic tools are developed by C / C + + language. which is the main tool for Windows programming[3]. have closely integrated with Windows, a powerful class library support and transformation capabilities, efficient speed,etc. SQL Server is relational database systems (DBMS)developed and promoted by Microsoft [4], and it is currently the most popular medium-sized database; therefore, In this paper, Visual C + + and SQL Server database as a tool for development and application of systems.

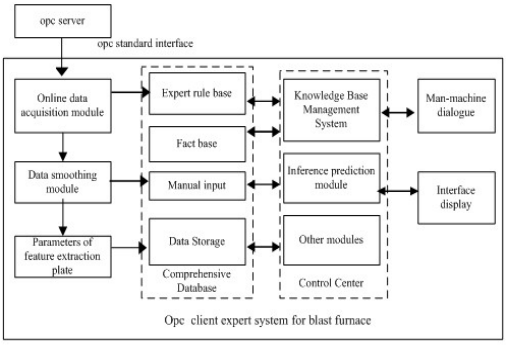


Fig. 1. Frame of the 4# BF expert system

3 System Menu

Starting the system software required to enter a user name and password. System based on the user's user name and password,testing the legitimacy of users and permissions, after legitimate user logs into the system of main interface,which including system, parameter settings, view, abnormal furnace conditions forecast, graphs and help menu. System Menu was composed of adding users, deleting users and recededing submenu. Parameter setting menu includes an expert rule base management system, the fact database management systems and artificial parameter input submenu. The status of the View menu includes oven to heat, cold furnace conditions and center pipe to the furnace conditions of various plans and rules of inference prediction query node configuration submenu. Abnormal conditions furnace furnace conditions to forecast the menu includes hot stove to cool conditions, the center pipe and hanging various furnace conditions and forecast total furnace conditions forecast submenu.

4 Data Sheet and Database

The expert system is developed based on fuzzy reasoning produce type rules expert system.

The database to store mainly includes rule base, fact base, the original database, pre-database and other parts, where the first focuses on the design of the database storage structure, the data dictionary definition and in practice use.

4.1 Data Sheet

① the original data sheet and pretreatment data sheet and corresponding field definition. Site operation data is stored in the data sheet $Data\_Rc$ $Data\_Rp$ $Data\_Rt$; Pretreatment of field data sheet the same with $Data\_Rc$.

② the point of parameters correspond the name of the data sheet. the named parameters are used to Running in the system, in order to link make the user to know the selected point on the representative parameters; set up the sheet $Para\_Fidld\_EXPL$.

③ the data sheet of rule base. Rules are the core of the knowledge base, fuzzy reasoning is the main basis for computing [5]. Rule base is mainly used to store various furnace conditions all the rules of reasoning. In the specific implementation rules with the data sheet for storage.

④ the module of feature extraction to complete the different parameters of the eigenvalue calculation, storage. The stored data sheet $Para\,\mathrm{Re}\,asoniong$.

⑤ the configuration sheet of rule reasoning figure the leaf nodes. According to reasoning figure, the root node, the value of the credibility of the intermediate nodes must be started by the layers of leaf nodes derived reasoning. If in the "collapse of material" logic diagram, the " the situation of air flow " is an intermediate node, "large fluctuations in the air" is a leaf node, the system used to represent the membership.

4.2 Base of Expert Knowledge and Management System

Expert knowledge is an important database in "expert system"; knowledge base management system mainly for rules, facts and other knowledge management operations [6]. Expert system to achieve the data collection, storage, processing, analysis and so on. Knowledge management system is mainly to complete the expert knowledge of storage, the rectification work. The process of knowledge stored, the various sheets are associated. Mid-leaf node node name of the configuration table to set the field only from the knowledge base management system to select, and knowledge management systems from the rule base is automatically listed in the rule base of the leaf nodes need to be configured. Rules leaf node configuration is part of knowledge base management system. Data in "Leaf node parameters configured rules list" from the LeafLib, "not configuration parameter list of a leaf node " was not configured in the rule base table of the leaf nodes, "rules leaf node" data source option in all the leaf nodes of sheet RuleLib. "Corresponds to the parameters" for the ParaName and Parameaning field records in Para Reasoning. "Feature selection" for the ParaReasoning definition of the meaning of the characteristic parameters. "Ensure Settings" button to complete the rules of the characteristic parameters of leaf node selection, membership function selection and threshold setting storage capabilities. Interface shown in Figure 2.

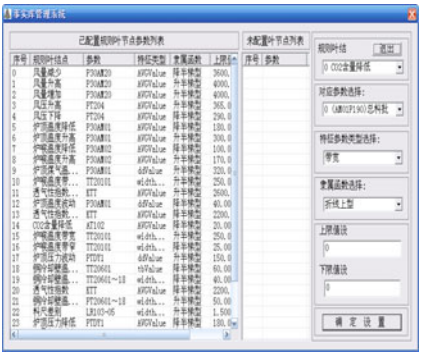


Fig. 2. Procedure interface of setting the leaf node of rule reasoning

5 The Realization of the Reasoning Machine

This design of the reasoning machine first establish an export target of the reasoning given tree (that is unusual given the name of furnace abnormal, in the search of records relevant rules from the rule base target, and rules stored in a structured variable node), later, adopt a positive way of reasoning, the reasoning tree node starting from the lowest layers of reasoning, update fact base. System using hybrid reasoning agency, the first goal of the target-based approach, in the reasoning process of the establishment of an export target of the reasoning given tree, and then the knowledge base of rules for forward reasoning. The expert system In this paper, rule tree was created based on the target search process, and ways to traverse the definition node number, destination node number is defined as 0.

Based on the information stored, reasoning process established as follows: ① the establishment of rules stored record and position pointer to the last node. ② read the node, and determine the number of the node's children. ③ If the number of child nodes to 0, from the fact that the library to read the node's credibility and to give Com_cf. Otherwise, according to Child_num and Child_index determine the child node information, call the fuzzy reasoning sub-module, calculate the node's credibility value. Then, set Com_flag effective,the flag is 1. ④ position pointer is advanced by one. • Check the position pointer is empty, if yes then complete reasoning. Otherwise, turn step (2).

6 System Software Running and Forecast the Results

After a long site running debug parameters to determine the parameters of the leaf node, features of the membership function and threshold. Table 1 shows the characteristics of some of the parameters and threshold parameters.

Table 1. Character and threshold of leaf-node of parameter

Leaf_Name	ParaType	Max_Para	Min_Para
Reduce air	AVGValue	3600	1000
Increased flow	AVGValue	4000	3000
Top gas temperature there still	ddValue	220	100
Furnace temperature throat with width	widthValue	150	50
Permeability index rise	AVGValue	2600	1750
Permeability index decreased	AVGValue	2000	1650
Top pressure fluctuations	ddValue	150	20
Top temperature fluctuations	ddValue	40	20
Air volume rising sharply	dtValue	450	200
Air volume fell sharply	dtValue	500	250
Copper cooling wall mean temperature	AVGValue	60	25

System with three ranges based on guidance section operating, they are CF in the [0,0.35] is a safe area; in (0.35,0.65] when warning area; in (0.65,1] when the danger zone. 3 2010-01-08 10:02:00 when the system is given a comprehensive prediction results:

Fig. 3. Map of all-around forecast

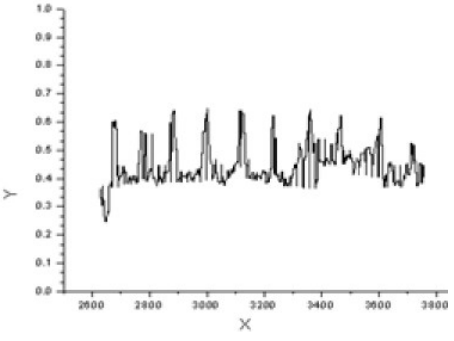


Fig. 4. Forecast result of BF warming

After a long run in the field showed better prediction results, Figure 4(X-axis is time, Y-axis is the ore to the furnace heat forecasting;), Figure 5 (X-axis is time, Y-axis is the expected forecast bias)shows the results of two exceptions furnace abnormal conditions, the CF level of confidence (from 2010-01-07 09:44:43 to time 2010-01--01-08 09:04:12):

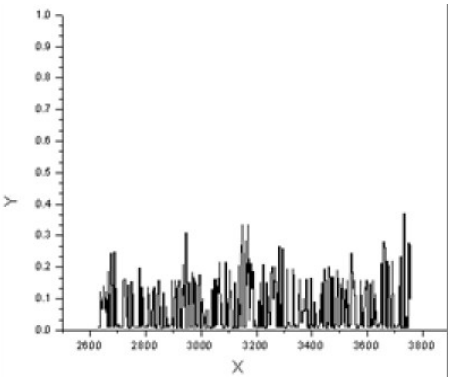


Fig. 5. Forecast result of fringe material

7 Conclusion

Blast furnace abnormal conditions forecasting expert system based on fuzzy reasoning is put forward. Through the iron-making process, the abnormal situation analysis, data smoothing filter processing, parameter correlation analysis, parameter eigenvalue calculation,reasoning to determine the credibility of a prerequisite,the establishment of inference engine and knowledge base,the use of visualization technology, and finally based on Visual C + + and SQL Server2000, the hot levels of blast furnace and abnormal stove condition forecasting expert system is developed, achieving in real-time monitoring. The system is already running in the blast furnace system seamlessly , integrating into an organic whole.

References

1. Yi, C.Q., Yi, H.: Artificial intelligence and expert systems. China Water Power Press, Beijing (2002)
2. Ye, S.: Blast furnace mathematical model to improve its practical application. The World's Iron and Steel (5), 13–21 (2003)
3. BF technical operation (internal data). Ironworks Panzhihua New Steel & Vanadium Co., Ltd. (2006)
4. Ebrahim, R.: Fuzzy logic programming. Fuzzy Sets and Systems 117, 215–330 (2001)
5. Zadeh, L.A.: Fuzzy sets. Information and Control 8, 338–353 (1965)
6. Ji, M., Xu, J.S.: Laota Luji Steel Plant Blast Furnace Expert System Lage function, operation and effect. Foreign Steel (4), 13–16 (1997)

Reasoning in Air Traffic Control Using Prolog

Dancheng Li, Zhiliang Liu, Cheng Liu, Binsheng Liu, and Wei Zhang

Software College, Northeastern University,
Northeastern University, NEU,
Shenyang, China
ldc@mail.neu.edu.cn

Abstract. With the acceleration of aircrafts, the situation of air traffic becomes increasingly complex. So the next generation air traffic control system should be more intelligent and flexible to deal with the changing environment. Logic reasoning, as a very important part of Artificial Intelligent, should be taken into consideration seriously in the air traffic control field, since logic reasoning can not only implement the intelligence of the system but it can also make quick and accurate prediction. Additionally, prolog, as a logical programming language, is quite popular and powerful in implementing logic reasoning. So in this paper we propose to use prolog to reason various situations in the field of air traffic.

Keywords: Air traffic control, reason, prolog, prediction.

1 Introduction

Anticipatory reasoning reacting systems natively focus on predicting issues and taking actions beforehand. A reasoning component is necessary in the system, where a system is considered as a smart agent who has to make decisions by its own to achieve its goals, and it also plays an important role in predicting future events for the system [1, 2].

Therefore, we need to choose a fundamental programming language to implement the reasoning component. We consider that this language should not only possess the ability of exactness and efficiency in predicting, but it should also be transplantable and compatible.

In this paper, we propose to use prolog as the fundamental programming language for the reasoning component of the Anticipatory reasoning reacting system. Prolog enables an entirely different approach to computer programming, using a declarative rather than procedural method. This language is compatible with varies of programming languages and also widely used in reasoning.

In the next section, we give some basic concepts in the research field of air traffic and reasoning. Then, in the third section, we introduce the idea of using prolog as the programming language. And through an example, we describe the working flow of prolog processing. In the last section, we make a conclusion of our previous work and describe the view of our future work.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 499–505.
springerlink.com

2 Logic Reasoning in Air Traffic Control

2.1 Basic Concepts

Air Traffic Control System is a system which allows aircrafts to detect other aircrafts in their vicinity so that they can be aware of them and it also allows an aircraft to maintain a particular spacing with respect to a selected aircraft to alleviate the load of the ground air traffic controllers [3].

Reasoning is the process of drawing new conclusions from given premises, which are already known facts or previously assumed hypotheses (Note that how to define the notion of "new" formally and satisfactorily is still a difficult open problem until now). Therefore, reasoning is intrinsically ampliative, i.e., it has the function of enlarging or extending some things, or adding to what is already known or assumed [4].

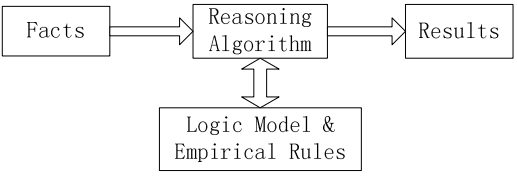


Fig. 1. Structure of predicting model through reasoning

Prediction is the process to make some future event known in advance, especially on the basis of special knowledge. The structure of the predicting model is shown in figure 1. For any prediction, both the predicted thing and its truth must be unknown before the completion of prediction process. Since reasoning is the only way to draw new conclusions from given premises, there is no prediction process that does not invoke reasoning [5], and in figure 2 the prediction model is shown.

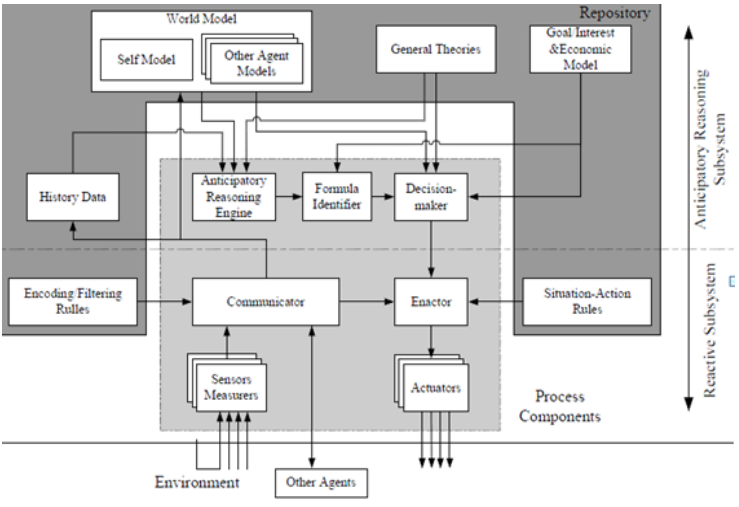


Fig. 2. Prediction Model

2.2 Why Prolog?

A long-ago interest in AI led to our awareness of Prolog and its ability to manipulate symbols while emphasizing a strict logical plan. Prolog enables an entirely different approach to computer programming, using a declarative rather than procedural method. The declarative method doesn't require the programmer to supply the computer with step-by-step instructions. Instead, we must present rules to satisfy and data to be tested against those rules. The Prolog "engine" then works to find relationships among the data that are consistent with the rules. Already this sounds a lot like how we solve a physics problem [6, 7].

3 Reason Using Prolog

Prolog enables an entirely different approach to computer programming, using a declarative rather than procedural method. The declarative method doesn't require the programmer to supply the computer with step-by-step instructions. Instead, we must present rules to satisfy and data to be tested against those rules [8].

In the following, we describe the mechanism of prolog processing through an example, and compare the results deduced manually and by prolog, according to which we can conclude that prolog possesses the ability of reasoning as good as human.

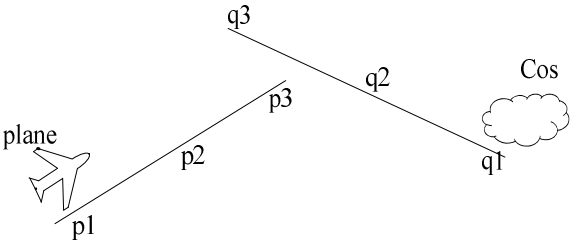


Fig. 3. Scene of air traffic

As is shown in figure 3, there is an airplane and a storm in our scene and the plane is flying from point p1 to point p3. Meanwhile the storm is flying from point q1 to point a3.

3.1 Definition of Empirical Theorem

Based on the knowledge and experience of air traffic controllers, we define some logical predicates and empirical theorems used for reasoning and prediction as follows:

Examples of predicates:

(1) DisEq(p1, p2, d) means "the distance between p1 and p2 is equal with d",
(2) Equal(m, n) means "m is equal with n",
(3) MulSp(o1, o2, n)means "o1's speed is n times of the o2's speed",

Examples of empirical theorems:

(1) $\forall p_1 \forall p_2$ Arc(p1, p2)⇒Reachable(p1, p2)
If point p1 is connected with p2, then we say that p1 is reachable to p2.

(2)∀p1∀p2∀p3(Arc(p1,p2)∧Reachable(p2,p3))⇒Reachable(p1,p3)
If point p1 is reachable to p2 and p2 is reachable to p3, then point p1 is reachable to p3.

(3)∀o1∀p1∀p2(Arrive(o1,p1)∧Reachable(p1,p2))⇒F(Arrive(o1,p2))

The main idea of the reasoning function is that we go through all the facts and theorems to get all the pair of reachable points, and then match the current location point of the moving object with the first point of each point pair. If the match is successful, then we can get the second point to generate the result according to the form of the result.

In the following, we give some facts of the current situation, given facts are as follows:

Arrive(AIRPLANE, P1), Arc(P1, P2), Arc(P2, P3)
Arrive(COS, Q1), Arc(Q1, Q2), Arc(Q2, Q3)

From the facts above, we can get the following conclusions by reasoning:

1) Reachable(P1,P2), Reachable(P1,P3), Reachable(P2,P3),
2) Reachable(Q1,Q2),Reachable(Q1,Q3),Reachable(Q2,Q3),
3) F(Arrive(AIRPLANE,P2)), F(Arrive(AIRPLANE,P3))
4) F(Arrive(COS,Q2))), F(Arrive(COS,Q3)))

3.2 Reason by Prolog

Since the reasoning processes above are all executed by hand, the results need to be examined through computer. And therefore, we will go through the scene in prolog and show how the prolog handles this example.

First, we input the facts and theorems mentioned in section A above according to the syntax of prolog as is shown in figure 4.

```
1 ?- listing.

arrive(airplane, p1).
arrive(cos, q1).

arc(p1, p2).
arc(p2, p3).
arc(q1, q2).
arc(q2, q3).

reachable(A, B) :-
        arc(A, B).
reachable(A, C) :-
        arc(A, B),
        reachable(B, C).

f_arrive(A, C) :-
        arrive(A, B),
        reachable(B, C).
true.
```

Fig. 4. List of facts and theorems in prolog

As we can see, the input of facts in the prolog file is exactly the same as in the hand reasoning process. Second, we ask prolog to deduce the conclusion of "reachable" for us, which means the relations of two connected point. The results are shown in figure 5.

```
2 ?- reachable(X,Y).
X = p1,
Y = p2 ;
X = p2,
Y = p3 ;
X = q1,
Y = q2 ;
X = q2,
Y = q3 ;
X = p1,
Y = p3 ;
X = q1,
Y = q3 ;
false.
```

Fig. 5. List of reachable points

The specific data evaluated to Xs and Ys are the connected points we want, which means exactly the same as the conclusions, including line 1) and 2), deduced manually.

At last, we input "f_arrive(O,P)." to ask prolog to get us some results about what objects will arrive at which point in the future. The results are shown in figure 6.

```
3 ?- f_arrive(O,P).
O = airplane,
P = p2 ;
O = airplane,
P = p3 ;
O = cos,
P = q2 ;
O = cos,
P = q3 ;
```

Fig. 6. Relation of future arrive between points and objects

The specific data evaluated to Os and Ps are the related Objects and points we want, which means exactly the same as the conclusions , including line 3) and 4), deduced manually.

3.3 The Process in Prolog

The results of prolog reasoning shown in section B indicate the correctness of our hand reasoning process, and here we will describe the process of how prolog get the "f_arrive" results.

At first, prolog reads the knowledge base, and tries to match f_arrive(O,P) with either a fact, or the head of a rule. It searches the knowledge base top to bottom, and carries out the matching. Here is only one possibility: it must match f_arrive(O,P) to the head of the rule "f_arrive(A, C) :-arrive(A, B), reachable(B, C)."

When prolog matches the variable in a query to a variable in a fact or rule, it generates a brand new variable to represent that the variables are now sharing. So the original query now reads: f_arrive(_G335,_G336) and prolog knows that f_arrive(_G335,_G336):-arrive(_G335,_G337), reachable(_G337, _G336).

So Prolog replaces the original query with the following list of goals: arrive(_G335,_G337), reachable(_G337, _G336).

That is, our original goal is to prove f_arrive(O,P). When matching it with the head of the rule in the knowledge base O and P and the internal variable _G335 and _G336 are make equal and we are left with the goals arrive(_G335,_G337), reachable(_G337, _G336).

Now, whenever it has a list of goals, prolog tries to satisfy them one by one, working through the list in a left to right direction. The leftmost goal is arrive (_G335,_G337), which reads: 'I want two objects with property arrive'. Prolog tries to satisfy this goal by searching through the knowledge base from top to bottom. The first thing it finds that matches this goal is the fact arrive(airplane,p1), and we are left one more goal to go. When matching arrive(_G335,_G337) to arrive(airplane,p1), O is instantiated to airplane. This applies to all occurrences of variables in the list of goals. So the list of remaining goals is: reachable(p1, _G336).

The fact reachable(p1,p2) is in the knowledge base, so the next goal we have to prove is satisfied too and P is instantiated to p2.

If we type " ; " to ask prolog for another solution, it forces Prolog to backtrack to the last choice point, to try and find another possibility. In this example, the last choice point is at _G336 since there is also reachable(p1,p3) deduced according to the first rule.

The result above is exactly the same as the second conclusion prolog has got. And also the third and fourth conclusions are deduced in the same way. So finally we can achieve the conclusion tree as is shown in figure 7, which contains 4 results.

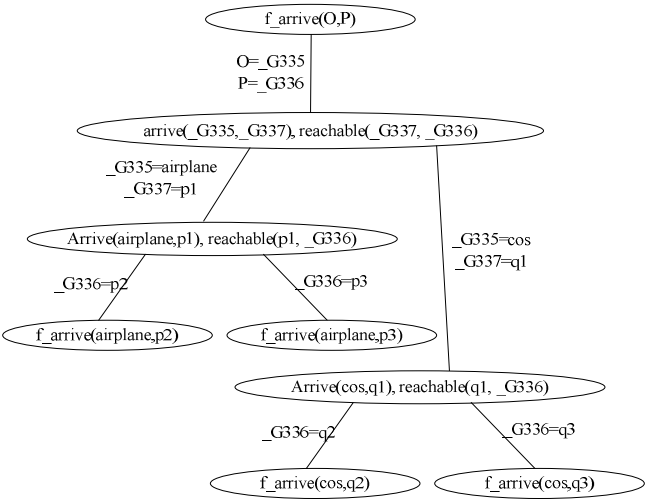


Fig. 7. Final results

In conclusion, through the process of prolog, we can obviously see that prolog works very well in the air traffic control field and it can not only handle the process of matching directly, but it can also handle multi-results issues.

4 Concluding Remarks

In the previous phase of our research work, we proposed to use prolog to implement the reasoning part of an anticipatory reasoning reacting system. And after analyzing the mechanism of prolog, we describe the reasoning process of prolog through an example.

However, the application of prolog is still not sufficient. In our future work, we will focus on combining prolog with other programming languages to implement a better reasoning mechanism.

Acknowledgment. This research work is supported by "Natural Science Foundation of LiaoNing Province" (20092006).

References

1. Cheng, J., Goto, Y., Kitajima, N.: Anticipatory Reasoning about Mobile Objects in Anticipatory Reasoning-Reacting Systems. In: Dubois, D.M. (ed.) Computing Anticipatory Systems: CASYS 2007, AIP Conference Proceedings, to appear, American Institute of Physics (2008)
2. Cheng, J.: Temporal Relevant Logic as the Logical Basis of Anticipatory Reasoning-Reacting Systems. In: Dubois, D.M. (ed.) Computing Anticipatory Systems: CASYS 2003 - Sixth International Conference, AIP Conference Proceedings 718, American Institute of Physics, Melville, NY, pp. 362–375 (2004)
3. Gilbert, G.A., Hobbs, J.: Air traffic control system. US Patent (1975)
4. Cheng, J.: A Strong Relevant Logic Model of Epistemic Processes in Scientific Discovery. In: Kawaguchi, E., Kangassalo, H., Jaakkola, H., Hamid, I.A. (eds.) Information Modelling and Knowledge Bases XI, pp. 136–159. IOS Press (2000)
5. Cheng, J.: Adaptive Prediction by Anticipatory Reasoning Based on Temporal Relevant Logic. In: Proceedings of the 8th International Conference on Hybrid Intelligent Systems, Barcelona, pp. 410–416 (2008)
6. Bensky, T.J., Taff, C.A.: Computer-Guided Solutions to Physics Problems Using Prolog. Computing in Science & Engineering (2010)
7. Stoianov, A., Sora, I.: Detecting patterns and antipatterns in software using Prolog rules. In: Computational Cybernetics and Technical Informatics, ICCC-CONTI (2010)
8. Gilman, E., Sánchez, I., Saloranta, T., Riekki, J.: Reasoning for Smart Space Application: Comparing Three Reasoning Engines CLIPS, Jess and Win-prolog. In: 2010 IEEE 10th International Conference on 2010 Computer and Information Technology (CIT), pp. 1340–1345 (2010)

Comparison of Baseband Modulations in Visual Light Positioning

Yao Ji-Yun

Department of Electronic Engineering,
Tsinghua University,
Beijing, 100084, China
goodboyyjy@gmail.com,
yunsheng@public.wh.hb.cn

Abstract. In this paper, author examines the use of on-off keying (OOK) and pulse-position modulation (PPM) and differential pulse-position modulation (DPPM) for visual light positioning systems, presents expressions for the error probability and synchronization failure probability and power spectral density of OOK, PPM and DPPM, gives structure design to realize PPM and DPPM based on existing OOK systems. Finally, it is shown that for a given bandwidth, the performance of 2M-DPPM is better than the performance of M-PPM, while the system structure of DPPM is simpler.

Keywords: Modulation schemes, power efficiency, synchronization failure probability.

1 Introduction

LED is much better than existing incandescent in terms of high energy efficiency, long life expectancy, high tolerance to humidity and minimal heat generation. Recently a new type of white LED, white phosphorescent LED has been developed. This technology cuts down the cost of LED lighting considerably. An underway national program in Japan has already suggested that white LED deserves to be considered as a general lighting technology in the 21st century owing to electric power energy consumption [1].

The visible light communication is drawing increasing attention because of the decreasing cost of Light-emitting diodes (LEDs) and the fast switching of them. Many applications have been proposed. Among them, one of the most likely to be commercialized is positioning.

The global positioning system (GPS) is widely used as a mature technology, but in some conditions such as indoor, underground and places surrounded by buildings GPS signal cannot reach the objectire, so the convantional positioning is also needed. Many technologies have been proposed, such as Active Badge system from AT&T and Cambridge [2] and RADAR system from Microsoft [3]. But none of existing technology has been widely used because of many problems such as the cost and

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 507–514.
springerlink.com

reliability of the system. However visible light positioning system based on LED is much more commercially available and reliable. In fact Masao Nakagava from Keio University had made a product of LED light positioning [4]. On the other hand, if higher accuracy is needed, we have to use the differential three dimensional coordinates and relative locations to user of multi LEDs [5].

This paper compares the three widely used baseband modulation schemes in optical communication systems. The comparisons of the error probability, synchronization failure probability and power spectral density of OOK, PPM and DPPM I demonstrate that in a visible light communication system, for a given bandwidth, the performance of OOK is the worst and the performance of 2M-DPPM is relatively better than the performance of M-PPM, while the system structure of DPPM is simpler. Below, the author gives a possible solution to realize PPM and DPPM on existing OOK system and it will be described in five sections:

a. The need of indoor visual light communication positioning system
b. Comparison of the error probability and power spectral density between OOK, PPM, DPPM
c. Theoretical analysis and simulation of synchronization failure of OOK, PPM, DPPM
d. Structure design to realize PPM on OOK system
e. Discussion and conclusion

2 The Need of Indoor Visible Light Communication Positioning System

The bandwidth of commercial thin film high-power phosphorescent white LED can reach tens of MHz. The bandwidth is a limit to high speed wireless communication [6] but the information which needs positioning is far less, so the bandwidth efficiency is not a problem in visual light positioning system. However, if we want to achieve higher accuracy of positioning using multi-LED, the coverage area of one LED should be as large as possible. In most visible light communication system, the working distance from a LED to receiver is less than 10 meters but this is not enough for practise[7]. In other words, power efficiency is the most important criterion to assess a modulation method.

In positioning systems, bit rate is not a primary limit, but the power efficiency and bandwidth are. If we want to make the power efficiency as high as possible, we have to make the most use of widest bandwidth as we can. In a visual light communication system using LED, the bandwidth can be several MHz or tens of MHz [8]. Limited by this bandwidth, the maximum width of pulse is fixed to a length. And in most of existing system for visible light communication, using OOK modulation at the beginning, if we want make the best use of existing function blocks when we adopt PPM or DPPM technology, we have to use the same driving circuit and pulse generator.

For these two reasons, in this work, the author uses the same time slot to compare the OOK, PPM and DPPM. An example is shown in figure 1.

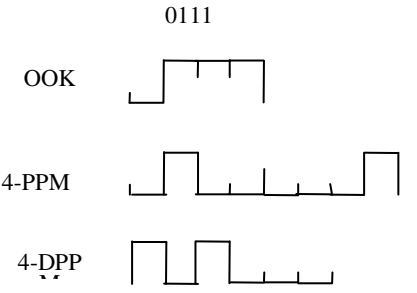


Fig. 1. Example of modulation pattern of '0111' in OOK, PPM, DPPM

3 Comparison of the Error Probability and Power Spectral Density between OOK, PPM, DPPM

Basic information [9] for the modulation schemes are as follows.

For OOK, the OOK transmission emits a rectangular pulse of duration $\frac{1}{R_b}$ and power intensity of $2P_t$. with a bandwidth B, the bit rate of OOK transmission is B.

For PPM, the length of a symbol of M-PPM is $\frac{\log_2^M}{R_b}$, the length of a pulse is $\frac{\log_2^M}{MR_b}$. The power intensity is MP_t. The bandwidth is approximately the inverse of the pulse width, so with a bandwidth B, the bit rate of PPM transmission is $\frac{\log_2^M B}{M}$.

For DPPM, the Length of a symbol of M-DPPM is $\frac{\log_2^M}{R_b}$. If possibilities of different symbols are the same for M-DPPM with no gap, the average length of DPPM pulse is $\frac{2\log_2^M}{(M+1)R_b}$, the power intensity is $(M+1)P_t/2$. If the M-DPPM is with T gaps between symbols, the average length is $\frac{2\log_2^M}{(M+1)R_b}+T$ and the power intensity is $[(M+1)/2+T]P_t$. For DPPM, the bandwidth is also approximately the inverse of the pulse width, so with a bandwidth B, the bit rate of DPPM transmission is $\frac{2\log_2^M B}{(M+1)}$.

The power requirement for the modulation schemes is as follows.

The BERs of OOK, PPM, DPPM for optical communication have been studied by Da-shan Shiu and Joseph M. Kahn [10].

The error probability for an F-chip OOK/PPM/DPPM packet is following respectively:

$$P_{F,OOK} \approx NQ\left(\frac{RP_t}{\sqrt{R_b N_0}}\right) \qquad P_{F,PPM} \approx NQ\left(RP_t\sqrt{\frac{M\log_2^M}{4R_b N_0}}\right)$$

$$P_{F,DPPM} \approx NQ\left(RP_t\sqrt{\frac{(M+1)\log_2^M}{8R_bN_0}}\right)$$

We can normalize all power requirements to the power required by OOK to send a 1-kB packet at average packet-error rate of 10^{-6}., then we can obtain the power requirement of the three modulation schemes.

The average optical power and bandwidth-requirements of OOK, PPM, and DPPM are displayed in figure 2.

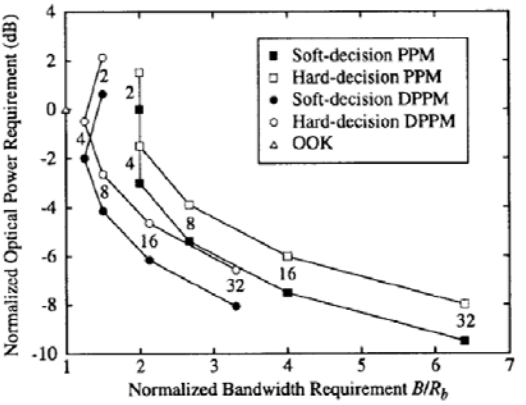


Fig. 2. Comparison of optical power requirement of OOK, PPM, DPPM [10]

This is a comparison of average-power efficiency and bandwidth-efficiency of PPM, DPPM, and OOK on non-dispersive channels. Both soft-decision and hard-decision detection are considered. The normalized power and bandwidth requirements of OOK are set to 0 dB and 1, respectively. The numbers represent the orders of M, the PPM, and DPPM.

Therefore, the power-efficient of both PPM and DPPM are much more than that of OOK, and the performance of 2M-DPPM is close to that of M-PPM. However, as the bandwidth efficiency is not an important issue in visual light positioning, we can not give a conclusion easily.

4 Theoretical Analysis and Simulation of Synchronization Failure of OOK, PPM, DPPM

All these three modulation schemes suffer from synchronization failures, Symbols may be decided with error if there are too long time intervals between adjacent pulses. However, DPPM does not require symbol synchronization as each symbol begins or ends with a pulse while PPM needs symbol synchronization to stop error spread. The theoretical analysis of synchronization of time slots is shown below.

We assume that the error of a time slot obeys normal distribution of T and σ. When an edge comes, the timer is synchronized. If no edge comes than the time of n-th

sampling point, will be as $T_n \sim \frac{e^{-\frac{(t-(n-0.5)T)^2}{2n\sigma^2}}}{\sqrt{2\pi n}\sigma}$. An interval with n time slot sampled without error needs exactly n sampling in the interval, thus the possibility is:

$$P_{C,n} = P\left(T_n < nT, T_{n+1} > nT\right) = \int_0^{nT} Q\left(\frac{(n-1)T - t}{\sigma}\right) \frac{e^{-\frac{[t-(n-0.5)T]^2}{2n^2\sigma^2}}}{\sqrt{2\pi n}\sigma} dt \cdot$$

For OOK, If a data stream is completely random, the possibility of synchronization failure is $P_{F,OOK} = \sum_{n=1}^{+\infty} (1 - P_{C,n}) \frac{1}{2^n}$. The average length of synchronized bits is $1/P_{F,OOK}$.

For M-PPM, If the M is larger than 2, the synchronization failure mostly occurs in even zeros. $P_{F,PPM} \approx \sum_{n=1}^{M-1} \frac{n}{M^2}(1 - P_{C,n}) + \sum_{n=M}^{2M-2} \frac{(2M - n - 1)}{M^2}(1 - P_{C,n})$

The average length of synchronized bits is $1/P_{F,PPM}$.

For M-DPPM, If the M is larger than 2, the synchronization failure mostly occurs in even zeros. We can obtain an approximate possibility of synchronization failure.

$$P_{F,DPPM} \approx \sum_{n=1}^{M-1} \frac{1}{M}(1 - P_{C,n})$$

The average length of synchronized bits is $1/P_{F,DPPM}$.

The comparison of possibility of synchronization between OOK, PPM and DPPM is as follows.

This is comparison of possibilities of synchronization failure of PPM, DPPM, and OOK in one time slot. The numbers represent the orders of M, the PPM, and DPPM. According to figure 3, M-PPM and 2M-DPPM has the similar performance of synchronization of time slots, which is far better than that of 2M-DPPM.

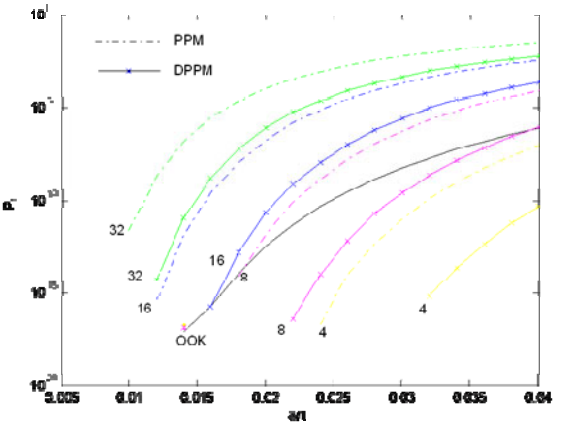


Fig. 3. Possibilities of synchronization failure of OOK, PPM and DPPM

Symbol synchronization: PPM needs symbol synchronization which can make the system complex and less bit rate efficient. There are many technologies for symbol synchronization, and most of them are inserting synchronous back-to-back symbols into the PPM sequence.

As an example, a synchronous symbol of 8-PPM is shown in figure 4.

Fig. 4. An example of synchronous symbol of 8-PPM

This symbol is not a valid symbol and it also has the best slot synchronization performance of all symbols so it can be used as a synchronous symbol.

5 Structure Design to Realize PPM and DPPM on OOK System

In most of systems proposed for visible light communication OOK is used because it's the simplest modulation schemes for realizing. So, a simple function block to change OOK modulation into PPM or DPPM can save lots of efforts to design a thoroughly new structure.

Structure design for realizing PPM&DPPM on OOK is shown in figure 5 and figure 6.

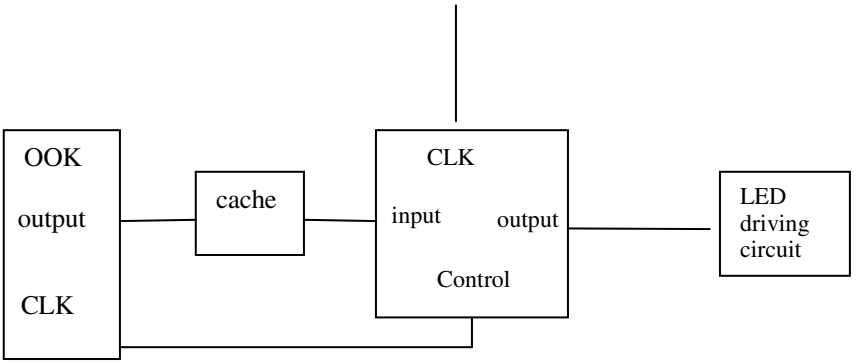


Fig. 5. T-converter

The transmit converter generates a clock signal $\log_2 M$ slower than original clock to control OOK encoder, and use a $\log_2 M$-M decoder and shift register to convert OOK signal to PPM signal. For PPM, the converter should also insert a synchronous symbol into every K normal symbols, and K can be random or fixed.

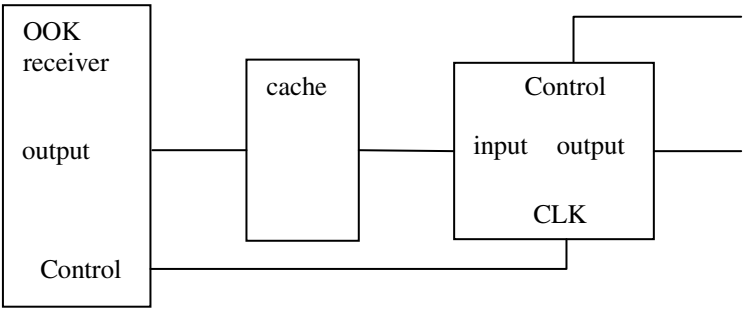


Fig. 6. R-converter

This receive converter generates a clock signal log2M slower than clock signal from receiver to control next blocks, and use a M-log2M encoder and shift register to convert PPM signal to OOK signal. For PPM, the converter should also detect the synchronous symbols and set the beginning of a symbol.

As synchronous symbols are inserted into original symbol sequences when convert OOK signals to PPM signals and symbol length of DPPM signals is not fixed, a cache is needed in both systems.

For PPM's R-converter, An symbol synchronization block is need, its possible structure is shown in figure 7.

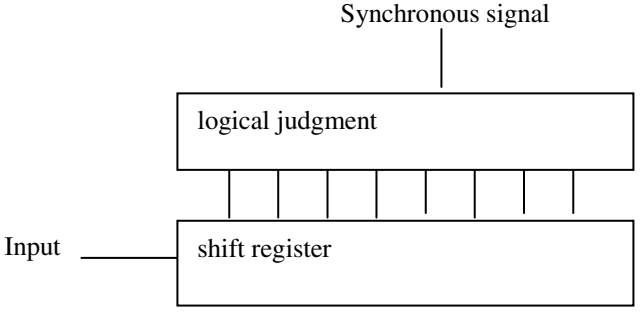


Fig. 7. A possible structure to generate synchronous signal

6 Conclusion

The author has compared the three common baseband modulation schemes used for the visible light positioning. To extend the coverage area of one LED transmitter, DPPM is the most promising modulation scheme and its system complexity is lower than that in PPM modulation scheme. If this modulation technology is adopted in visible light positioning system, the receiver can link many LEDs at the same time. In this way, the accuracy of positioning can be highly promoted.

References

1. Komine, T., Nakagawa, M.: Fundamental analysis for visible-light communication system using LED lights. IEEE Transactions on Consumer Electronics 50(1), 100–107 (2004)
2. Want, R., et al.: The Active Badge Location System. ACM Trans. Information Systems, 91–102 (January 1992)
3. Bahl, P., Padmanabhan, V.: RADAR: An In-Building RF-Based User Location and Tracking System. In: Proc. IEEE Infocom 2000, pp. 775–784. IEEE CS Press, Los Alamitos (2000)
4. Komine, T., Nakagawa, M.: Integrated system of white LED visible-light communication and power-line communication. IEEE Transactions on Consumer Electronics 49(1), 71–79 (2003)
5. Yoshino, M., Haruyama, S., Nakagawa, M.: High-accuracy positioning system using visible LED lights and image sensor. In: 2008 IEEE Radio and Wireless Symposium, January 22-24, pp. 439–442 (2008)
6. Vučić, J., Fernández, L., Kottke, C., Habel, K., Langer, K.-D.: Implementation of a real-time DMT-based 100 Mbit/s visible-light link. In: 2010 36th European Conference and Exhibition on Optical Communication (ECOC), September 19-23, pp. 1–5 (2010)
7. Vucic, J., Kottke, C., Nerreter, S., Habel, K., Buttner, A., Langer, K.-D., Walewski, J.W.: 125 Mbit/s over 5 m wireless distance by use of OOK-Modulated phosphorescent white LEDs. In: 35th European Conference on Optical Communication, ECOC 2009, September 20-24, pp. 1–2 (2009)
8. Grubor, J., Randel, S., Langer, K.-D., Walewski, J.W.: Bandwidth-efficient indoor optical wireless communications with white light-emitting diodes. In: 6th International Symposium on Communication Systems, Networks and Digital Signal Processing, CNSDSP 2008, July 25, pp. 165–169 (2008)
9. Mahdiraji, G.A., Zahedi, E.: Comparison of Selected Digital Modulation Schemes (OOK, PPM and DPPM) for Wireless Optical Communications. In: 4th Student Conference on Research and Development, SCOReD 2006, June 27-28, pp. 5–10 (2006)
10. Shiu, D.-S., Kahn, J.M.: Differential pulse-position modulation for power-efficient optical communication. IEEE Transactions on Communications 47(8), 1201–1210 (1999)

The Achieving of Gas Turbine's Gas Path Fault Criterion

Jian-hua Liu, Yong-bao Liu, Xiong-fei Zhao, and He Xing

1 403 Laboratory, College of Naval Architecture and Power,
Naval University of Engineering,
430033 Wuhan, P.R. China
452672622@qq.com

Abstract. In this paper the mathematical model of a certain kind of gas turbine was built based on parts thermodynamic characteristics and matching of components considering outside and boundary conditions. On the basis of this model, small deviation equations were produced by using small deviation method. Then, the standard fault model was acquired through introducing fault factors into the small deviation equations, which, further imported with actual operating data, educed the small deviation influence coefficient matrix. This matrix, combined with gas path fault variable criterion of performance parameters, would then ascertain the membership size relationship between typical gas path faults and measurement parameters. Finally, the typical gas path fault variable criterion of measurement parameters was deduced, which lays the foundation for gas path fault diagnosis of gas turbine.

Keywords: Fault criterion, small deviation method, fault factor, gas turbine, standard fault model.

1 Introduction

As a new power plant, gas turbine has many merits, such as good flexibility, compact structure, reliable and stable operation, high efficiency and so on. It is widely used in departments like space, navigation, land transportation and electric energy sector, and eventually it becomes the irreplaceable power equipment. Therefore, once breakdown, it will cause serious consequences to users. Meanwhile, in the fault diagnosing process, the same causation of the fault may show a variety of fault symptoms, and signs of the same fault may be caused by a variety of causations. The relationship between fault symptoms and causations is very complex; therefore, it is very hard to conduct a precise and quantitative analysis. For this reason, the maintenance of gas turbine in normal operation timely and the accurate diagnosis of gas turbine failure appear to be particularly important. Scholars have proposed many diagnosing methods in recent years, which have been further applied to fault diagnosis of gas turbine [1, 2, 3].

In the fault diagnosing process, a practical problem come into view is that how to determine the gas path fault criterion, for an effective confirmation of which would lay actual foundation for the application of various diagnostic methods. This paper adopts fault equation method to establish mathematical model of gas turbine, then combined with typical gas path fault performance parameters criterion of gas turbine,

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 515–521.
springerlink.com

the gas path fault diagnosis criterion was finally obtained, which lays a powerful foundation for the application of various diagnostic methods.

Gas path fault diagnosis came into existence in 1972. It is Urban [4, 5] who first applied fault equation method to gas path fault diagnosis of gas turbine. This method is mainly used in fault diagnosis of engine's gas path; therefore, it is called gas path analysis. Through introducing fault factor to the normal operation model, the fault model of gas turbine was obtained, which is:

$$B\delta y = C\delta \tilde{x} \tag{1}$$

Where $\delta \tilde{x}$ is the fault factor vector, δy is the response vector. B , C respectively correspond to the coefficient matrix of δy and $\delta \tilde{x}$. In the solving process, standard fault model was dealt with small deviation method to get small deviation fault equation, which then generated a fault coefficient matrix significantly connected with the gas path diagnosis, the matrix $B^{-1}C$. Then combined with typical gas path fault performance parameters criterion of gas turbine, the typical gas path fault measurement criterion was finally obtained, which therefore lays the foundation for gas path fault diagnosis of gas turbine.

2 The Establishment of Standard Fault Model

2.1 The Determination of Characteristic Equation of the Engine Parts

For acquiring fault criterion, the mathematical model of a certain kind of gas turbine should be built under normal working operation condition. In the process of modeling the crucial step is how to fit characteristic curve of each component in order to get the characteristic equation. Hence the mathematical model of the kind of gas turbine was established. It is important for accuracy to choose proper fitting order times n in fitting progress. In normal case, if the value of n is too small, it will led to a rough fitting result that would not meet the requirement of accuracy showed by plenty of documents and experiment results. But in the other extremity, the over-fitting that lead to large error between fitting and measured data will occur under this circumstance and produce high noises. Reference literature [6] proposed a method to use mean square deviation index $RMSE$ and test index Q_2 to test fitting results and determine the fitting order times n .

(1) The Mean Square Deviation Index $RMSE$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{N}} \tag{2}$$

Where: N is the number of observation sample ; when it closer to 0, the effect of fitting is better

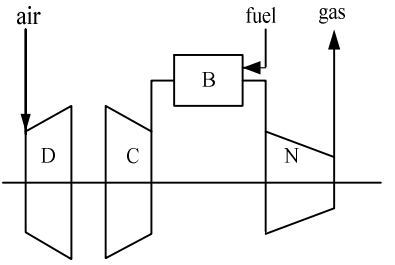


Fig. 1. The structure of the gas turbine

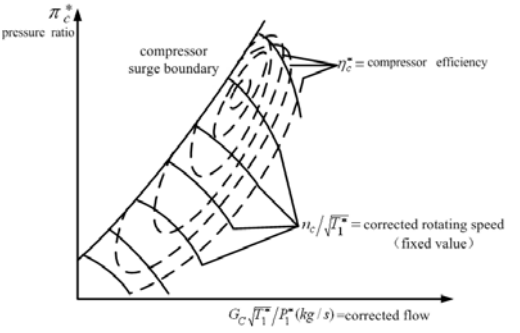


Fig. 2. The compressor general characteristic curve

(2)The Test Index Q_2

$$Q_2 = 1 - P\left(\chi^2 < N\right) = 1 - \int_0^{\chi^2} \frac{t^{(N-2)} e^{-t/2}}{2^{N/2}\Gamma\left(N/2\right)} dt \tag{3}$$

Where: $\Gamma(.)$ is gamma function and $\Gamma(\alpha) = \int_0^{\infty} e^{-t} e^{\alpha-1} dt$. When Q_2 close to 0.5, it indicates that the fixed value of n is proper.

Generally speaking, it often lead to worse morbid of the coefficient matrix of the results when $n \geq 6$. So the chosen fitting order times in this paper is $n \leq 5$. The following calculation results of the indexes in table 1 based on the compressor general characteristic curve of the certain kind of gas turbine.

Table 1. The fitting results under the different fitting order times n

n	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$
RMSE	0.1224	0.0428	0.0105	0.0067	0.0067
Q_2	0.00015	0.00338	0.4745	0.3921	0.8956

It is known by calculation that the mean square deviation index $RMSE$ shows a better trend with the increasing of n. But the over-fitting of Q_2 occurred as $n \geq 4$. Considering the adequate accuracy and the low noises, $n = 3$ was chosen as the fitting order times. Then the component characteristic equation fits as follow combined with the fixed n and the coefficient fitting method proposed in reference [7].

$$\eta_C = -0.0014351\pi_c^3 + 0.035341\pi_c^2 + 0.37562\pi_c - 324.52 \ldots \quad .(4)$$

$$q_m = -0.03565\pi_c^3 + 4.3246\pi_c^2 + 97.564\pi_c - 323.67 \ldots \quad .(5)$$

2.2 Small Deviation of the Mathematical Mode

Combine with thermodynamic characteristics and matching of components considering outside and boundary conditions, the mathematical model of the kind of gas turbine was built. Looking the guide critical section area of the first level turbine A_T and the turbine efficiency η_T as the simple performance parameters, fault factors are introduced into the nonlinear standard model. Seeking the differential of the model, then it can be gotten by substitution of the data as follow.

$$\delta\eta_C - 0.15999\delta\pi_C = \delta\tilde{\eta}_C \quad (6)$$

$$\delta q_m + 0.499974\delta\pi_C = \delta\tilde{q}_m \quad \ldots (7)$$

$$\delta T_2 - 0.307\delta\pi_C + 0.546\delta\eta_C = 0 \quad \ldots (8)$$

$$\delta q_m - 0.5\delta T_3 - \delta\pi_C = \delta\tilde{A}_T \quad (9)$$

$$0.562\delta\pi_C - \delta\eta_C - \delta T_3 - 0.57\delta\varepsilon_T = \delta\tilde{\eta}_T \quad (10)$$

$$\delta T_3 - 0.217\delta\varepsilon_T - \delta T_5 = 0.3811\delta\tilde{\eta}_T \circ \quad \ldots (11)$$

$$\delta q_m + 2.363\delta T_3 - 1.363\delta T_2 - \delta q_f = 0 \quad (12)$$

$$\delta\varepsilon_T - 0.5\delta T_3 + 0.5\delta T_5 = -\delta\tilde{A}_T \quad \ldots (13)$$

Where, η_C is compressor efficiency, π_C is compressor pressurization ratio, ε_T is turbine expansion ratio, T_2 is the air total temperature of compressor export, T_3 is gas total temperature of turbine import, T_5 is gas total temperature of turbine export interface, q_m is air flow, q_f is fuel flow, A_T is the guide critical section area of the first level turbine, $\delta\tilde{\eta}_C$ is the compressor fault factor. $\delta\tilde{q}_m$ is the air flow fault factor, $\delta\tilde{A}_T$ is the guide critical section area of the first level turbine fault factor, $\delta\tilde{\eta}_T$ is the turbine efficiency fault factor.

2.3 The Determination of Fault Coefficient Matrix

By determining the fault model in the above section, we can get the coefficient matrix B and C of Table 2 and Table 3 in Eq (1). Both are the coefficient matrix obtained by fault model. In the actual fault diagnosing process, the pneumatic thermodynamic parameters we can usually monitor are compressor outlet pressure π_C, turbine compressor outlet pressure π_T, air compressor export total temperature T_2, gas turbine import total temperature T_3, and gas turbine export section total temperature T_5. Therefore, we select the above five monitoring parameters as the response measurement parameters when establishing the fault coefficient matrix.

Table 1. Coefficient matrix B

$\delta\eta_C$	$\delta\pi_C$	$\delta\pi_T$	δT_2	δT_3	δT_5	δq_m	δq_f
1	-0.16	0	0	0	0	0	0
0	0.50	0	0	0	0	1	0
0.546	-0.307	0	1	0	0	0	0
0	-1	0	0	0.5	0	1	0
-1	0.562	-0.546	0	-1	0	0	0
0	0	-0.217	0	1	-1	0	0
0	0	0	-1.363	2.363	0	1	-1
0	0	1	0	-0.5	0.5	0	0

Table 2. Coefficient matrix C

$\delta\tilde{\eta}_C$	$\delta\tilde{q}_m$	$\delta\tilde{A}_T$	$\delta\tilde{\eta}_T$
1	0	0	0
0	1	0	0
0	0	0	0
0	0	0	1
0	0	1	0
0	0	0.3811	0
0	0	0	0
0	0	0	-1

Having obtained the coefficient matrix B and C, we can then try to solve the vector y through Eq (1). The fault equation we obtained is:

$$\delta y = B^{-1}C\delta x \tag{14}$$

Where, δy is the response vector of m dimension, δx is the fault factor of n dimension, and $A = B^{-1}C$ is the fault coefficient matrix. According to the established model, we can finally get the fault coefficient matrix as shown below:

Table 3. The fault coefficient matrix of fault model

Fault factor	The response measurement parameters of small deviation capacity				
	$\delta\pi_C$	$\delta\pi_T$	δT_2	δT_2	δT_5
$\delta\tilde{\eta}_C$	-0.3849	0	-0.6804	-1.1548	-1.1548
$\delta\tilde{q}_m$	0.7699	0	0.169	0.3095	0.3095
$\delta\tilde{\eta}_T$	-0.4318	0.2138	-0.0948	-1.2955	-1.723
$\delta\tilde{A}_T$	-0.5236	-1.1219	-0.1149	0.4291	0.6729

3 The Acquisition of Fault Criterion

After the establishment of the fault coefficient matrix, combined the gas path fault performance parameters criterion of gas turbine in reference [9], calculation result of a example of a typical gas path fault on turbine's blade scaling is obtained as follow where gas flow and efficiency of compressor decline by 7% and 2%.

$$\begin{bmatrix} \delta\pi_C \\ \delta\varepsilon_T \\ \delta T_2 \\ \delta T_3 \\ \delta T_5 \end{bmatrix} = \begin{bmatrix} -0.3849 & 0.7669 \\ 0 & 0 \\ -0.6804 & 0.169 \\ -1.1548 & 0.3095 \\ -1.1548 & 0.3095 \end{bmatrix} \cdot \begin{bmatrix} \delta q_m \\ \delta\eta_C \end{bmatrix} \qquad \text{…….. (15)}$$

$$= \begin{bmatrix} -0.3849 & 0.7669 \\ 0 & 0 \\ -0.6804 & 0.169 \\ -1.1548 & 0.3095 \\ -1.1548 & 0.3095 \end{bmatrix} \cdot \begin{bmatrix} -0.07 \\ -0.02 \end{bmatrix} = \begin{bmatrix} 0.0115 \\ 0 \\ 0.0442 \\ 0.0746 \\ 0.0746 \end{bmatrix}$$

Similarly, the measuring parameters criterion of rest 7 typical gas path fault of gas turbine can also be calculated. The fault parameters' criteria obtained through calculation are shown below:

Table 4. The gas path fault variable criterion of gas turbine

Typical gas path fault	Measuring parameters' changing criterion (with % for the unit)				
	$\delta\pi_{CL}$	$\delta\pi_{CH}$	δT_2	δT_2	δT_5
Fault 1	0.0115	0	0.0442	0.0746	0.0746
Fault 2	-0.0308	0	-0.0068	-0.0124	-0.0124
Fault 3	0.0077	0	0.0136	0.0231	0.0231
Fault 4	0.0192	0	0.034	0.0577	0.0577
Fault 5	-0.0259	0.0128	-0.0057	-0.0777	-0.1034
Fault 6	0.0345	-0.0171	0.0076	0.1036	0.1378
Fault 7	-0.0173	0.0086	-0.0038	-0.0518	-0.0689
Fault 8	0.0216	-0.0107	0.0047	0.0648	0.0862

Table 4 has calculated the corresponding fault criterion of the eight kinds of typical gas path fault of gas turbine, which has laid foundation for the application of a variety of fault diagnosis methods. In the meantime, the method above can also be similarly applied to the calculation of fault criteria of other types of gas turbine under different working conditions and has great generality.

4 Conclusion

The paper takes a certain type of gas turbine as the research example and proposes two criterion indexes which provide theoretical basis for fixing fitting order times from the qualitative to quantitative. Researching the choosing of fitting order times n, concludes the characteristic equations of the gas turbine parts by applying the least square fitting method. By using the relationship between typical gas path fault and performance parameter criteria in reference [9], and combining the derived fault coefficient matrix, the paper has finally determined the criterion relationship between the fault and the measured parameters, which provides the fault diagnosis with reliable basis.

References

1. Xie, C., Dai, J.: The Research Overview and Prospects of Gas Turbine Fault Diagnosis Technique. Turbine Technology 52(1), 1–3 (2010) (in Chinese)
2. Yu, D., Liu, J., Xu, J.: The Development of Gas Turbine Technology Faced to Twenty-First Century. Gas Turbine Technology 14(1), 14–21 (2001) (in Chinese)
3. Peng, Y.: Strive to Develop Gas Turbomachine. Gas Turbine Experiment and Research 14(4), 57–60 (2001) (in Chinese)
4. Urban, L.A.: Gas Path Analysis Applied to Turbine Engine Condition Monitoring. Journal of Engineering for Power 10(7), 400–406 (1972)
5. Urban, L.A.: Parameter Selection for Multiple Fault Diagnosis of Gas Turbine Engines. Journal of Engineering for Power 97(2), 225–230 (1975)
6. He, X.: Study on gas turbine performance deterioration based on the thermodynamic work potential. Naval University of Engineering, Wu Han (2010) (in Chinese)
7. Cui, M.: A Coefficient Fitting Method for Compressor Characteristic Curves. Journal of Engineering for Thermal Energy and Power 1(1), 1–8 (1999) (in Chinese)
8. Fan, Z., Sun, C., Bai, J.: Introduction to Aircraft Engine Fault Diagnosis, p. 392. Science Press, Bei Jing (2004) (in Chinese)
9. Yang, Y.: Axis Gas Turbine Performance based Fault Simulation of Thermal Parameters. ShangHai Jiao Tong University (1975) (in Chinese)

Finite-Time Stabilization of Switched Time-Delay System via Dynamic Output Feedback Control

Yi Shen and Hao Liu

School of Astronautic, Harbin Institute of Technology
Shen@hit.edu.cn

Abstract. problem of finite-time stabilization of switched time-delay system via dynamic output feedback is investigated.The definition of finite-time stabilization for switched system via dynamic output feedback is given. Based on the method of average dwell-time and Lyapunov-like function, sufficient conditions which can ensure the switched time-delay system finite-time stabilization via dynamic output feedback are given. An example is given to illustrate the efficiency of the proposed method.

Keywords: Switched system, time-varying delay, finite-time control, dynamic output feedback control.

1 Introduction

A switched system is a special kind of hybrid system, which is composed of a family of subsystems and a switching sequence orchestrating the switching between the subsystems [1]. Recently, switched systems have attracted considerable attention. And for recent progress, readers can refer to survey papers [2-4]. For switched system, average dwell time (ADT) [5-6] and dwell time (DT) [7-8] methods were employed to study the stability and stabilization of switched systems.

Time-delay, which is a common phenomenon encountered in various practical systems, such as chemical engineering systems, distributed networks, and neural networks, is known to be great sources of poor performance and instability. The current methods of stabilization for time-delay systems can be classified into two categories: delay-independent and delay-dependent stabilization [13-15].

Up to now, most of existing literature related to stability of switched systems investigates Lyapunov asymptotic stability, which is defined over an infinite time interval. However, in practice, we may be interested in a bound of system trajectories over a fixed short time. The finite-time stability is a different stability concept which admits the state does not exceed a certain bound during a fixed finite-time interval. Some early results on finite-time stability can be found in [14-15]. In [16] finite-time bounded and finite-time weighted L_2-gain for a class of switched delay systems with time-varying external disturbances is addressed. In [17], the problems of finite-time stability analysis and stabilization for switched nonlinear discrete-time systems are addressed.

However, to the best of authors' knowledge, there is no result available yet on finite-time stabilization of switched systems with time-varying delay via dynamic

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 523–528.
springerlink.com

feedback output. Thus, it is necessary to investigate finite-time stabilization for a class of switched linear systems with time-delay via DOF. In this paper, dynamic output feedback control is designed to ensure the closed-loop switched system finite-time stabilization.

2 Preliminaries and Problem Formulation

Consider the following switched time-varying delay system:

$$\begin{cases} \dot{x}(t) = A_\sigma x(t) + D_\sigma x(t - d(t)) + B_\sigma u(t) + G_\sigma \omega(t) \\ y(t) = C_\sigma x(t) \\ x(t) = \varphi(t), \quad t \in [-\tau, 0] \end{cases} \tag{1}$$

where $x(t) \in \mathrm{R}^n$ is the state, $u(t) \in \mathrm{R}^m$ is control input, $\omega(t)$ is the exogenous noise signal and satisfies Assumption 1, $y(t) \in \mathrm{R}^q$ is measured output, $d(t)$ is the time-varying delay and satisfies Assumption 2, $\sigma : [0, \infty) \to I = \{1, 2, \cdots, N\}$ is the switching signal, $\varphi(t)$ is the given initial condition, A_σ, B_σ, C_σ and D_σ are real known constant matrices with appropriate dimensions, respectively.

Assumption 1: The external noise signal $\omega(t)$ satisfies

$$\int_0^\infty \omega^T(t)\omega(t)dt < d, \quad d > 0 \tag{2}$$

Assumption 2: The time-varying delay $d(t)$ satisfies

$$0 < d(t) \le \tau, \ \dot{d}(t) \le h < 1 \tag{3}$$

Assumption 3: Suppose that the matrices $C_{\sigma(t)}$ have full row rank, that is, $rank(C_\sigma) = q$.

For convenience of discussion, the singular decomposition of C_i as follows:

$$C_i = U_i[S_i, 0]V_i^T \tag{4}$$

where $S_i \in R^{q \times q}$ is a diagonal matrix with positive diagonal elements, $U_i \in R^{q \times q}$ and $V_i \in R^{n \times n}$ are unitary matrices.

In this paper, our goal is to design the following dynamic output feedback (DOF) controller, which can guarantee the switched time-delay system finite-time stabilization.

$$\begin{cases} \dot{x}_d(t) = A_{d\sigma} x_d(t) + B_{d\sigma} y(t) \\ u(t) = C_{d\sigma} x_d(t) + D_{d\sigma} y(t) \\ x_d(t) = 0, \quad t \le 0 \end{cases} \tag{5}$$

where $A_{d\sigma}$, $B_{d\sigma}$, $C_{d\sigma}$ and $D_{d\sigma}$ are matrices to be determined.

Substituting Eq. (5) into Eq. (1), we have

$$\begin{cases} \dot{\tilde{x}}(t) = \tilde{A}_\sigma \tilde{x}(t) + \tilde{D}_\sigma \tilde{x}(t - d(t)) + \tilde{G}_\sigma \omega(t) \\ \tilde{x}(t) = \phi(t), \quad t \in [-\tau, 0] \end{cases} \tag{6}$$

where $\tilde{A}_\sigma = \begin{bmatrix} A_\sigma + B_\sigma D_{d\sigma} C_\sigma & B_\sigma C_{d\sigma} \\ B_{d\sigma} C_\sigma & A_{d\sigma} \end{bmatrix}$, $\tilde{G}_\sigma = \begin{bmatrix} G_\sigma \\ 0 \end{bmatrix}$, $\tilde{D}_\sigma = \begin{bmatrix} D_\sigma & 0 \\ 0 & 0 \end{bmatrix}$, $\tilde{x}(t) = \begin{bmatrix} x(t) \\ x_d(t) \end{bmatrix}$.

The definitions of finite-time bounded and finite-time stabilization of switched system and switched discrete-time time-delay system are given in [16] and [17], respectively. Now, according to those definitions, we propose the following definition for continuous switched time-delay system.

Definition 1[16]: Switched system (1) with $u(t) \equiv 0$ is said to be finite-time bounded with respect to $(\delta, \varepsilon, T_f, d, R, \sigma)$, where $0 \le \delta < \varepsilon$ and $d \ge 0$, R is positive definite matrix and $\sigma(t)$ is a switching signal. If $x^T(t)Rx(t) < \varepsilon$, $\forall t \in [0, T_f]$, $\forall \omega(t): \int_0^{T_f} \omega^T(t)\omega(t)dt < d$, whenever $\sup_{-\tau \le \theta \le 0}\{x^T(\theta)Rx(\theta)\} < \delta$.

Definition 2: Switched system (1) with controller $u(t)$ is said to be finite-time stabilization with respect to $(\delta, \varepsilon, T_f, d, R, \sigma)$, where $0 \le \delta < \varepsilon$ and $d \ge 0$, R is positive definite matrix and $\sigma(t)$ is a switching signal. If $\tilde{x}^T(t)R\tilde{x}(t) < \varepsilon$, $\forall \omega(t): \int_0^{T_f} \omega^T(t)\omega(t)dt < d$, $\forall t \in [0, T_f]$, whenever $\sup_{-\tau \le \theta \le 0}\{\tilde{x}^T(\theta)R\tilde{x}(\theta)\} < \delta$.

Remark 1: It can be seen from Definition 1 that the concepts of finite-time bounded and asymptotic stability are different. A Lyapunov asymptotically stable switched system may not be finite-time stable if its states exceed the prescribed bounds. For observer-based finite-time stabilization, it is necessary for Definition 2 to keep the states and observer residuals in the given bounds.

3 Finite-Time Stabilization

Consider the following non-switched time-delay system:

$$\begin{cases} \dot{\tilde{x}}(t) = \tilde{A}\tilde{x}(t) + \tilde{D}\tilde{x}(t - d(t)) + \tilde{G}\omega(t) \\ \tilde{x}(t) = \phi(t), \quad t \in [-\tau, 0] \end{cases} \tag{7}$$

Lemma 1. Suppose that there exists matrices $P > 0$, $Q > 0$ and $W > 0$ such that

$$\begin{bmatrix} \tilde{A}^T P + P\tilde{A} - \alpha P + Q & P\tilde{D} & P\tilde{G} \\ * & -(1-h)e^{\alpha\tau}Q & 0 \\ * & * & -W \end{bmatrix} < 0 \tag{8}$$

Then, along the trajectory of system (7), we obtain

$$V(t) < e^{\alpha(t-t_0)}V(t_0) + \int_{t_0}^{t} e^{\alpha(t-s)}\omega^T(s)W\omega(s)ds \tag{9}$$

Theorem 1. For $i \in I$, let $\bar{P}_i = R^{-1/2}\tilde{P}_i R^{-1/2}$, $\bar{Q}_i = R^{-1/2}\tilde{Q}_i R^{-1/2}$. Suppose that there exist matrices $\tilde{P}_i > 0$, $\tilde{Q}_i > 0$, $Y_{1,i}$, $Y_{2,i}$, $Y_{3,i}$, $Y_{4,i}$ and constants $\alpha_i > 0$ such that

$$\Psi < 0 \tag{10}$$

$$(\lambda_3 + \lambda_4 \tau e^{\alpha_i \tau})\delta + d\lambda_1 < \lambda_2 e^{-\alpha_i T_f}\varepsilon \tag{11}$$

where

$$\Psi = \begin{bmatrix} \varphi_{11,i} & \varphi_{12,i} & D_i\bar{Q}_{1,i} & 0 & G_i & \bar{P}_{1,i} & 0 \\ * & \varphi_{22,i} & 0 & 0 & 0 & 0 & \bar{P}_{2,i} \\ * & * & \varphi_{33,i} & 0 & 0 & 0 & 0 \\ * & * & * & \varphi_{44,i} & 0 & 0 & 0 \\ * & * & * & * & -W_i & 0 & 0 \\ * & * & * & * & * & -\bar{Q}_{1,i} & 0 \\ * & * & * & * & * & * & -\bar{Q}_{2,i} \end{bmatrix}, \varphi_{12,i} = C_i^T Y_{2,i}^T + B_i Y_{3,i}, \bar{Q}_i = \begin{bmatrix} \bar{Q}_{1,i} & \\ & \bar{Q}_{2,i} \end{bmatrix}$$

$$\varphi_{11,i} = A_i\bar{P}_{1,i} + \bar{P}_{1,i}A_i^T + B_iY_{1,i}C_i + C_i^TY_{1,i}^TB_i^T - \alpha_i\bar{P}_{1,i}, \quad \varphi_{22,i} = Y_{4,i}^T + Y_{4,i} - \alpha_i\bar{P}_{2,i},$$

$$\varphi_{33,i} = -(1-h)e^{\alpha_i\tau}\bar{Q}_{1,i}, \ \varphi_{44,i} = -(1-h)e^{\alpha_i\tau}\bar{Q}_{2,i}, \ \bar{P}_{1,i} = \begin{bmatrix} \hat{P}_{11,i} & \\ & \hat{P}_{12,i} \end{bmatrix}, \ \bar{P}_i = \begin{bmatrix} \bar{P}_{1,i} & \\ & \bar{P}_{2,i} \end{bmatrix}.$$

If the average dwell time of the switching signal σ satisfies

$$\tau_a > \tau_a^* = \frac{T_f \ln\mu}{\ln(\lambda_2\varepsilon) - \ln\rho - \alpha T_f - N_0 \ln\mu} \tag{12}$$

where $\rho = ((\lambda_3 + \tau e^{\alpha_i\tau}\lambda_4)\delta + \lambda_1 d)$, then the parameters of dynamic output feedback controller are given as follows: $D_{di} = Y_{1,i}U_iS_i\hat{P}_{1,i}S_i^{-1}U_i^T$, $B_{di} = Y_{2,i}U_iS_i\hat{P}_{1,i}S_i^{-1}U_i^T$, $C_{di} = Y_{3,i}\bar{P}_{2,i}^{-1}$, $A_{di} = Y_{4,i}\bar{P}_{2,i}^{-1}$, and it can guarantee time-delay switched system (1) is finite-time stabilization with respect to $(0, \varepsilon, T_f, d, R, \sigma)$, where $\mu \geq 1$, $\bar{P}_j \leq \mu\bar{P}_i$, $\bar{Q}_j \leq \mu\bar{Q}_i$, $C_i\bar{P}_i = \hat{P}_iC_i$, $\forall i, j \in I$, $\lambda_1 = \sup_{i\in I}\{\lambda_{\max}(W_i)\}$, $\lambda_2 = \inf_{i\in I}\{\lambda_{\min}(\tilde{P}_i^{-1})\}$, $\lambda_3 = \sup_{i\in I}\{\lambda_{\max}(\tilde{P}_i^{-1})\}$, $\lambda_4 = \sup_{i\in I}\{\lambda_{\max}(\tilde{Q}_i^{-1})\}$, $\alpha = \max_{i\in I}\{\alpha_i\}$.

4 Simulation Results

In this section, an example is utilized to illustrate the efficiency of the method proposed above.

Consider the time-delay system (1) with the following parameters:

$$A_1=\begin{bmatrix}-2.0 & -1.5 & -1.2\\ 0.7 & -1.6 & 0.5\\ -1.3 & 0.5 & -1.1\end{bmatrix},\ A_2=\begin{bmatrix}-1.5 & -1.2 & -1.5\\ 0.2 & -1.5 & 0.4\\ -0.7 & 1.1 & -1.2\end{bmatrix},\ B_1=\begin{bmatrix}1\\ 0.5\\ 2\end{bmatrix},\ B_2=\begin{bmatrix}0.5\\ 0.7\\ 1.5\end{bmatrix},$$

$$G_1=\begin{bmatrix}0.3\\ 0.5\\ 0.2\end{bmatrix},\ G_2=\begin{bmatrix}0.4\\ 0.2\\ 0.3\end{bmatrix},\ D_1=\begin{bmatrix}0.2 & 0.0 & 0.1\\ 0.1 & 0.3 & 0.1\\ 0.3 & 0.1 & 0.2\end{bmatrix},\ D_2=\begin{bmatrix}0.2 & 0.0 & 0.0\\ 0.1 & 0.2 & 0.1\\ 0.1 & 0.1 & 0.3\end{bmatrix},$$

$C_1=[-1.2, 0.5, 0.9]^{\mathrm{T}}$, $C_2=[-1.0, 1.2, 0.5]^{\mathrm{T}}$, $\tau=0.2$, $h=0.01$.

The values of δ, ε, d, T_f and R are given as follows:

$\delta=1.0$, $\varepsilon=30$, $d=0.01$, $T_f=10$, $R=I$.

According to Theorem 1, the controller (5) can be obtained by solving (10) – (12) with $\alpha_i=0.05$, and the feasible solutions are given as follows: A_{d1}=diag(-1.0823, -1.0823, -1.0823), $B_{d1}=[0, 0, 0]^{\mathrm{T}}$, A_{d2}=diag(-0.8068, -0.8068, -0.8068), $B_{d2}=[0, 0, 0]^{\mathrm{T}}$, $C_{d1}=[0\ \ 0\ \ 0]$, $C_{d2}=[0\ \ 0\ \ 0]$, $D_{d1}=-0.3878$, $D_{d2}=-0.5346$, $\tau_a^*=1.9931$.

4 Conclusions

In this paper, unlike most existing research results focusing on Lyapunov stabilization property, we mainly discussed the finite-time bounded of switched linear systems with time-delay via dynamic output feedback. The concepts of finite-time bounded and finite-time dynamic output feedback stabilization are given. Based on the method of average dwell-time and Lyapunov-like function, sufficient conditions which can ensure the switched time-delay system finite-time stabilization via dynamic output feedback are given. An example is given to illustrate the efficiency of the proposed method.

References

1. Song, Y., Fan, J., Fei, M.R., Yang, T.C.: Robust control of discrete switched system with time delay. J. Appl. Math. and Comp. 205, 159–169 (2008)
2. Lin, H., Antsaklis, P.J.: Stability and stabilizability of switched linear systems: A survey of recent results. IEEE Trans. on Autom. Contr. 54(2), 308–322 (2009)
3. Liberzon, D., Morse, A.S.: Basic problems in stability and design of switched systems. IEEE Contr. Syst. Maga. 19(5), 59–70 (1999)
4. Sun, Z.D., Ge, S.S.: Analysis and synthesis of switched linear control systems. Automatica 41(2), 181–195 (2005)
5. Wu, L., Qi, T., Feng, Z.: Average dwell time approach to L2 –L∞ control of switched delay systems via dynamic output feedback. IET Control Theory Appl. 10(3), 1425–1436 (2009)
6. Zhang, L., Shi, P.: Stability, L2 gain and asynchronous control of discrete-time switched systems with average dwell time. IEEE Trans. on Autom. Contr. 54, 2193–2200 (2009)

7. Zappavigna, A., Colaneri, P., Geromel, J.C., Shorten, R.: Dwell time analysis for continuous-time switched linear positive systems. In: Proc. of the 2010 American Control Conference, Baltimore, USA, pp. 6256–6261 (2010)
8. Wu, A., Feng, G., Zeng, X.L.: State feedback stabilizing switching strategies with dwell time for switched discrete-time linear systems. In: Proc. of the 8th World Congress on Intelligent Control and Automation, Jinan, China, pp. 975–980 (2010)
9. Sun, Y.G., Wang, L., Xie, G.M.: Delay dependent robust stability and stabilization for discrete-time switched systems with mode-dependent time-varying delays. J. Appl. Math. and Comp. 180, 428–435 (2006)
10. Moon, Y.S., Park, P.Y., Kwon, W.H., Lee, Y.S.: Delay-dependent robust stabilization of uncertain state-delayed systems. Int. J. of Control 74(14), 1447–1455 (2001)
11. Sun, X.M., Liu, G.P., Rees, D., Wang, W.: Delay-dependent stability for discrete systems with large delay sequence based on switching techniques. Automatica 44, 2902–2908 (2008)
12. Zhang, Y.P., Kang, S.R., Loguinov, D.: Delay-independent stability and performance of distributed congestion control. IEEE/ACM Trans. on networking 15(5), 838–851 (2007)
13. Chen, W.H., Zheng, W.X.: Delay- independent minimum dwell time for exponential stability of uncertain switched delay systems. IEEE Trans. on Autom. Contr. 55(10), 2406–2413 (2010)
14. Weiss, L., Infante, E.F.: Finite time stability under perturbing forces and on product spaces. IEEE Trans. on Autom. Contr. 12(1), 54–59 (1967)
15. Michel, A.N., Wu, S.H.: Stability of discrete systems over a finite interval of time. Int. J. of Control (9), 679–693 (1969)
16. Lin, X.Z., Du, H.B., Li, S.H.: Finite-time boundedness and gain analysis for switched delay systems with norm- bounded disturbance. J. Appl. Math. and Comp. 217, 5982–5993 (2011)
17. Xiang, W.M., Xiao, J.: finite-time control for switched nonlinear discrete-time systems with norm-bounded disturbance. J. of the Franklin Institute 348, 331–352 (2011)

Hyperspectral Image Classification Based on Ensemble Empirical Mode Decomposition

Yi Shen and Min Zhang

Department of Control Science and Engineering,
Harbin Institute of Technology,
PO Box 327, 92# XiDaZhi Street, Harbin 150001, P.R. China
zhangminhit01@163.com

Abstract. Ensemble empirical mode decomposition (EEMD) has been shown in the literature to be a more suitable for nonlinear and non-stationary signals than empirical mode decomposition (EMD). This paper proposes support vector machine classification based on EEMD for hyperspectral images (2D-EEMD-SVM). EEMD is utilized to improve the performance of SVM classification by effectively exploiting the feature of hyperspectral images. First, EEMD is used to decompose hyperspectral images into several intrinsic mode functions (IMFs) and the trend of hyperspectral images. Then, new hyperspectral images are reconstructed with IMFs and the trend. Finally new hyperspectral images are used for support vector machine classification to show the performance of the proposed approach. Experimental results show that 2D-EEMD-SVM can significantly increase the classification accuracy and the classification speed.

Keywords: Classification, ensemble empirical mode decomposition, hyperspectral images, support vector machine.

1 Introduction

Classification of hyperspectral images is an important issue in remote sensing (RS) applications. Recently, Support Vector Machines have been proposed as an efficient method for hyperspectral images classification[1,2]. The results from these studies indicate that SVMs provide a promising approach for classification of hyperspectral images. In this letter, SVM classification based on EEMD is proposed to improve the performance of hyperspectral images classification.

Empirical mode decomposition (EMD) is a novel and powerful method for analyzing nonlinear and non-stationary signal, which is firstly proposed by Huang et al.[3] in 1998. It has been applied in many places, and has shown its effectiveness, for instance, detection and monitoring of crack in financial data analysis[4], vibration signal detection[5], mechanical engineering[6], and so on. Recently, 2D-EMD is utilized to improve the performance of SVM classification (2D-EMD-SVM) for hyperspectral images[7]. Although interesting results are obtained from 2D-EMD-SVM, noise and high-spatial-frequency components can't be separated well in the first IMF, because of the problem of EMD which call mode mixing. In [8], denoising is applied to the first IMF of each band to resolve this problem (2D-EMD-WAV-SVM).

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 529–536.
springerlink.com

In this paper, SVM classification based on EEMD (2D-EEMD-SVM) is proposed to overcome the problem of mode mixing in 2D-EMD-SVM classification.

In the proposed approach, 2D-EEMD is applied to individual hyperspectral image bands as a preliminary decomposition step. Then, new hyperspectral images are reconstructed as the sum of IMFs and the trend. Finally, the new hyperspectral images are classified with SVM classification. Experimental results show that the proposed approach (2D-EEMD-SVM) significantly increases the classification accuracy and the classification speed compared to 2D-EMD-SVM and 2D-EMD-WAV-SVM.

2 EEMD for SVM Classification

The proposed 2D-EEMD-SVM classification includes ensemble empirical mode decomposition, hyperspectral images reconstruction and SVM classification. This section describes the algorithm of empirical mode decomposition, the algorithm of ensemble empirical mode decomposition, and the reconstruction of hyperspectral images with 2D-EEMD.

2.1 Empirical Mode Decomposition

The main idea of the EMD technique is to decompose the original signal into a set of IMFs. The algorithm of empirical mode decomposition for the hyperspectral can be summarized as follows:

Step 1) Set input of the 2D-EMD. $X_k(m,n)$ is the k th band hyperspectral image. The sifting process starts form the original band, so input = $X_1(m,n)$.

Step 2) Find all the local maxima points of input and connect all the maxima points by a cubic spline line as the upper envelope $e_{\max}(m,n)$. Then find all the local minima points of input, and connect all the minima points by a cubic spline line as the lower envelope $e_{\min}(m,n)$. The mean of the upper envelope and the lower envelope is designated as $m_1(m,n)$

$$h_1(m,n) = input_{ij} - m_1(m,n) \tag{1}$$

If $h_1(m,n)$ satisfies all the requirements of the IMF, $h_1(m,n)$ is the first IMF.

Step 3) If $h_1(m,n)$ does not satisfy the conditions of the IMF, $h_1(m,n)$ is treated as the data, step 2) is repeated, $m_{11}(m,n)$ is the mean of the upper and the lower envelopes of $h_1(m,n)$, then

$$h_{11}(m,n) = h_1(m,n) - m_{11}(m,n) \tag{2}$$

If $h_{11}(m,n)$ satisfies all the requirements of the IMF, $h_{11}(m,n)$ is the first IMF. If $h_{11}(m,n)$ does not satisfy all the requirements of the IMF, step 2) is repeated continually up to j times until $h_{1j}(m,n)$ is a IMF, that is

$$h_{1j}(m,n) = h_{1(j-1)}(m,n) - m_{1(j-1)}(m,n) \tag{3}$$

It is designated as the first IMF component $c_1(m,n)$ from the $X(m,n)$, that is $c_1(m,n) = h_{1j}(m,n)$. Then $c_1(m,n)$ is removed from the $X(m,n)$ to obtain the residue $r_1(m,n)$, that is

$$r_1(m,n) = X(m,n) - c_1(m,n) \tag{4}$$

Step 4) The residue $r_1(m,n)$ is treated as the new data, then steps 2) and 3) are repeated to obtain the second IMF $c_2(m,n)$ and new residue $r_2(m,n)$. This procedure is repeated to obtain all the IMFs, the result is

$$\begin{aligned} &\mathrm{IMF}_2 = c_2(m,n) \\ &r_2(m,n) = r_1(m,n) - c_2(m,n) \\ &\quad\vdots \\ &\mathrm{IMF}_i = c_i(m,n) \\ &r_i(m,n) = r_{i-1}(m,n) - c_i(m,n) \end{aligned} \tag{5}$$

The decomposition can be stopped by any of the following predetermined criteria: (1) either component $c_q(m,n)$ or the residue $r_q(m,n)$ becomes so small that it is less than the predetermined value of substantial consequence; (2) the residue $r_q(m,n)$ becomes a monotonic function from which no more IMF can be extracted, which is the trend of the signal. Thus, the original data is the sum of the IMF components plus the trend,

$$X_k(m,n) = \sum_{i=1}^{I} c_i(m,n) + r_n(m,n) \tag{6}$$

The condition of IMF is a standard deviation (SD) by calculating

$$\mathrm{SD} = \sum_{m=1}^{M}\sum_{n=1}^{N}\left[(h_{ij-1}(m,n) - h_{ij}(m,n))^2 / h_{ij-1}^2(m,n)\right] \tag{7}$$

Step 5) Set input of the EMD for the next band hyperspectral image, and repeat Step 1), step 2), step 3) and step 4) until all band hyperspectral images are decomposed.

2.2 Ensemble Empirical Mode Decomposition

The principle of the EEMD is simple [9]: the added white noise would populate the whole time-frequency space uniformly with the constituting components of different scales separated by the filter bank. When signal is added to this uniformly distributed white background, the bits of signal of different scales are automatically projected onto proper scales of reference established by the white noise in the background. Of course, each individual trial may produce very noisy results, for each of the noise-added decompositions consists of the signal and the added white noise. Since the noise in each trial is different in separate trials, it is canceled out in the ensemble mean of enough trails. The ensemble mean is treated as the true answer. In the end, the only persistent part is the signal as more and more trials are added in the ensemble.

The algorithm of 2D-EEMD for hyperspectral image is separated into the following steps:

Step 1) Add a white noise series to the target data

$$X_k(m,n) = X(m,n) + m_k(m,n) \tag{8}$$

Step 2) Decompose the data with added white noise into IMFs with 2D-EMD;

Step 3) Repeat steps 1) and 2) with different white noise series in each trial;

Step 4) Obtain the means of corresponding IMFs of the decompositions as the final result

$$\mathrm{IMF}(i) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} c_{ik}(m,n) \tag{9}$$

The relationship among the ensemble number, the amplitude of the added white noise and the effect of the added noise are given in the following equation [10]:

$$e = \frac{a}{\sqrt{N}} \tag{10}$$

where N is the number of ensemble, a is the amplitude of the added white noise, and e is the standard deviation of decomposition error.

2.3 Hyperspectral Images Reconstruction

After EEMD, mode mixing in 2D-EMD-SVM is removed; the high-frequency noise and components of hyperspectral images are separated well in the IMFs. So hyperspectral images $X(m,n)$ can be reconstructed by the sum of higher IMFs and the final trend, for the lower IMFs contain high-frequency noise. This can be formulated as:

$$X(m,n) = \mathrm{IMFs}_J^K = \sum_{i=J}^{K} \mathrm{IMF}_i + r(m,n) \tag{11}$$

where $1 \le J < K \le N$.

3 Experimental Results

The experimental hyperspectral image set is the 92AV3C dataset collected by AVIRIS sensor taken over northwest Indian Pine test site in 1992[11]. The dataset consists of 145x145 pixels with 224 contiguous spectral bands, with 10 nm intervals in the spectral region from 0.40 to 2.45μm. After removing 4 bands containing all zero values and 20 bands affected by atmosphere, 200 bands that have 16 different land-covers are used for simulation[12]. Select 9 classes of land-covers with the highest number of pixels for experimental samples (see Table 1). In this experiment, the proposed method (2D-EEMD-SVM) is compared with 2D-EMD-SVM and 2D-EMD-WAV-SVM.

Table 1. Number of samples (NOS) for each class of the Indian Pine data

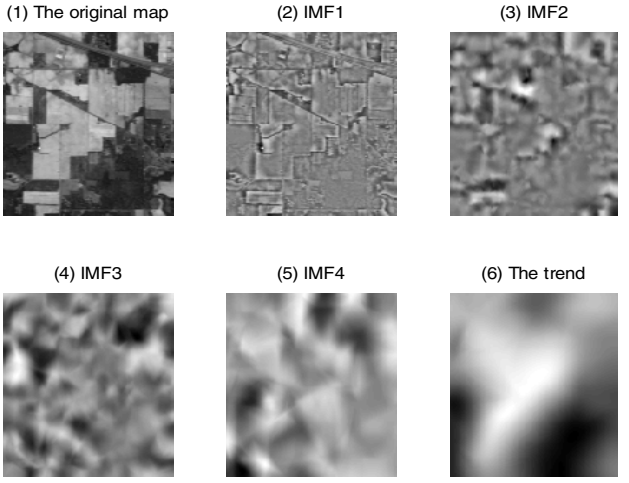
Label	Class	Number of samples
A	Corn-no till	1434
B	Corn-min till	834
C	Grass/Pasture	497
D	Grass/Trees	747
E	Hay-windrowed	489
F	Soybeans-no till	968
G	Soybean-min till	2468
H	Soybean-clean till	614
I	Woods	1294

In the 2D-EEMD, the number of ensemble N is selected as 8; the amplitude of the added white noise a is selected as 0.01. In the SVM classification, five folds (i.e. 20% as training samples, the remaining 80% as testing samples) is used as the classification training model. And the Gaussian radial kernel function is choosed as:

$$K(\mathbf{x},\mathbf{x}') = \exp(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\sigma^2}) \tag{12}$$

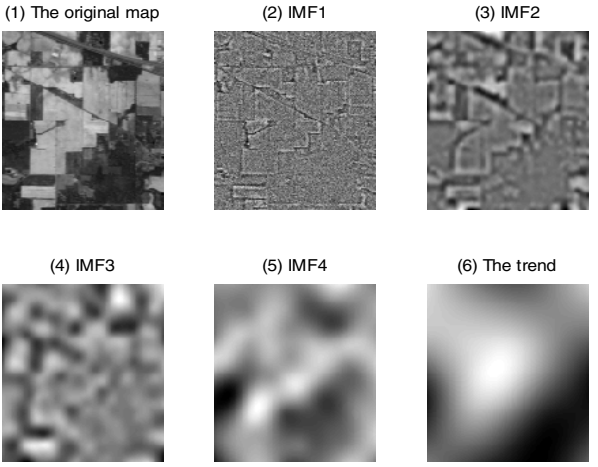
where the parameter σ is selected as 0.4, the penalty parameter is equal to 60.

The results of decomposition of 120th band hyperspectral image with 2D-EMD and 2D-EEMD are shown in Fig.1. It is found that the IMF3 and IMF4 of decomposition with 2D-EMD contain some high-spatial-frequency components, as shown in Fig.1 (a). In other words, IMF3 and IMF4 have the problem of mode mixing. And the IMF3 and IMF4 of decomposition with 2D-EMD contain little high-spatial-frequency components, as shown in Fig.1 (b). So the mode mixing in IMF3 and IMF4 of 2D-EMD are solved by 2D-EEMD well.



(a) Decomposition with 2D-EMD

Fig. 1. Solution of mode mixing in 120th band hyperspectral image in 2D-EMD with 2D-EEMD



(b) Decomposition with 2D-EEMD

Fig. 1. (Continued)

The classification results of 2D-EEMD-SVM, 2D-EMD-SVM and 2D-EMD-WAV-SVM are shown in table 2. In 2D-EEMD-SVM, $IMFs_J^K$ is the reconstruction of the hyperspectral images, which reconstructed by formulation (11). In 2D-EMD-SVM[7], nIMFs is the reconstruction of the hyperspectral images, which reconstructed by formulation (13). In 2D-EMD-WAV-SVM[8], nIMFs is the reconstruction of the hyperspectral images, which reconstructed by formulation (14). In table 2, it is found that the propose method (2D-EEMD-SVM) get the best classification accuracy. In the reconstruction of $IMFs_4^4$, the proposed method gets the highest accuracy, where the overall accuracy of classification is 99.7009% and the average accuracy of classification is 99.7700%. In the classification accuracy of 2D-EMD-SVM, the method 2D-EMD-SVM get the highest accuracy in 5IMFs, where the overall accuracy is 94.8368% and the average accuracy of classification is 95.2432%. In the classification accuracy of 2D-EMD-WAV-SVM, the method 2D-EMD-WAV-SVM get the highest accuracy in 5IMFs, where the overall accuracy is 96.1610% and the average accuracy of classification is 96.0516%. So the classification accuracy of 2D-EEMD-SVM is increased by more than 4% compared to 2D-EMD-SVM, and more than 3% than 2D-EMD-WAV-SVM. In order to analyze the performance of 2D-EEMD-SVM fully, the number of support vectors and the classification time are compared in 2D-EEMD-SVM with $IMFs_4^4$, 2D-EMD-SVM with 5IMFs and 2D-EMD-WAV-SVM with 5IMFs, as shown in table3. It is found that, the number of SVMs of the proposed method (2D-EEMD-SVM) is less than 1800 compared to 2D-EMD-SVM, and less than 1000 compared to 2D-EMD-WAV-SVM. The classification time of the proposed method (2D-EEMD-SVM) is less than 9s compared to 2D-EMD-SVM, and less than 7s compared to 2D-EMD-WAV-SVM.

$$n\text{IMFs} = \sum_{k=1}^{n} n\text{IMF} \tag{13}$$

where nIMF is the nth IMF of the hyperspectral.

$$n\text{IMFs} = \text{DIMF1} + \sum_{k=1}^{n} n\text{IMF} \tag{14}$$

where DIMF1 is the first IMF which is denoised by wavelet, nIMF is the nth IMF.

Table 2. Classification accuracy of different methods

Method		Accuracy (%)	
		Overall Accuracy	Average Accuracy
2D-EEMD-SVM	$IMFs_2^2$	99.1476	99.0269
	$IMFs_3^3$	99.0130	99.0183
	$IMFs_4^4$	99.7009	99.7700
	$IMFs_2^3$	99.0579	99.0207
	$IMFs_2^4$	96.4409	95.9825
	$IMFs_3^4$	98.5644	98.5607
2D-EMD-SVM	1IMFs	91.4259	92.0586
	2IMFs	93.3788	93.8421
	3IMFs	94.6094	94.9825
	4IMFs	94.7164	95.1362
	5IMFs	94.8368	95.2432
	6IMFs	94.7967	95.2294
2D-EMD-WAV-SVM	1IMFs	94.8769	94.7754
	2IMFs	95.6260	95.4747
	3IMFs	96.1209	95.7902
	4IMFs	96.0407	95.9474
	5IMFs	96.1610	96.0516
	6IMFs	96.0674	95.9646
SVM		88.9112	90.2435

Table 3. Classification time and number of SVMs with different methods

Method	Number of SVMs	Test time(s)	Train Time(s)	Classification Time(s)
2D-EEMD-SVM with $IMFs_4^4$	1040	5.2183	0.7862	6.0045
2D-EMD-SVM with 5IMFs	2918	13.9282	1.3348	15.2630
2D-EMD-WAV-SVM with 5IMFs	2137	11.7873	1.3906	13.1779
SVM	3537	16.5557	1.5837	18.1394

4 Conclusions

In this paper, we have proposed SVM classification based on 2D-EEMD for hyperspectral images. Simulations are carried out on 92AV3C dataset with the proposed method (2D-EEMD-SVM), 2D-EMD-SVM and 2D-EMD-WAV-SVM. The experiment results show the proposed method (2D-EEMD-SVM) can solve the mode fixing problem in 2D-EMD successfully. And the performance of the proposed method (2D-EEMD-SVM), which improves the classification accuracy and speed obviously, is better than 2D-EMD-SVM and 2D-EMD-WAV-SVM.

Acknowledgments. This work was financially supported by National Natural Science Foundation of China (No. 60975009) and Research Fund for the Doctoral Program of Higher Education of China (No. 20092302110037).

References

1. Melgani, F., Bruzzone, L.: Classification of hyperspectral remote sensing images with support vector machines. IEEE Trans. Geosci. Remote Sens. 42, 1778–1790 (2004)
2. Camps-Valls, G., Bruzzone, L.: Kernel-based methods for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 43(6), 1351–1362 (2005)
3. Huang, N.E., Shen, Z., Long, S.R., et al.: The Empirical Mode Decomposition and The Hilbert Spectrum for Nonlinear and Non-stationary Time Series Analysis. Proceeding of Royal Society A 454, 903–995 (1998)
4. Huang, N.E., Wu, M.L., Qu, W.D., Long, S.R., Shen, S.P.: Applications of Hilbert-Huang transform to non-stationary financial time series analysis. Appl. Stochastic Models Bus. Ind. 19, 245–268 (2003)
5. Peng, Z.K., Peter, W., Chu, F.L.: An improved Hilbert–Huang transform and its application in vibration signal analysis. J. Sound Vib 286, 187–205 (2005)
6. Babu, T.R., Srikanth, S., Sekhar, A.S.: Hilbert-Huang transform for detection and monitoring of crack in a transient rotor. Mech. Syst. Signal Pr. 22, 905–914 (2008)
7. Demir, B., Ertürk, S.: Empirical Mode Decomposition of Hyperspectral Images for Support Vector Machine Classification. IEEE Trans. Geosci. Remote Sens. 48, 4071–4084 (2010)
8. Demir, B., Ertürk, S., Güllü, M.K.: Hyperspectral Image Classification Using Denoising of Intrinsic Mode Functions. IEEE Geosci. Remote Sens. 8, 220–224 (2011)
9. Wu, Z.H., Huang, N.E.: Ensemble Empirical Mode Decomposition: a Noise-assisted Data Analysis Method. Advances Adaptive Data Analysis 1, 1–41 (2009)
10. Shen, Z.Y., Wang, Q., Shen, Y., Jin, J., Lin, Y.R.: Accent extraction of emotional speech based on modified ensemble empirical mode decomposition. In: 2010 IEEE International Conference on Instrumentation and Measurement Technology, pp. 600–604 (2010)
11. `ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/92AV3C/`
12. Zhang, M., Shen, Y., Wang, Q.: Hyperspectral Datum Classification Using Kernel Method Based on Mutual Information of Neighbor Bands. Optoelectronics Letters 5, 309–312 (2009)

Globalization in Cross-Cultural Communication and Chinese Civil Society

Shengyong Zhang

Foreign Language Department of Dezhou University, Dezhou, Shandong, China
jackyzhangshengyong@gmail.com

Abstract. This paper discusses five major sets of developments that have had a profound relationship to intercultural, cross-cultural, international, and global communication: (1) globalization and culture universals; (2) eastern and western values; (3) culture and modernity; (4) international and global communication; and (5) the Chinese civil society.

Keywords: Civil society, Globalization, cross-cultural communication.

1 Globalization and Cultural Universals

We are well aware through the concept of globalism and the actual process of globalization that a new world order has begun to emerge in the recent past several decades, much different than international relations experts might have predicted even in the mid-1980s. *The Encarta World English Dictionary* (1999) defines globalism as "the belief that political policies should take world-wide issues into account before focusing on national or state concerns, or the advocacy of this concept: and globalization as "the process by which social institutions become adopted on a global scale" or "the process by which a business or company becomes international or starts operating at the international level." In Europe, the term 'social cohesion' is preferred, while in Asia, the phrase 'the harmonious society' is more commonly used. Whatever the phrasing, the meaning behind the words is solidarity among all of humankind for the common good.

Samuel P. Huntington in his 1996 book, *The Clash of Civilizations: and The Remaking of World Order*, asks whether the world cultures are fundamentally different and at odds with each other. Can we engage in a "dialogue of civilizations?" If China is a civilization state more than a nation state, as Jacques Martin postulates, it is an important intercultural issue for China and for other states to pursue, as well as for individuals interacting with Chinese citizens and the reverse. Martin proposes, "It is this civilizational dimension which gives China its special and unique character" (2009, p.374). He argues that "Previously, the US was regarded as the overwhelming agent and beneficiary of globalization. Now the main beneficiary is perceived to be East Asia and especially China" (p. 352). In this case, we may have to consider globalization, not as moving Asians towards westernization, but increasingly as the West moving more toward East Asia and China.

The theory of cultural universals offers important intercultural and cross-cultural communication insights. However, Clifford Geertz warns that having identified such

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 537–542.
springerlink.com

concepts as cultural universals, we must test the most significant ones which are thoroughly grounded in particular biological, sociological, or psychological processes, empirically across cultures and cross-culturally. He believes that saying that all people have a religious impulse (and we are aware that most Chinese do not recognize such an impulse), or have reasonably similar views on mating or marriage, or the concept that all people have a common interest in private property, for example, then "the question still remains whether such universals should be taken as the central elements in the definition of man [and woman], whether a lowest common denominator of humanity is what we want anyway…. In short, we need to look for systematic relationships among diverse phenomena, not for substantive identities among similar ones" (1973, pp. 39-44).

2 Eastern and Western Values

Without briefly considering the basis of Eastern and Western values broadly in their ancient and contemporary context, it is difficult to understand the pull between tradition and modernity in Confucian societies such as China, Japan, Korea, and Southeast Asia versus western Graeco/Roman, Judao/Christian societies. Robert T. Oliver, in his book, *Communication and Culture in Ancient India and China* (1971), argues: "For centuries, the 'Confucian industry' of China has matched the 'Shakespearean industry" of Anglo-American scholarship in its production of books, lectures, and educational programs. What Confucius sought above all was a society in which harmony would prevail, because propriety and loyalty would be practiced by the rulers and the people….This philosophy was cogent, clear, consistent, and practical" (1971, pp.123-144). Oliver concluded his thoughts about Confucius on harmony and justice: "The focus of his inquiry was upon effective means of adjusting people to ideas, ideas to people, and people to people. This was the humanist way, as he conceived it" (p.144). Basically, Confucius was concerned about the social order, based on love for one's kind and family, authority, social stability and harmony: "Goodness: in private life, courteous; in public life, diligence; in relationships, loyal."

Among the Greek philosophers, Socrates, Plato and Aristotle, there was an ongoing debate about whether the ultimate values were truth, wisdom, goodness, and justice, leading toward the good life from the perspective of Plato, or from the point of view of Aristotle, his student, whether these values must be complemented by happiness, then leading to the good life. Socrates, the main protagonist in Plato's dialogues, always modestly, but firmly, identified with the idea of what a philosopher should be, an individual with wisdom, truth, and a sense of justice. In a more practical way, Aristotle called for the speaker to have qualities of ethos (credibility, knowledge, and good will for the audience), logos (reasoned logic) and pathos (appropriate emotional appeals for the audience). Where Plato compared argument with dialectic having the favored place over rhetoric philosophically, Aristotle identified dialectic as the counterpoint of politics—thus promoting happiness and movement for the community toward the good life. The Greek teachers of rhetoric saw themselves as guiding young Greek male citizens to become active members of the civil society and able to argue directly, forcefully, and persuasively their well-considered points of view, in contrast to the Confucian hierarchical, authoritarian based reasoning—modeling justice and goodness by the leaders, and thereby naturally but implicitly encouraging the populace to follow.

In terms of the comparison of Eastern and Western Values, K.S. Sitaram (1995) identifies two primary values and value orientations: Eastern responsibility VS Western individuality. He regards many easterners as seeing themselves collectively, with modest respect for responsibility, authority, benevolence, and propriety for their groups and others as well as loving care for family members, while westerners see themselves first as individuals which leads to the importance of values such as competitiveness, aggressiveness, challenges to authority, public opinion polls, political differences, court-protected individual rights, success, high personal earnings, private property, personal identity, self-centeredness, and ethnocentrism. To be more specific, Sitaram lists the following contrasting values and goals in Asian and western cultures. In Asian culture, he believes that the key values are authoritarianism [authoritarian consensus], brotherhood, collective responsibilities, education, gratefulness, loyalty, respect for elders, and hospitality. While in western culture, the primary values are human dignity, individuality, firstness, frankness, directness, punctuality as well as respect for youth. According to hisillustration, the following are key values stressed in the Eastern cultures. Naturally, culture is not static, and so some of these characteristics also change as the cultures evolve.

3 Culture and Modernity

The traditional society remains static in time and place and tends to most often culturally accept submission to authority, filial piety, conservatism, fatalism, pessimism, and a patriarchal society, with authoritarian consensus, or a potentially multicultural and harmonious integration, while the modernizing society moves, often only gradually, toward egalitarianism and open mindedness, gender equality, social isolation, individualism, self reliance, optimism and new assertiveness. Asia's future may include an integration of an individualistic and collectivistic society, with an increasingly modern scientific education which leads to western cognitive intelligence and rationality, identity as global citizens and international stakeholders, and more hybrid identities. The new cross- cultural prism may be Asians in western dress and customs, and westerners in traditional Asian dress and customs (Martin, 2009).

Jules Henry, in his *Culture Against Man* (1963), Jacques Ellul in *The Technological Society* (1964), and B. F. Skinner, in *Beyond Freedom and Dignity* (1971), were all "cultural determinists". In every cultural generation and time, the irreversibility of technical progress and its geometric cultural diffusion and expansion across space, according to Henry, Ellul, and Skinner, offer the conclusion that technology in modern life is so pervasive that it has produced a global society in which human culture has become the subject rather than the master over technology. Henry has argued that in contemporary society, the technological society is a driven one, seeking to become expansive, competitive, individualistic, and consuming, all cultural factors that are usually subscribed to westernization and even global Americanization (1963, pp. 23-25). Although some value theorists such as Shalom Schwartz include such drives among universal values, it would seem that they are the antithesis of universal values such as those espoused by the ancient Greeks of truth, wisdom, happiness and the good life or by Confucius of benevolence, ritual, right thinking, and harmony.

One of the most interesting recent western books about Chinese culture is Jacques Martin's *When China Rules the World: The End of the Western World and the Birth of a New Global Order* (2009) in which he suggests that modernity in East Asia and the West can be measured by several different characteristics: language, the body, food, power, and politics. Martin draws two general conclusions about East Asian modernity: "First, if the impact of Westernization is limited, it follows that these societies—and their modernities—remain individual and distinctive, rooted in and shaped by their own histories and culture. It also follows that their modernization has depended not simply or even mainly upon borrowing from the West, but on their ability to transform and modernize themselves." Second, he notes, "if the process of modernization is simply a transplant then it cannot succeed. A person must believe that modernity is there in order to take root and flourish" (pp. 113-138). In terms of China, we can then surmise from Martin's arguments that Chinese modernization, while borrowing from the West is useful; it must be with "Chinese characteristics".

4 International and Global Communications

In contrast to the cross-cultural national characteristics or dimensions, as are suggested in the studies of Geert Hofstede, when we discuss international and global communication, Thomas L. McPhail, in his book, *Global Communication* (2010) defines it as referring "to the cultural, economic, political, social, and technical analysis of communication and media patterns and effects across and between nation-states. International communication focuses more on global aspects of media and communication systems and technologies and, as a result, less on local or even national aspects or issues" (2010, p.2). Noting a number of critical issues relating to international communication, he argues that they can be explained through three major theories or movements: NWICO (New World Information and Communication Order), electronic colonialism, and world system theories: "International communication will have a greater impact on the future of the planet than exploration and transportation combined" (p.35).

McPhail identifies the Electronic Colonialism Theory (ECT) as passing through four epochs of empire-building: the Greco-Roman period; the Crusades of the Middle Ages; the mercantile colonialism; and finally, with the rise of nationalism and decolonization, the recent and current electronic colonialism represents the dependent relationship of poorer regions on the post-industrial nations, especially in the area of communication transfer.

McPhail labels what he considers the third major aspect in considering both the international and global communication revolution as the World System Theory (WST), which identifies the core, periphery, and semiperiphery zones in today's global setting. He defines the core zone as "Capital intensive, high-wage, high-technology production, involving lower labor exploitation and coercion"; the semi-periphery zone as "Core-like activities, Peripheral-like activities"; and the periphery zone as "Labor-intensive, low-wage, low-technology production involving high labor exploitation and coercion." He contends that the "World system theory states that global economic expansion takes place from a relatively small group of core-zone

nation-states [the industrialized West] out to two other zones of nation-states, these being in the semi-peripheral and peripheral zones." He identifies China as among the semi-peripheral nations, along with Brazil and India, all of which can expect in the near future to become core nation-states, rivaling both the US and European Union's initial ten nation-states.

In this sense, we can also see the merger of both international and global communication through what McPhail calls "three new strong hegemonic communication forces stemming from: (1) expansion of cable and satellite broadcasting systems; (2) An avalanche of Western, primarily American, television and movie programming, and (3) The collective rules of the WTO, the World Bank, and the International Monetary Fund." McPhail notes that issues facing both international and global communication are explained through the three major theories or movements which he has proposed: "Collectively, they help organize or frame the trends, economics, technologies, and stakeholders involved in the dynamic, globally significant, and expanding role of international communication" (pp. 31-35).

5 Chinese Civil Society

John H. Powers and Randy Kluver suggest that earlier periods of China were framed by "fragmentation, civil war, and Japanese occupation, wherein little progress was made toward the evolution of civil society in the modern sense. From the founding of the People's Republic, the development of 'civil society with Chinese characteristics floundered while the nation lurched from one top-driven political movement to another." They propose that "From a communication framework" civic society may develop on either a top down or bottom up basis, or within various groups in the society, and also between the central government and other international and global actors. Not only has there been the possibility of internal and external civic communication but also between the internal Chinese society and the world-wide Chinese diaspora" (1999, pp. 1-2).

Kluver writes that civil society allows citizens a voice in influencing social and political life outside the power of the state itself. He calls civic discourse the ability to define the nature of the society and its people, including economics, cultural and social issues, and popular culture, by which the national identity can be expressed. Civic discourse helps to create the society, and the civil society helps to promote civic discourse, much like Hall's claim that culture is communication and communication is culture. Civic discourse is seen not only in political matters in China, where those in power in the Party and government provide the substantial part of the discourse, but also in popular culture including the arts, music, books, TV, advertisements, the internet both as a form of exercising communicative interactions and in a major way in China as entertainment. Given the ever increasing importance of education at all levels in China, not only do the political leaders exercise power in the civic society, but also intellectuals from the major universities (1999,pp. 12-13). Just as Jacques Martin (2009) has argued, "The future is likely to bring an even more diverse set of ideas into a culture and a society that is rapidly constructing a new identity, and a new Chinese world" (pp. 20-22).

References

1. Chang, K.C.: Food in Chinese Culture–Anthropological and Historical Perspectives. Yale University Press, NewHaven (1977)
2. Deng, Y.: Ancient Chinese inventions, Trans. Wang, P. Intercontinental Press, Beijing (2005)
3. Ellul, J.: The Technological Society. Introduction by Merton, R. K., Trans. Williamson, J. Knopf, New York (1964)
4. Geertz, C.: The Interpretation of Cultures. Basic Books, New York (1974)
5. Henry, J.: Culture Against Man. Penguin, Middlesex (1963)
6. Huntington, S.P.: The Clash of Civilizations and the Remaking of the World Order. Simon and Schuster, NewYork (1996)
7. Kluver, R.: Elite-based Discourse in Chinese Civil Society. In: Powers, J.H., Kluver, R. (eds.) Civic Discourse, Civil Society and Chinese Communities. Ablex, Stamford (2009)
8. Kuhn, R.L.: How China's Leaders Think: The Inside Story of China's Reform and What This Means for the Future. Wiley, New York (2010)
9. Martin, J.: When China Rules the World: The End of the Western World and the Birth of a New Global Order. The Penguin Press, New York (2009)
10. McPhail, T.L.: Global Communication: Theories, Stakeholders and Trends, 3rd edn. Wiley-Blackwell, Malden (2010)
11. Oliver, R.T.: Communication and Culture in Ancient India and China. Syracuse University Press, Syracuse (1971)
12. Powers, J.H., Kluver, R.: Civic Discourse, Civil Societies, and Chinese Communities. Ablex, Stamford (1999)
13. Sitaram, K.S.: Communication and Culture: A World View. McGraw Hill, New York (1995)
14. Skinner, B.F.: Beyond Freedom and Dignity. Knopf, New York (1971)

Design of Glycol Circulation Control System about Engine Hot Test Line

Yufeng Lian[1], Chonghe Tang[1], and Qiang Wang[1,2]

[1] School of Electrical & Electronic Engineering,
ChangChun University of Technology, Changchun, 130012, China
[2] Jilin Provincial Communist Youth league Youth Education Center,
Changchun,130061, China
{lianyufeng_1982,tangchonghe}@126.com,
lianyufeng_1982@163.com

Abstract. To ensure the quality and performance of factory engine, hot test is essential. Fueling engine to test its quality and performance was a traditional method, which wasted much fuel with high cost. At present, it does not meet the trend of energy saving. To this end, instead of fuel, glycol was used as energy to test engine in engine hot test line, so as to achieve energy savings. Glycol circulation control system about engine hot test line was designed, which used glycol as test energy.

Keywords: Engine, hot test line, glycol circulation system.

1 Introduction

Fuel is non-renewable resources, and it has been one of the important global energy. Many conflicts in the world are caused by fuel; it goes without saying that fuel is important to human. Auto is one of the major transports for human, engine is the heart of auto, and engine test technology has always been the main study goal of researchers [1][2][3]. To ensure the quality and performance of the factory engine, starting the engine in different speeds and exerting certain load to test engine performance, it is called hot test. Hot test need energy, and energy comes from fuel combustion. Fuel was used as test fuel at first, but because of its high cost, there is an urgent need for a substitute with low cost. With the people's continuous exploration and efforts, glycol can be used to test engine performance as a test fuel instead of fuel. Compared with fuel, Glycol is not only a renewable and available resource with low cost, but is also easy to be produced [4][5].

2 Glycol Circulation System Process

Glycol circulation system process was composed of mixing process, heating process, circulation process, precipitation process and purification process. The order of their work was shown in Fig.1.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 543–547.
springerlink.com

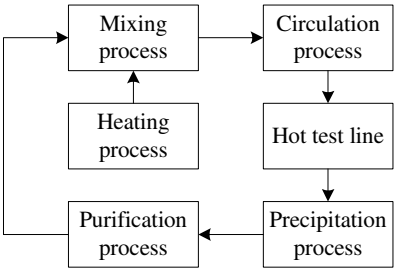


Fig. 1. Glycol circulation system process

The mixture with a certain percentage of running water was made by mixing running water and glycol in mixing tank, so as to meet the requirements of engine hot test line. The mixture was heated by jacket water that was heated by heating process and the temperature of mixture must be suit for engine hot test. In circulation process, the mixture with a certain temperature was pressured by circulation pump, and it was sent to engine hot test line. There are three circulation pumps used in this process. According to the actual pressure, mixture was provided with constant pressure by controlling three circulation pumps.

Mixture came back from engine hot test line flow into the precipitation tank to be precipitated, and then it would be pressured by two pressure pumps, and then purified through two filters to go back into mixing tank. At this point, the whole glycol circulation system process was completed.

3 Glycol Circulation System Hardware Design

S7-300 PLC was used as a system controller in control system, and mixing process and circulation process played a very important role in the whole processes.

3.1 Mixing Process Controller Design

Mixing process requires that glycol and running water ware mixed with a certain percentage; the structure was shown in Fig.2. G_{c1} was transfer function of glycol flow loop controller; G_{a1} was transfer function of actuator; G_{p1} was transfer function of controlled object; G_{s1} was transfer function of detection part; G_{c2} was transfer function of running water flow loop controller; G_{a2} was transfer function of actuator; G_{p2} was transfer function of controlled object; G_{s2} was transfer function of detection part; K was proportional coefficient.

In glycol flow loop, the settings (r_1) ware got from touch screen to set the flow control settings and instrumentation, according to detected feedback value, the

corresponding control signals ware calculated by PID algorithm to control flow valve opening; Running water flow loop was similar to glycol flow loop, the difference was that settings ware got by multiplying between r_1 and K in running water flow loop.

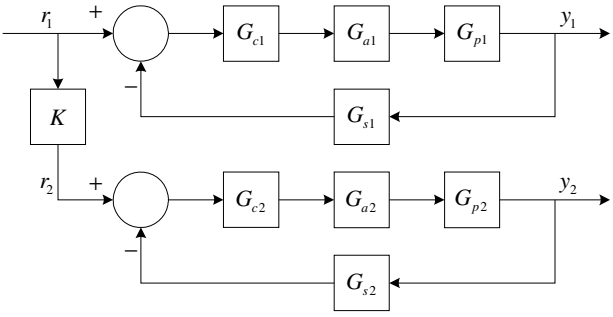


Fig. 2. Mixing process control system structure

3.2 Circulation Process Control Project

In order to improve the test efficiency of engine hot test line, the whole system must non-stop work, so the three circulation pumps used loop-soft start. When pump run to 50HZ, and actual pressure has not reached the set pressure, not directly to start another pump, but start another pump with frequency conversion to achieve the purpose of maintaining a constant pressure. Cut out the pump, in accordance with the first start first stop manner. This working manner has the advantages that all the pumps ware used at work time, according to a cycle way conversion. Circulation pumps started with frequency conversion, so this manner had no effect on grid, which can prevent the circulation pumps from rust that resulted from setting aside for a long time effectively.

4 Glycol Circulation System Software Design

System software included PLC programming software and touch screen configuration software. STEP 7 V5.4 was used to program PLC to achieve controlling of pumps and valves; MCGS was used to program touch screen configuration to achieve monitoring working status.

4.1 Cycle Program Design

According to the loop-soft start works, three circulation pumps work flow was shown in Fig.3. In Fig.3, FC: frequency conversion; PF: power frequency; Ps: pressure measurement; Pr: Pressure set.

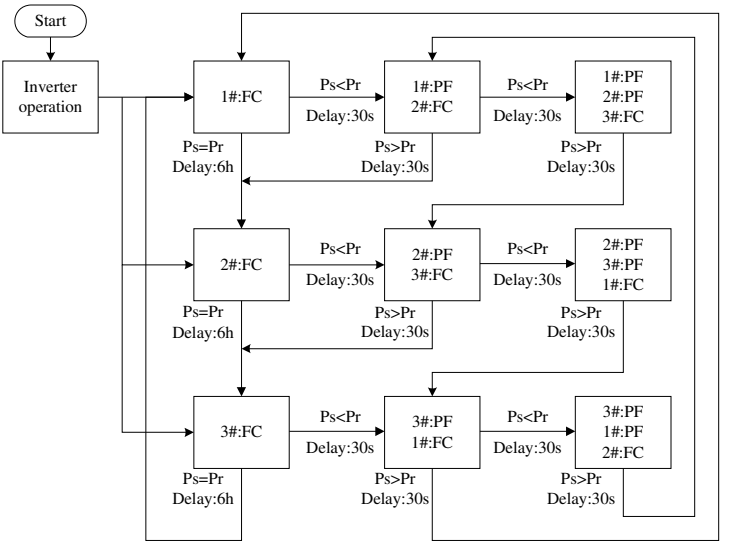


Fig. 3. Circulation pumps work flow

4.2 Touch Screen Configuration Software Design

Touch screen configuration included two configurations mainly, mixing process configuration and precipitation process configuration. The liquid level, temperature, pressure and opening of valves ware monitored in touch screen configurations. Touch screen configurations ware shown in Fig.4 and Fig.5.

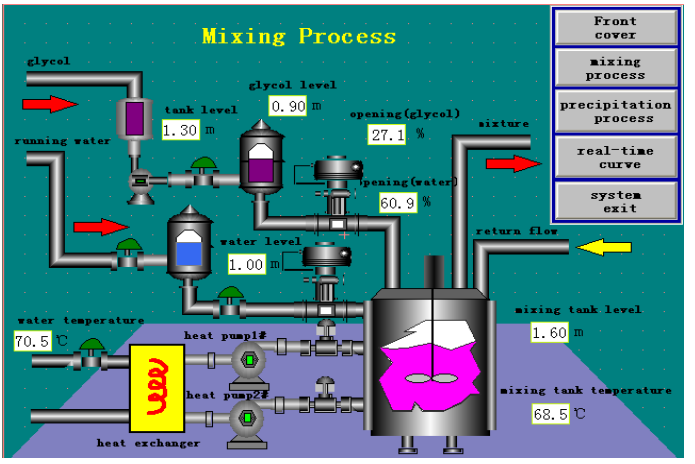


Fig. 4. Mixing process configuration

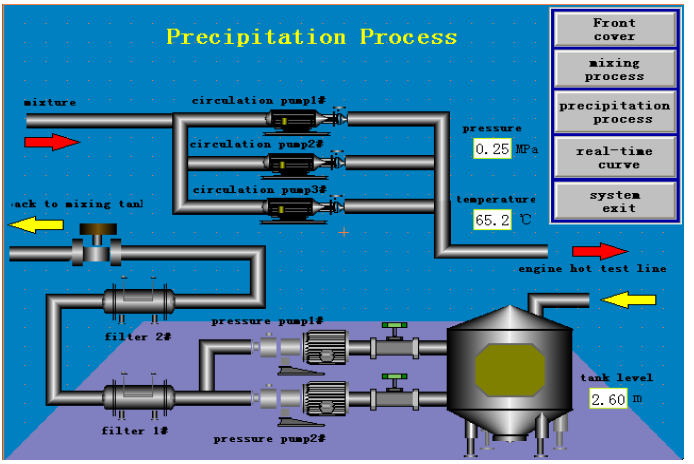


Fig. 5. Precipitation process configuration

5 Conclusion

Glycol circulation control system about engine hot test line was designed and it can control equipments to run automatically according to the practical conditions. It can achieve energy saving and productivity-enhancing purposes and the technology and method adopted in this system is practical and worthy of using abroad.

References

1. Angelov, P., Giglio, V., Guardiola, C., et al.: An approach to model-based fault detection in industrial measurement systems with application to engine test benches. Measurement Science & Technology 17(7), 10 (2006) doi:10.1088/0957-0233/17/7/020
2. Alvey, A., Jay, G.C.: Application of NVH Techniques to Engine Production Line Test. In: 2005 SAE Noise and Vibration Conference and Exhibition (June 2005)
3. Advanced hot test cells for end-of-line engine testing. Diesel progress North American 66 (7) _2 (2000)
4. Wu, J.-J., Tang, H.-W.: On-line test technology of engine technology state. In: 3rd International Symposium on Test and Measurement, vol. 1(1) (1999)
5. Yang, Y.: Engine and automotive theory, vol. 5. People's Communications Press (1993)

A Hybrid Fault Detection and Diagnosis System Based on KPCA and DDAG

Qiang Gao[1,2], Guojing Wang[2], and Xiaopeng Hao[2]

[1] School of Electrical Engineering & Automation of Tianjin University, 300072, China
[2] Tianjin Key Laboratory for Control Theory & Applications in Complicated Systems, Tianjin University of Technology, Tianjin, 300384, China
{gaoanbei,Chaoyue91}@163.com,88279093@qq.com

Abstract. In order to improve the capability of the fault diagnosis, this paper introduces the Decision Directed Acyclic Graph (DDAG) algorithm and establishes a new detection and diagnosis system combing the DDAG with the Kernel Principal Component Analyses (KPCA) method. The hybrid system uses KPCA and DDAG to detect and identify the fault. A specific description of the principles and procedures about how to use KPCA method and DDAG is given. The new detection and diagnosis system has an excellent performance in the fault detection and diagnosis of the Tennessee-Eastman (TE) process. This paper gives a new way to research the fault detection and diagnosis in industrial nonlinear system.

Keywords: Fault diagnosis, KPCA, DDAG, TE, nonlinear system.

1 Introduction

With the rapid development of mass production and new technology, automatic control systems for modern industrial processes have been expanding to increasing complexity and larger scale. Its timely and accurate diagnosis can reduce the chemical production downtime, increase production safety. In terms of efficiency point of view or perspective of safety, fault diagnosis has the necessity and urgency. For complex industrial processes, variables often have a nonlinear relationship. To solve this problem, B.Schölkopf and others proposed a nonlinear process of detection of kernel PCA (KPCA [1-3]), with high practical value.

DDAG [4,5], a new learning machine based on Statistical Theory, is a Multi-category Support Vector Machines (M-SVM).It is a new learning method based on a limited sample of learning theory. Taking advantage of the original data, it can be effective for fault identification. In this paper, KPCA and DDAG are combined to enhance the fault detection and diagnostic capability.

2 KPCA

2.1 The Principle of KPCA

Through the nonlinear mapping φ , firstly the original input space $(x_1, x_2, \cdots, x_n \in R^m$, where n is the number of samples, m is the dimension of

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 549–555.
springerlink.com

measured variables) is mapped to a high-dimensional feature space F (as is $\varphi : R^m \to F$). Then the nonlinear problem of input space is transformed into the linear problem of feature space. The mapping x_i will be denoted by $\varphi(x_i) = \varphi_b$, and the covariance matrix of the feature space F can be expressed as:

$$C^F = \frac{1}{N}\sum_{i=1}^{N} \varphi_i \varphi_i^T \tag{1}$$

Set λ as the eigenvalue of the matrix C^F and v as the eigenvector, then:

$$\lambda v = C^F v \tag{2}$$

Eigenvector v can be obtained by mapping the sample of the feature space:

$$v = \sum_{i=1}^{N} \alpha_i \varphi_i \tag{3}$$

From equation (1), the v corresponding to the maximum value of λ is the first principal component of the feature space and the v corresponding to the minimum is the last main principal, then $\lambda v = C^F v$ is equivalent to:

$$\lambda < \varphi_k, v > = < \varphi_k, C^F v >, \varphi_k = \varphi(x_k), k = 1, \cdots, N \tag{4}$$

$< x, y >$ is the dot product of x x and y. Combing with equation (3) and (4), we can get:

$$\lambda \sum_{i=1}^{N} \alpha_i < \varphi_k, \varphi_i > = \frac{1}{N}\sum_{i=1}^{N} \alpha_i < \varphi_k, \sum_{j=1}^{N} \varphi_i >< \varphi_p \varphi_i > \tag{5}$$

Define the matrix $K \in R^{N \times N}$, and set that $[K]_{ij} = k_{ij} = < \varphi_{ij} \varphi_j >$, so by equation (3) we can get:

$$\lambda N \alpha = K\alpha, \alpha = [\alpha_1, ..., \alpha_k]^T \tag{6}$$

Before the analysis of principal component in the feature space F, you should make standardized. Substitution matrix K with the following equation:

$$\hat{K} = K - I_N K - K I_N + I_N K I_N \tag{7}$$

In the equation (7), I_N is equal to the multiplication of $1/N$ and a matrix $E = R^{N \times N}$. So the analysis of principal component in the feature space is equivalent to the eigenvalue of equation (6). Combined with equation (6) and (3), eigenvectors v of matrix C^F can be obtained by the eigenvector α of the matrix K, and satisfies:

$$\langle v_k, v_k \rangle = 1 \ , \ k = 1, \cdots, p \ . \tag{8}$$

And the p is the principal number. Then it should calculate the principal component by calculating the projection of the mapping data on the eigenvectors v_k .

$$t_k < v_k, \varphi(x) >= \sum_{i=1}^{N} \alpha_i^k < \varphi(x_i), \varphi(x) >= \sum_{i=1}^{N} \alpha_i^k k(x_i, x) \tag{9}$$

To solve the eigenvalue problem of equation (6), and calculate the main principal vector of feature space from the input space directly by using (9), we introduce the kernel function of the form of dot product in the feature space, namely $k(x, y) =< \varphi(x), \varphi(y) >$, to avoid a direct calculation of non-linear mapping.

2.2 Fault Detection Strategies of KPCA

KPCA-based fault detection method uses the T^2 and SPE statistics to detect faults in the feature space. T^2 statistic is the square of the standard principal component vector and is defined that:

$$T^2 = [t_1, ..., t_p] \Lambda^{-1} [t_1, ..., t_p]^T \tag{10}$$

In the equation (10), t_i can be get by equation (9), Λ^{-1} is the eigenvalue of principal composition.

The statistical control limit of T^2 can be obtained by F-distribution:

$$T_{\lim}^2 = \frac{p(N-1)}{N(N-p)} F_\alpha(p, N-p) \tag{11}$$

In the equation above: p is the number of the principal component, N is the number of samples.

Squared prediction error (SPE) is a measure of changes in external data of the model. It is defined that:

$$SPE = \left\| \varphi(x) - \varphi(y) \right\|^2 = \sum_{i=1}^{N} t_i^2 - \sum_{i=1}^{p} t_i^2 \tag{12}$$

The control limits of the SPE statistic is:

$$SPE_{\lim} = g X_h^2 \tag{13}$$

In the above equation: g and h are the constant coefficients of the mean and variance about the SPE .

Application of KPCA for fault diagnosis of industrial processes, firstly, pretreats the collected data from normal condition, and carries on kernel principal component

analysis to establish the kernel principal component model which reflects normal operation. Secondly calculate the T^2 and SPE statistic to determine the control limits, and then calculate the observed sample statistics. Excess of control limits stands for system may have faults at the time. [6]

3 Applied Research

3.1 Tennessee Eastman Process

TE (Tennessee Eastman) process is proposed by Downs and Vogel of the U.S. Eastman Chemical Company in 1990. It is a real chemical process model for the development, research and evaluation of process control technology and monitoring methods. Many foreign scholars and experts use it as a data source to research the control algorithm, optimization and fault diagnosis. [7, 8]

TE process has five major unit operations: the reactor, the product condenser, a vapor-liquid separator, a recycle compressor cycle and a product stripper. There are four reactions and two kinds of products. The process has 12 manipulated variables and 41 measurements (including 22 continuous measurement variables and 19 composition measurements). TE process control structure shown in Fig. 1.

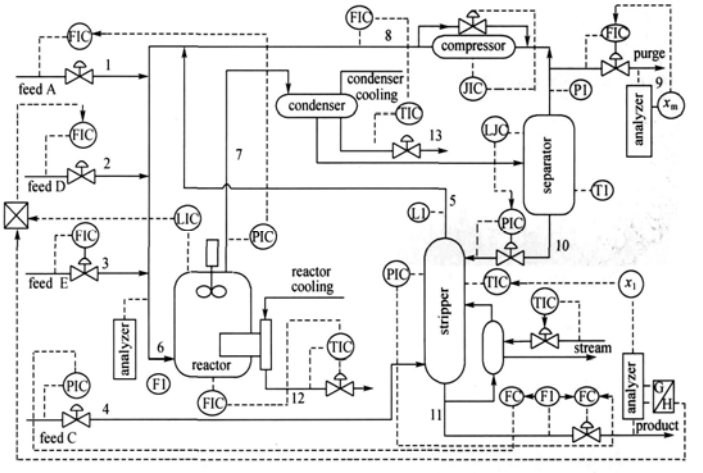


Fig. 1. Tennessee - Eastman process control structure

3.2 Simulation Results and Analysis

KPCA algorithm is applied to the fault detection of TE process. In the simulation, the kernel function uses Gaussian kernel: $k(x, y) = \exp(-\|x - y\|^2 / \sigma)$, kernel function σ is 500, and monitor the SPE statistic. Figures 2 to 5 are the SPE charts of the fault 5, 7,8,12.

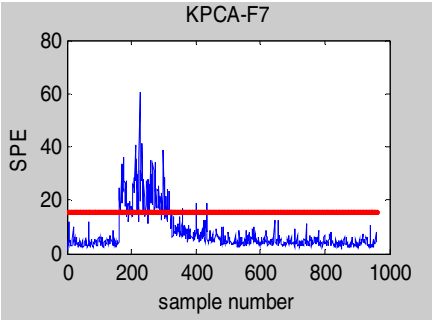


Fig. 2. Detection of fault 5

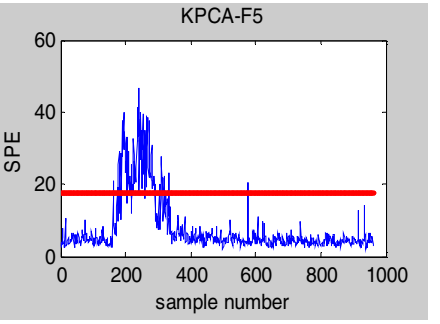


Fig. 3. Detection of fault 7

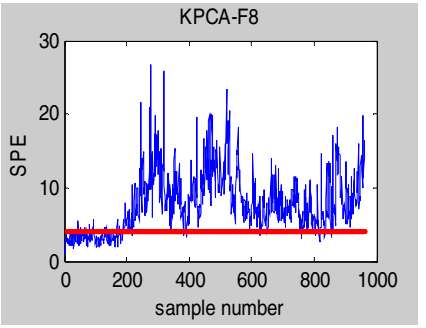


Fig. 4. Detection of fault 8

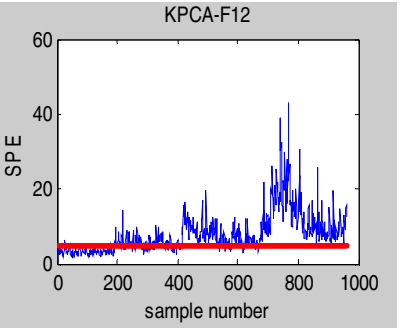


Fig. 5. Detection of fault 12

Fault 5, 7 belong to a step change and fault 8 and 12 belong to random changes in the system fault. When the fault occurs (after the 160 data points), SPE statistics produced significant changes. The result of KPCA in the fault diagnosis can clearly show whether the fault occurred. The use of KPCA method for chemical process fault detection can achieve better results, but it is Lack of capacity in the identification of fault types.

4 Improved Research

4.1 DDAG

DDAG is a new learning framework for support vector machine. [9] As a more complete multi-class classification algorithm, it can solve the identification of fault type excellently.

DDAG is proposed to solve the phenomenon of refusal and misclassification of “One-Against-One "SVM. In the training phase, the sub-classification of DDAG is like the sub-classification of the “One-Against-One” SVM, and need construct the surface of each classification between two. For a K-class classification problem, $k(k-1)/2$ classifiers are needed. In the testing phase, the approach constructs an

acyclic graph with two value-oriented and a root, which have $k(k-1)/2$ leaf nodes and k internal nodes. Each internal node corresponds to a binary support vector machine classifier, and each leaf node corresponds to a class mark. When an unknown sample needed to be the classified, it starts from the top of the root and reaches a leaf node until the bottom, which is the type of unknown sample. Node of the next layer is selected According to the classification of root node. [10]

Before the Fault diagnosis of DDAG, code known fault in accordance with binary numbers, and give the corresponding explanation. To Tennessee - Eastman process, for example, a known fault libraries can be established. According to T^2 and SPE charts, selecting the fault data as DDAG training input vectors and selecting binary number as the target vector, Stable DDAG fault diagnosis model can be obtained after training. When we enter a sample of fault, the type of fault can be drawn.

4.2 Hybrid diagnostic system of KPCA-DDAG

Because of the deficiency of the KPCA method in fault identification and diagnosis, DDAG is introduced to establish hybrid diagnostic system of KPCA-DDAG. DDAG fault classification method is precisely to make up for lack of KPCA. Hybrid diagnostic system of KPCA-DDAG is divided into two parts, KPCA detection and DDAG fault Identification. Flow Chart of Hybrid fault diagnosis system is shown in Figure 6.

For the hybrid diagnostic system, process data is firstly transported to the detection module of KPCA for fault detection. If no fault is detected, the next set of data will be transported to the detection module. When T^2 and SPE statistic exceed its threshold, it indicates there is a fault. Then the program enters the DDAG classification model, type of the fault will be identified out.

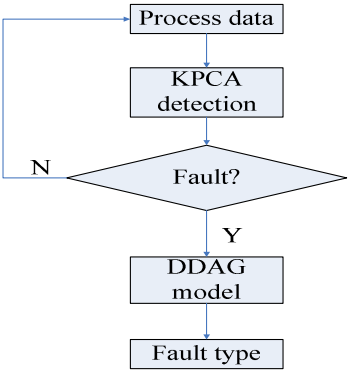


Fig. 6. Flow Chart of Hybrid fault diagnosis system

5 Conclusion

In this paper, the DDAG and KPCA have been combined to improve the fault diagnosis system. It uses DDAG classification model to identify the type of fault

when there is a fault detected by KPCA. The new way is given to enhance the capability of fault detection and diagnosis by the establishment of the new hybrid diagnostic system. And it shows excellent performance in the fault detection and diagnosis of Tennessee - Eastman process.

References

1. Schölkopf, B., Smola, A., Müller, K.-R.: Kernel principal component analysis. In: Advances in Kernel Methods-Support Vector Learning, pp. 196–201. MIT Press, Cambridge (1999)
2. Kim, K.I., Park, S.H., Kim, H.J.: Kernel principal component analysis for texture classification. IEEE SPL 8(2), 39–41 (2001)
3. Lee, J.M., Yoo, C., Choi, S.W., et al.: Nonlinear process monitoring using kernel principal component analysis. Chemical Engineering Science 59(1), 223–234 (2004)
4. Platt, J.C., Cristianini, N., Shawe-Taylor, J.: Large margin DAGs for multiclass classification. Advances in Neural Information Processing Systems 12, 547–553 (2000)
5. Cristianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines. University Press, Cambridge (2000)
6. Choi, S.W., Lee, C., Lee, J.M., et al.: Fault detection and identification of nonlinear processes based on kernel PCA. Chemometrics and Intelligent Laboratory Systems 75(1), 55–67 (2005)
7. Downs, J.J., Vogel, E.F.: A plant-wide industrial process control problem. Computers and Chemical Engineering 17(3), 245–255 (1993)
8. Lyman, P.R.: Plant-Wide Control Structures for the Tennessee Eastman Process, Lehigh University (1992)
9. Hsu, C., Lin, C.: A Comparison of Methods for Multiclass Support Vector Machines. IEEE Trans. on Neural Networks 13(2), 415–425 (2002)
10. Chapelle, O., Vapnik, V., Bousquet, O., Mukherjee, S.: Choosing Multiple Parameters for Support Vector Machines. Machine Learning 46(1-3) (2002)

Optimized Formation Control for Motor Schema-Based Multiple Robots

Qiang Gao[1,2], Yi Pang[2], and Dong Hao Lv[2]

[1] School of Electrical Engineering & Automation of Tianjin University, 300072 Tianjin, China
[2] Tianjin Key Laboratory for Control Theory & Applications in Complicated Systems, Tianjin University of Technology, 300384 Tianjin, China
gaoanbei@163.com

Abstract. Motor Schema-based formation control system is chosen to be the main research direction. Five sub-behaviors called move-to-goal, avoid-obstacle, swirl, noise and keep-formation are introduced for the line formation control mission. Using Teambots simulation system, the line formation is simulated. Neighbor-referenced method for formation position determination has been adopted. Aiming at the line formation, single-neighbor-referenced and double-neighbor-referenced methods are used. The simulations experiments data indicate that double-neighbor-referenced method reduces the running time, improves performance and achieve stability.

Keywords: Motor Schema, multiple robots, formation control.

1 Introduction

In the research of robot motion control, the robot behavior is decomposed into a series of sub-behaviors in the behavior-based control system in accordance with the mission requirements. This system establishes direct contact between sensors and actuators to achieve fast response to unknown environment. An effective behavioral synthesis strategy is provided in the Motor Schema-based system which is more flexible than common behavior-based control system [1]. The formation missions are implemented as motor schemas in this paper. Each schema generates a vector representing the desired behavioral response (direction and magnitude of movement) given the current sensory stimuli provided by the environment. A gain value is used to indicate the relative importance of the individual behaviors. The high-level combined behavior is generated by multiplying the outputs of each primitive behavior by its gain, then summing and normalizing the results. Motor Schema-based system takes into account all the sub-behaviors, which facilitates real-time performance, and the modular construction of schemas for ease in the development and debugging new behavior patterns.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 557–560.
springerlink.com

2 Motor Schema-Based Formation Control

2.1 Selection of Sub-Behaviors

The Motor Schema-based system includes control structure and control parameters. The control parameters (the gain value of the sub-behaviors) will be given in the simulation. The control structure chosen in accordance with formation control mission. Five motor schemas, move-to-goal, avoid-obstacle, swirl, noise and keep-formation implement the overall behavior for a robot to move to a goal location while avoiding obstacles, collisions with other robots and remaining in formation.

Move-to-goal: This sub-behavior generates a vector which is in the direction towards perceived goal.

Avoid-obstacle: The direction of this vector is along a line from robot to center of obstacle, moving away from obstacle.

Swirl: This schema will generate the vector which makes the robot move round obstacle from one side to the other side of the obstacle.

Noise: An additional background schema, noise, serves as a form of reactive "grease," dealing with some of the problems endemic to purely reactive navigational methods such as local maxima, minima and cyclic behavior.

Keep-formation: Once the desired formation position is known, the keep-formation motor schema generates a movement vector toward it. The vector is always in the direction of the desired formation position. Fig. 1 illustrates three zones, defined by distance from the desired position. Robot 3 attempts to maintain a position to the right of and abeam Robot 1. Robot 3 is in the controlled zone, so a moderate force toward the desired position (forward and left) is generated by keep-formation.

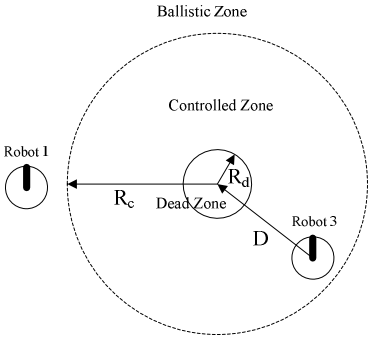


Fig. 1. Zones for the keep-formation

2.2 Optimized Neighbor-Referenced Method

Each robot computes its proper position in the formation based on the locations of the other robots. Three methods for formation position determination have been identified: Unit-center-referenced, Leader-referenced and Neighbor-referenced. Neighbor-referenced was chosen in this paper [2].

The Neighbor-referenced is a single-neighbor-referenced method. It means that robot i determine its desired position relative to robot (i-1). If robot 1 behind its desired position and robot 0 move toward the target point at maximum speed at the same time, in any case robot 1 can not get its desired position in time. This would make the formation broken.

Taking into account the above problem, double-neighbor-referenced method is adapted to solve it. It means that robot i and robot (i-1) in the formation determine its desired position relative to each other. For example, robot 1 computes its desired position relative to robot 0, while robot 0 maintains its desired position relative to robot 1.

3 Results in Simulation

The simulation results are generated using Georgia Tech's Teambots robot simulation environment. The simulation environment is a 60 by 30m two-dimensional field upon which various sizes and distributions of circular obstacles can be scattered. The origin of coordinates is in the center of the field. The robots are displayed as black circles whose radius is 0.35m, while the obstacles are gray circles. The robots' paths are depicted with solid lines. The robot's sensor radius is 12m in all directions. The gains used for the experimental simulations reported in this article are listed in Table 1. The running time is employed to evaluate the performances. The running time is the time of the robot move from starting point to the target. A lower value for this time indicates better performance.

The running time data in our simulation are shown in Table 2. It can be seen from the experimental data that double-neighbor-referenced method has lower value for the running time. The robot equipped single-neighbor-referenced method always can not catch its neighbor in time. This would lead to prolong the running time. The double-neighbor-referenced method can make the performance better.

Table 1. The gain values of the sub-behavior

Sub-behavior	move-to-goal	avoid-obstacle	swirl	noise	keep-formation
Gain value	0.8	1.5	2.0	0.1	1.8

Table 2. The running time in the simulation

The simulation times	Single reference point	Double reference point
1	174.0s	156.9s
2	174.8s	156.4s
3	178.7s	156.5s
4	175.2s	156.7s
5	174.7s	155.7s
6	174.9s	156.7s
7	178.0s	156.5s
8	174.6s	156.4s
9	174.9s	155.7s
10	178.4s	155.7s

Different performances between the two referenced methods can be seen as the line formation of robots moves around obstacles (Fig. 2). In Fig. 2, we can see robots need to keep formation and avoid obstacles simultaneously in the area [-12, 8]. The double-neighbor-referenced method has better performance in this area. For example, in the area [-8, -4], the distances among robot 3, 4, 5 equipped single-neighbor-referenced method change greatly, but the double-neighbor-referenced formation is remained well. Meanwhile, the double-neighbor-referenced method can make the formation recover more quickly. The robots equipped single-neighbor-referenced method has to recover their line formation at $X = -16$, but double-neighbor-referenced method can achieve it at $X = -12$. In general, double-neighbor-referenced method fares better than single-neighbor-referenced method.

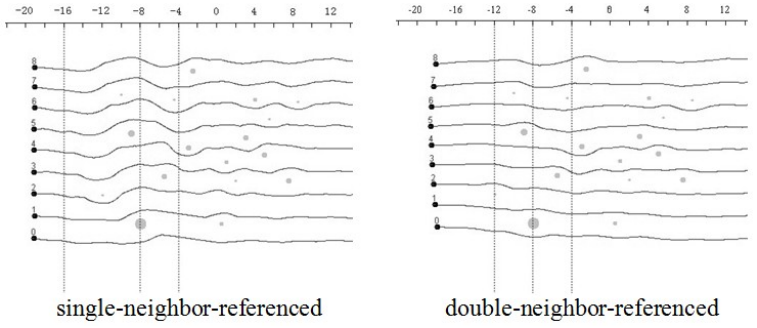


Fig. 2. A comparison of two referenced methods

4 Conclusion

The Motor Schema-based control is chosen in this simulation to achieve the mission that 9 robots keep line formation to arrive at the target area in the complex, unknown environment. According to the mission, we compare the performances of single-neighbor-referenced and double-neighbor-referenced method. The simulations experiments indicate that double-neighbor-referenced method reduces the running time, improves performance and achieve stability.

References

1. Arkin, R.C.: Motor Schema-based mobile robot navigation. The International Journal of Robotics Research 8, 92–112 (1989)
2. Balch, T., Arkin, R.C.: Behavior-based formation control for multi robots teams. IEEE Trans on Robotics and Automation 14, 926–939 (1998)

Stability of a Special Discrete Hopfield Neural Networks

Weigen Wu[1], Xing Yin[2], and Zhinong Miao[1]

[1] School of Information and Electronic Engineering, Pan Zhi Hua University, Panzhihua, 617000, China
[2] College of Computer, Pan Zhi Hua University, Panzhihua, 617000, China
weigenwu@126.com

Abstract. In paper [8] we mainly discuss a new kind of special discrete Hopfield neural networks (DHNNs) . Based on the paper [8], we consider stability of another special DHNNs. With the general permutation theory, a new permutation that two value is introduced. A new stability condition of this DHNNs is obtained.

Keywords: Discrete Hopfield neural networks, dominant matrix, stability.

1 Introduction

In regard to the work of stability research of Discrete Hopfield neural networks (DHNNs), most mainly discuss symmetric、anti-symmetric and nonnegative definite matrix[1-5]. Because most neural networks are asymmetric, the restriction on the symmetric is unpractical. Xu Z. B., eatl.[4,5] prove thatany Hop_eld networks will converge to a stable state when operating in the serial mode, if the connected matrix is main diagonally dominant matrix(MDDM); and the convergence of Hop_eld networks in the parallel mode is also veried under corresponding nonnegative deniteness conditions. In recent year, the best improvement of this work is that Donq-Liang Lee extends the serial(or parallel) updating mode to partial simultaneous updating mode[6] and R. N. Ma gives the weight matrix an increment matrix, which is row or column diagonally dominant[7].

However, we know that the dominant effect of every neuron is the connected weight from oneself to oneself. The restriction is not good for the dominant effect of one neuron may be other neuron. In this paper we will consider that the weight matrix of arbitrary DHNNs is the sum of the special DHNNs (that is $W = \sum_{1\le k\le n} W^k$) and the weight matrix of the special DHNNs is extended to nonnegative definite matrix. Let $P^k = \left(W^k, \theta\right)\left(1 \le k \le n\right)$ be a DHNNs. $W^k = \left\{w_{ij}^k\right\}_{n\times n}$ is denoted by

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 561–564.
springerlink.com

$$W^k = \begin{bmatrix} 0 & \cdots & 0 & w_{1k}^k & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & w_{(k-1)n}^k \\ w_{k1}^k & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & w_{n(k-1)}^k & 0 & \cdots & 0 \end{bmatrix} \tag{1}$$

2 Main Result

We mainly study the DHNNs combined with two special DHNNses. Some results are obtained.

We know that arbitrary DHNNs N may consist of P^k ($1 \le k \le n$). Now we consider P^k, P^u, and $P = \left(W^k + W^u, \theta\right)$. In this paper, P, P^k and P^u have the same neurons set $G = \{1, 2, \cdots, n\}$. According to paper , We know that in $P^k\left(1 < k \le n\right)$ the connection among neurons set $H = \{1, 2, \cdots, n\}$ is a permutation

$$d^k = \begin{pmatrix} 1 & 2 & \cdots & n-k & \cdots & n \\ k & k+1 & \cdots & n & \cdots & k-1 \end{pmatrix} = \left(\left(H_1^k\right)\cdots\left(H_b^k\right)\cdots\left(H_{\frac{n}{r_k}}^k\right)\right) \tag{2}$$

where ${}^k r$ is the minimum solution of equation

$$\left(k + \left(k-1\right)\left(r_k - 1\right)\right) \bmod n = 1 \tag{3}$$

${}^k H_b$ is a circle set. The permutation ${}^k d$ is divided up to $\frac{n}{{}^k r}$ cycles of ${}^k r$ order. Then neuron set $H = \{1, 2, \cdots, n\}$ is divided up to $\frac{n}{{}^k r}$ sets (that is $H = \bigcup_{1 \le b \le n/{}^k r} {}^k H_b$).In P, the behavior of one neuron effect other two neurons that d^k and d^u work simultaneously and it is denoted by

$$\begin{aligned} d^{ku} &= \begin{pmatrix} 1 & 2 & \cdots & n-u+1 & \cdots & n-k+1 & \cdots & n \\ k & k+1 & \cdots & k+n-u & \cdots & n & \cdots & k-1 \\ u & u+1 & \cdots & n & \cdots & u-k & \cdots & u-1 \end{pmatrix} \\ &= \left(\left(H_1^{ku}\right)\cdots\left(H_l^{ku}\right)\cdots\left(H_L^{ku}\right)\right) \end{aligned} \tag{4}$$

This means that starting with 1, we arrive at k and u by the connection weight w_{1k}^{k} and w_{1u}^{u}, then we arrive at $(k+k-1)\bmod n$, $(k+u-1)\bmod n$, and $(u+u-1)\bmod n$ from k and u, …, we arrive at $\left(k+(r_k-1)(k-1)\right)\bmod n$ and $\left(u+(r_u-1)(u-1)\right)\bmod n$, finally we arrive at 1. This process may integrate d^k with d^u. With (2) and (4), we obtain the union $d^{ku}=\left(\left(H_1^{ku}\right)\cdots\left(H_l^{ku}\right)\cdots\left(H_L^{ku}\right)\right)$ of sets as follow:

We obtain H_l^{ku}, as follow(**C Program 1**):

r is solved by $\left(k+(k-1)r\right)\bmod n=1$; $l=1$;

While ($H\neq\varnothing$)

{ define set $H_l^{ku}=\varnothing$;

$\vee p$ ($p\in H$);

$H_l^{ku}=H_l^{ku}\cup\{p\}$;

For($i=1$; $i\le r$; $i++$)

{For($j=1$; $j\le i$; $j++$)

{ $H_l^{ku}=H_l^{ku}\cup\left\{\left(p+i\times(k-1)+j\times h\right)\bmod n\right\}$;

$j=j+1$;

}

$G=G-H_l^{ku}$;

$l=l+1$;

}

$L=l-1$;

Set H divide into L groups. Based on the symmetry of P, we know that ${}^{ku}H_l$ is a circle set and the permutation ${}^{ku}d$ is divided up to n/L cycles of L order.

Thermo 1: Let $P^k=\left(W^k,\theta\right)$ and $P^u=\left(W^u,\theta\right)$ ($1\le k<u\le n$,) be DHNNses. $P=\left(W^k+W^u,\theta\right)$, where $w_{ij}^{k}\ge\theta_i\left(1\le i<k,k\le j\le n\right)\wedge w_{ij}^{k}\ge\theta_i\left(k\le i\le n,1\le j<k\right)$ and $w_{ij}^{u}\ge\theta_i\left(1\le i<u+1,u+1\le j\le n\right)\wedge w_{ij}^{u}\ge\theta_i\left(u+1\le i\le n,1\le j<u+1\right)$. To arbitrary initial state $X(0)$ excepted for $(1,1,\cdots1)^T$ and $(-1,-1,\cdots-1)^T$, P always converges to a cycle attraction factor with the length of n/L.

Proof: with $x_j(t+1)=\operatorname{sgn}\left(\sum_{j=1}^{n} w_{ij}x_j(t)-\theta_i\right)$ where $\operatorname{sgn}(x)=\begin{cases}1, x\geq 0,\\ -1, x<0.\end{cases}$ and the symmetry, we know that $H_l^{ku}\left(1\leq l\leq L\right)$ has n/L elements. Based on **thermo 1** of paper [8], the state of neuron of $H_l^{ku}\left(1\leq l\leq L\right)$ converges to a cycle attraction factor with the length of n/L. Now we consider the combination of $H_l^{ku}\left(1\leq l\leq L\right)$ that the initial state $X_l(0)$ is not $(1,1,\cdots 1)^T_{1\times n/L}$ or $(-1,-1,\cdots,-1)^T_{1\times n/L}$. $X_l(0)$ is the initial state of ${}^{ku}H_l$. According to **thermo 3** of paper (I), the least common multiple is n/L. So, if $\exists X_l(0)=(1,1,\cdots 1)^T_{1\times n/L}\ or\ (-1,-1,\cdots,-1)^T_{1\times n/L}$, P always converges to a cycle attraction factor with the length of n/L.

When $X(0)=(1,1,\cdots 1)^T\ or\ (-1,-1,\cdots -1)^T$, ${}^kP$ always converges to a stable state $(1,1,\cdots 1)^T\ or\ (-1,-1,\cdots -1)^T$. This completes the proof.

References

1. Hopfield, J.J.: Neural networks and physical systems emergent collective computational abilities. Proc. Nat. Acad. Sci., USA 79, 2554–2558 (1982)
2. Goles, E., Fogelman, F., Pellegrin, D.: Decreasing energy functions as a tool for studying threshold networks. Discrete Applied Mathematics 12, 261–277 (1985)
3. Goles, E.: Anti-symmetrical neural networks. Discrete Applied Mathematics 13, 97–100 (1986)
4. Xu, Z.B., Kwong, C.P.: Global Convergence and Asymptotic Stability of Asymmetric Hopfield Neural Networks. Mathematical Analysis and Applications 191(3), 405–427 (1995)
5. Xu, Z.B., Hu, G.Q., Kwong, C.P.: Asymmetric Hopfield-type Networks: Theory and Applications. Neural Networks 9(3), 483–501 (1996)
6. Lee, D.L.: New stability conditions for Hopfield neural networks in partial simultaneous update mode. IEEE Trans. on Neural Networks 10, 975–978 (1999)
7. Ma, R.N., Zhang, S., Lei, S.: Stability Conditions for Discrete Neural Networks in Partial Simultaneous Updating Mode. ISNN (1), 253–258 (2005)
8. Li, J., Yang, J., Wu, W.G.: Stability analysis of discrete Hopfield neural networks with column arbitrary-dominant weight matrix, revised manuscript submitted to Neurocomputing (2011)

Design and Realization of Real-Time Image Acquisition and Display System Based on FPGA

Honglu Hou, Wenfang Zhang, Dingjin Huang, and Tao Zhang

Department of Photoelectric, Xi'an Technological University, Xi'an, 710032, China
qie1213@163.com

Abstract. A real-time image acquisition and display system based on FPGA is proposed, to against the issue of the traditional image acquisition system, which includes: complex hardware circuit, high cost, long development cycle, not suitable for secondary development and design, and hard to meet the applications which are highly request for real time. The idea of the modular and integration design is adopted in the period. The real- time image acquisition and display is realized by combining with the technologies of color interpolation, line buffering, pipeline, data joining and SDRAM partition storage. The working principle and realization method of each function module is introduced in details. In the end, the system performance test is done, the result certificates that it's flexible, reliable, small volume, low cost, and meet requirement of the high performance real-time image acquisition system in field of industry and scientific research.

Keywords: Real-time Image Acquisition, Display, FPGA, CMOS Image Sensor; VGA.

1 Introduction

The traditional image acquisition system is based on PC and video capture card, due to the division element, the system is complex, huge, slow, hard to debug and not inconvenience for carrying, so it can't satisfy the modern image processing technology's demand for high real-time. We consider applying embedded technology to the real-time image acquisition system, which can improve its collection and transmission rate, enhance its stability and real-time and reduce its volume and power consumption. At present, several plans for embedded image acquisition system which are based on ASIC or DSP or FPGA, etc has existed in foreign and domestic. In comparison to the ASIC chip's poor flexibility, long design cycle and the DSP chip's complex circuit design, not suitable for secondary development, using the FPGA chip as video collection chip has its unique advantages-strong parallel processing ability and pipeline technology, flexible online dynamic configuration, processing speed, the system portability is strong, flexible online dynamic configuration, high processing speed, strong system portability, so it is very suitable for real time video collection, and has a vital significance for improving the economic benefit and shortening the development cycle and reducing cost.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 565–573.
springerlink.com

Based on the above reasons, by using the CycloneII series FPGA (EP2C35F672C6) on Altera DE2 Development and Education board as core control device and the board resources and Terasic CMOS Image Sensor TRDB-D5M as hardware configuration, we successfully realized the real-time image acquisition and VGA display on DE2 development board.

2 System Frame Work and Work Principle

The whole collection system mainly includes the following five modules: Timing driving module, image capture module, image format conversion module, SDRAM controller module and VGA sequential control module. Each module are integrated in the FPGA, and realized comprehensive through Verilog language in QuartusII compile environment. The system framework is shown in figure1.

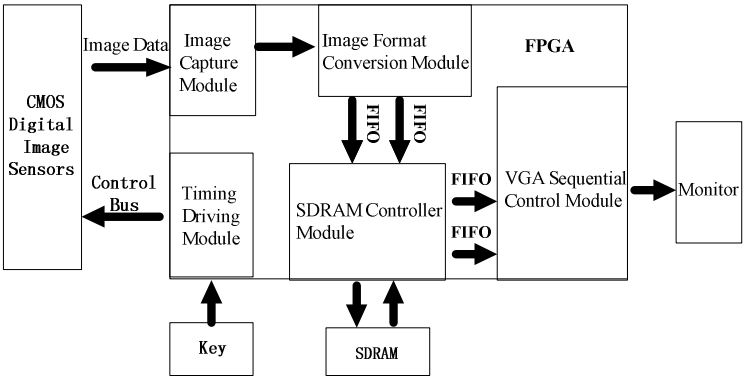


Fig. 1. The system framework

The front image sensor is Terasic CMOS Image Sensor TRDB-D5M (2592 x 1944 resolution), which its output are 12 roads of digital signal and get parameter setting through I2C bus. The display uses ordinary desktop computer monitor, other hardware are all use DE2 board resources, Main control chip is Altera EP2C35F672C6, SDRAM is PSC A2V64S40CTP. In the design we use LED, seven-segment digital tube, etc to help testing operation of the system.

Its working principle is: The key on board triggers camera starts, then the timing driving module transfers default parameters to CMOS image sensor through I2C bus to realize the TRDB-D5M internal registers' configuration. When the camera is initialized, the output data streams flow into the FPGA image capture module in digital form directly through the GPIO on DE2 development board. Then after line and frame identification, the collected Bayer type valid data convert to RGB format through the image format conversion module. In the end, the three road RGB video data streams are stored in SDRAM separately for reading and displaying of the

subsequent VGA sequential control module through two roads of FIFO opened up in FPGA by data joining. The keys and slide switches change the image sensor's registers value through I2C bus to adjust its focal length and exposure.

3 Design and Realization of Image Acquisition

3.1 Timing Driver Module

This module mainly responsible for sending configuration information to the image sensor, including exposure, RGB gain, row size and column size, to realize the image sensor's parameter configuration, drive image sensor to collect environmental information and transmit the collected digital video data streams to FPGA.

The configuration is realized through I2C bus protocol. This protocol is a serial bus consists by the data line (SDA) and the clock line (SCL). The SDA and SCL realize control through setting high/low separately according to the bus protocol. The control signal is divided into the address part and the control part. The address part is used for location, namely select the TRDB-D5M register that need to control. The control part is used to set the register parameter and realize the parameter adjustment such as resolution, brightness and gain.

Based on the I2C bus protocol above, the timing driving module writes image sensor register addresses and data in turn. We assigned the focal length iZOOM_MODE_SW, exposure iEXPOSURE_ADJ and gain iEXPOSURE_DEC_p at the module input port in turn to SW [16], KEY [1] 和SW [0] on DE2, While doing pins allocation. So that we can use the keys and slide switches to adjust exposure and focal length of the image sensor instantly.

3.2 Image Capture Module

Image capture module gets digital video signal from image sensor output port through the GPIO on DE2, which mainly contains: fame mark signal iFVAL, line mark signal iLVAL and pixel signal [11:0] iDATA, namely the input port signal of the image capture module.

FPGA judges the begin of each image by testing the rising edge of the fame mark signal iFVAL, then show the fame on the seven-segment digital tube through counting by frame counter.

While iFVAL is in high level and iLVAL is at the rising edge, the pixel turn to valid at iDATA port, and the column counter gets reset and start counting. While the counter reaches to 1280, it gets reset again, and the line counter increase by one. Repeat the above operation until the column counter reach to 960 each time when iLVAL is at the rising edge. Then clear the column counter, and a frame image transmission has finished. Repeat the above operation each time iFVAL is at the rising edge.

The image capture module collects valid pixels through fame and line judges as show above and then exports pixel information which mainly contains: rows/columns count value, namely pixel position oX_Cont, oY_Cont, valid pixel mark oDVAL, and pixel value [11:0] oDATA, namely the output port signal of the image capture module.

4 Design and Realization of Image Display

4.1 Image Format Conversion Module

Due to the pixel of image sensor DRB- D5M exports in Bayer format and VGA sequential control module in back-end displays in RGB format, it needs to convert the image format. On the premise that no loss of image detail, a method based on color interpolation is proposed. Pixels are merged with a pixel values to replace the adjacent four pixels. The combined method is shown in figure2.

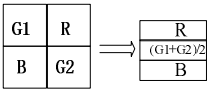


Fig. 2. Color interpolation transform

Due to the front image acquisition module acquisition pixels for 1280 * 960, the resolution of image reduce to half through image format conversion module, namely the VGA resolution 640 * 480.

According to the transform principle above, it needs to deal with at least two lines of data at the same time to make the Bayer format convert to RGB format. Therefore, we open up a 8-digits shift register by the FPGA resources M4K and set two tap spacing for 1280 (a line),and realize to process two rows at the same time with the technologies of line buffering, pipeline. The hardware diagram of shift register is shown in figure3.

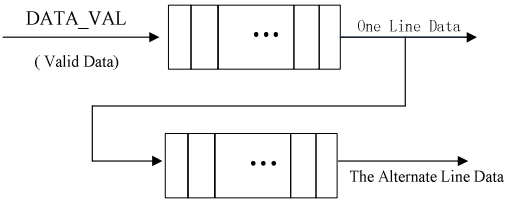


Fig. 3. The hardware diagram of shift register

Through opening two tap the shift register, combined with the algorithm thought of color interpolation above and convert it into verilong source code, the image format conversion module is finally realized.

4.2 SDRAM Controller Module

Through the collection of image capture module and the conversion of image format conversion module, the valid pixels are flown into VGA sequential control module for image show. Due to the different speed between front end and back end, it needs to cache images. With a SDRAM + FIFO form, the SDRAM controller module mainly

includes four FIFO that two is for reading and two is for writing and one SDRAM controller. Its structure diagram is shown in figure4.

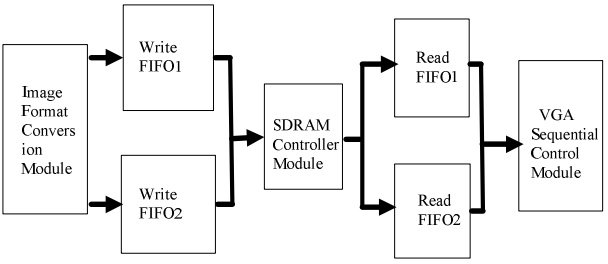


Fig. 4. The function diagram of SDRAM controller module

Here, the data port of write FIFO1 is correspond to {1'b0,sCMOS_G[11:7], sCMOS_B[11:2]}, the data port of write FIFO2 is correspond to {1'b0,sCMOS_G[6:2], sCMOS_R[11:2]}. That is to say, the top 10 digits of the 12 digits RGB pixels are taken from the image format conversion module, and then three roads of them are merged to two roads of 16 digits signal and transmitted to write FIFO1 and write FIFO2.

Due to the limited resources on DE2, the SDRAM is divided into two areas for reading and writing. In this way, we can do read/write in double simultaneously and speeding up the transmitting of pixels to VGA sequential control module.

The FIFO in this module is instantiated through QUARTUSII IP core. The realization of SDRAM controller is introduced in the following.

The pins of SDRAM device are usually divided into three kinds: control pins (LDQM, UDQM, CAS, CKE, CS, RAS, WR, BA[1:0]), address pins (ADDR[N:0]) and data pins (DATA[N:0]). LDQM and UDQM are used for masking, CKE is used for clock enable, CS clock is used for chip choosing, BA[1:0] is used for BANK choosing, CAS, CS, RAS, WR are used for operation control. The Specific instructions can be checked in SDRAM data manual, here no longer tautology.

Each time before it enters into the normal work state, SDRAM need initialization, which in turn includes: 200us input stable, precharge of all L-BANK, eight refresh cycles and mode register setting (MRS. The critical period is the mode register setting, which are used to set SDRAM operation mode, incubation period CAS, burst way and burst length.

After the mode register is set, SDRAM begin to enter into normal work state. First, enable the line that will write/read. Second, open the right column. Then make burst write/read to the choosed address. In addition, Due to the particularity of SDRAM storage structure, we need to rewrite all memories on the original effective line and reset its addresses, which is called precharge.

Based on the SDRAM working principle and time characteristics above, SDRAM controller should including three modules: command response module, command parse module and data path module. Command response module is used to receive instructions and produce SDRAM operating actions; CMD commands parsing module is used to receive CMD orders and decode them into operation instructions; Data path

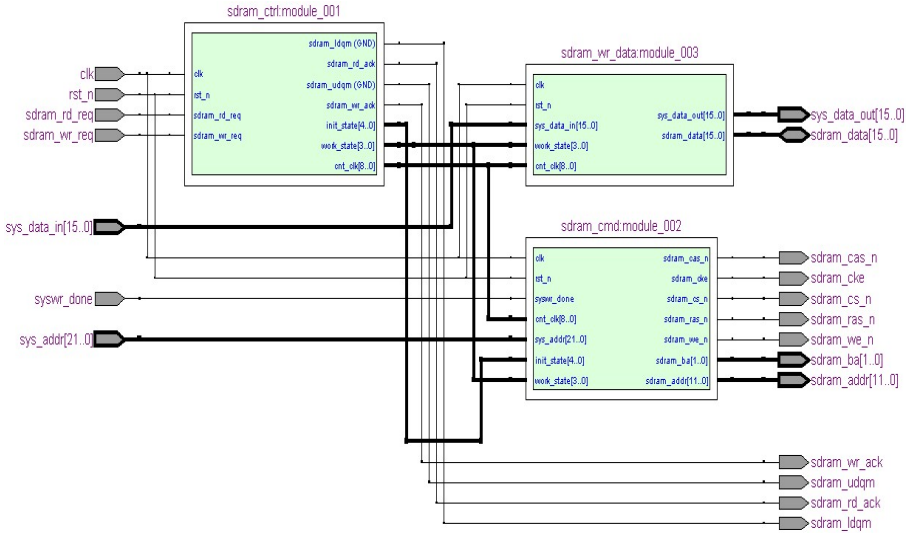


Fig. 5. The RTL view of SDRAM controller

module is used to control input/output of the valid data. The RTL view of SDRAM controller coding by FPGA is shown in figure5.

4.3 VGA Sequential Control Module

Image display module transmits video data stored in SDRAM temporarily to monitor for display. The VGA displays in the form of progressive scan, which start from the upper left of the screen and scans from left to right, top to bottom. After the scan of one line, it uses line synchronized signal for line synchronization. Then electron beams make a line vanish to monitor and return to the left starting position of the next line. When all lines are scanned, it uses frame synchronized signal for frame synchronization, and make the scan return to the upper left of the screen. At the same time, the electron beams make a frame vanish to monitor and prepare for scanning of the next frame.

Figure6 is the VGA industrial timing diagram (640 x 480, 60 Hz). According to the industrial timing, through FPGA compile, the VGA sequential control module is realized.

5 Image Acquisition Experiment and Test

The final code is got through code compiling, debugging and integrating of all modules above and downloaded to FPGA. The platform of real-time image acquisition and display system finally set up is show in figure7.

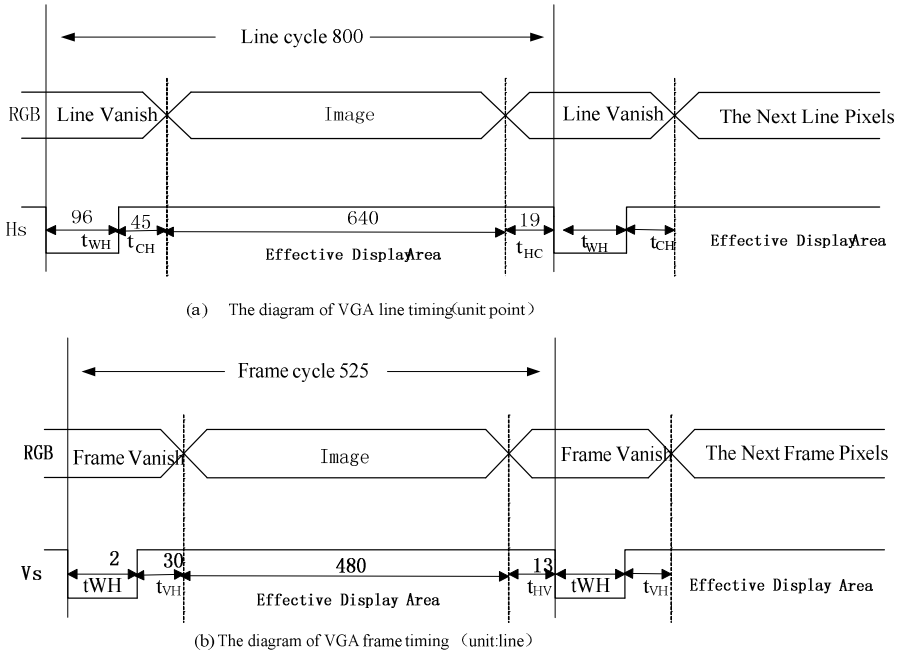


Fig. 6. The diagram of VGA industrial timing

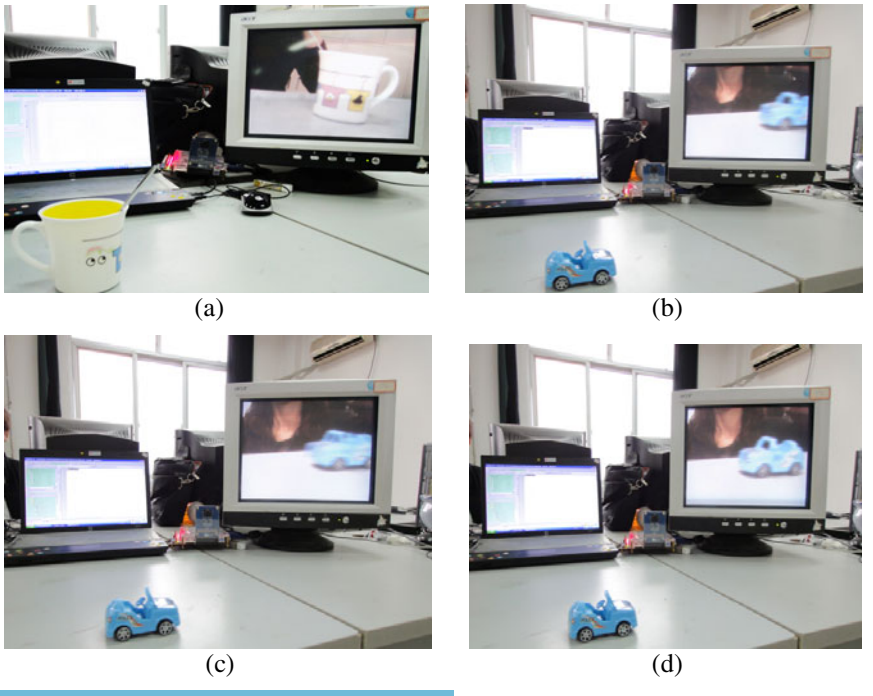


Fig. 7. The physical platform of real-time image acquisition and display system

Figure7 (a) is acquisition and display for static object (cup), figure7 (b), (c),(d) is acquisition and display for moving objects (car). We can do real-time adjust to the focal length of the image sensor through slide switches SW[16], and Real-time adjust the image sensor exposure through key KEY[1] and slide switches SW[0]. Through VGA Imaging, we can see the function of image acquisition and display is perfect, the image collected by system is clear and the real-time image acquisition and display is realized.

Through doing long time test, the system did not produce abnormal phenomenon. It shows that the system is stable and reliable, and meets the requirements of real-time image processing.

6 Conclusion

Through choosing high performance CycloneII series FPGA as image processing the core of the system, combined with Terasic company digital CMOS image sensor TRDB-D5M and PSC company SDRAM memory chip constructed a portable embedded system suitable for image acquisition. Based on overall design, through researching on interface control, resource allocation and algorithm processing, we realized the real-time video image acquisition, cache and display. The experimental results show that the system's external interface circuit is simple and is easy for use and transplantation. it can be applied to the real-time image acquisition and display system in the field of industrial and scientific research with advantages of small size, low power consumption and high speed. This system's design and implementation laid a foundation for the design of high frame frequency portable embedded image acquisition system in the next step.

References

1. Zhao, F.: Research on Image Capture and Preprocessing System Based on CycloneII series FPGA. Chongqing University, Chongqing (2009)
2. Gong, T.: Image Collection and Process System Based on FPGA. Huazhong University of Science&Technology, Wuhan (2005)
3. Si, X.: The design of FPGA real-time image acquisition system based on SDRAM. Journal of Southwest University (Natural Science Edition) 33(01), 128–132 (2011)
4. Yang, H., Tong, S.: Conversion from Bayer to RGB Based on FPGA. Modern Electronics Technique 2(313), 122–124 (2010)
5. The Research and design of Infrared Images PreProcessing System Based on CNN, http://wenku.baidu.com/view/3ef64f3f5727a5e9856a61bf.html
6. Master advanced-The ultimate memory technical guidelines, http://wenku.baidu.com/view/c81237ea81c758f5f61f67c2.html
7. Yang, Y.: Design of SDRAM Controller and Its Application on FPGA. Lanzhou University, Lanzhou (2007)

8. Cao, H., Bin, D.: Realization FPGA-based SDRAM Controller with Verilog. Electronic Products China (01), 53–55, 60 (2005)
9. Song, Y., Lei, J., Li, C.: Design of SDRAM Controller Based on FPGA. Electronic Engineer 29(09), 10–13 (2003)
10. Chen, B., Wu, Q., Liu, Y.: Design of VGA Control Module Based on FPGA. Microelectronics 38(2), 306–308 (2008)

A Coverage Algorithm for Multiple Micro Robots Inspection Based on Asynchronous Communication

Ling Mao, Dawei Zhang, Jiapin Chen, and Zhenbo Li

National Key Laboratory of Nano/Micro Fabrication Technology, Key laboratory for Thin Film and Microfabrication of the Ministry of Education, Research Institute of Micro and Nano Science and Technology, Shanghai Jiao Tong University, Shanghai 200240 P.R. China
maoling2290@126.com

Abstract. This paper presents a simple coverage algorithm for multiple micro robots inspection. The microrobot is designed as differential driven vehicle, and actuated by MEMS-based electromagnetic micromotors. It is required that the entire free space is max covered by the sensors of the robots. Miniaturization of robot dimension limits individual ability of each robot. This limitation leads us to employ simple formation coverage algorithm based on asynchronous communication. The proposed coverage algorithm has been tested and evaluated by experiments.

1 Introduction

Autonomous micro mobile robot is highly demanded to facilitate automatic operation in micro factory, narrow pipe, high risk etc environments, however, in practice, miniaturization of robot dimension limits development of individual ability of each robot. This limitation leads us to employ distributed mobile micro robot cooperation to finish the task. In distributed mobile micro robots cooperation systems, multiple micro robots(MMRs) can exchange sensor information, and collaborate to complete some task. MMRs are good candidates for remote inspection tasks, especially in man-made infrastructures where small structural dimensions stand in contrast to a large overall size of the entire system where small structural dimensions stand in contrast to a large overall size of the entire system[1],[2]. As a preliminary step towards realizing the long-term goal of an operational MMRs system capable of high confidence inspection, the current work focuses on several parallel technology development efforts such as search strategy, coverage completeness, obstacle avoidance and so on. The most important of these efforts is coverage algorithm which this paper focuses on. Traditional coverage algorithms become invalid for micro robot which is limited in calculation capacity, communication ability and especially energy. This paper presents a simple coverage algorithm for multiple micro robots inspection. Section 2 introduces the individual microrobot. The coverage algorithm of multiple micro robots is presented in section 3. Section 4 experimental evaluates the effective we present coverage algorithm. The conclusion is shown in section 5.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 575–579.
springerlink.com

2 Micro Robot

Our centimeter-scale mobile automous microrobot which mainly consists of three wheels, is designed as differential driven vehicle with two identical standard wheels and a caster. The volume is 28*28*30mm, as shown in Fig 1[3]. each standard wheel is directly actuated by a 6.8mm micromotor is designed as 3-phase permanent synchronous motor with the star-connected windings, and employs axial flux structure which can reduce the volume compared with radius flux[4]. The basic microrobot platform is designed and integrated, including the microcontroller base on ARM7TDMI-S CPU, which is characterized by tiny size (LQFP64: 10×10×1.4mm) and low power consumption (working mode: about 10mA at 10MHz); electromagnetic micromotor are derived by ATA6836 which integrates six independent half-bridge drivers(from Atmel, QFN24: 5×5mm) ; Long rang communication between robots is achieved by RF wireless transceiver (Nordic nRF24L01)which is a single chip radio transceiver for the world wide 2.4 - 2.5 GHz ISM band with very low current consumption only 9.0mA at an output power of -6dBm and 12.3mA in RX mode and power manager. To realize that the microrobot can carry out vision-based inspection task, an image acquisition module composed of Omnivision OV7670 and Mtek Vision MV3018 is designed. The OV7670 is a low The OV7670 is a low voltage 1/6 inch CMOS image sensor with 300K pixels (active array size: 640×480), which can provides the full functionality of a single-chip VGA camera and image processor in a small footprint package. Some additional input and output CPU ports are reserved which provide our scale application.

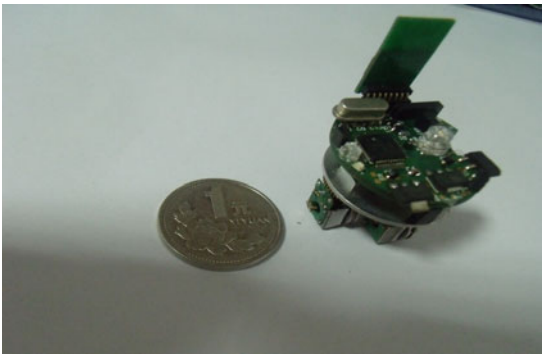


Fig. 1. Photograph of basic microrobot with a 22mm coin

3 Coverage Algorithm

3.1 Problem Statement

The problem space consists of an enclosed space, of which enclosing space is to be inspected continuously. The enclosing space is a continuous space. See Fig 2. Boundary is denoted by Black line and free space is shown by Grey solid regions. To

Fig. 2. Inspection Area with continuous free space

simplify, the inspection space doesn't have obstacles. It is required that the entire free space is max covered by the sensors of the robots.

3.2 Inspection Coverage Algorithm

Our micro robot is autonomous robot which can avoid obstacles which include inspection environment boundary and other robots, but there are some constraints on our micro robot which are as follows: Firstly, our micro robot doesn't have encoders for closed-loop displacement information from the wheels for miniature size. Instead, the electromagnetic micromotors allow to measure the number of steps in a open-loop manner. Secondly, for miniature size of micro robot, the power is limited. Finally, our micro robot has limited computation capability and memory ability, which is depicted above for robots' constrains. Our inspection coverage algorithm modifies Exact cellular decomposition like Boustrophedon Approach[5]. Whose core is keep a line formation and back and forth movement in inspection area.

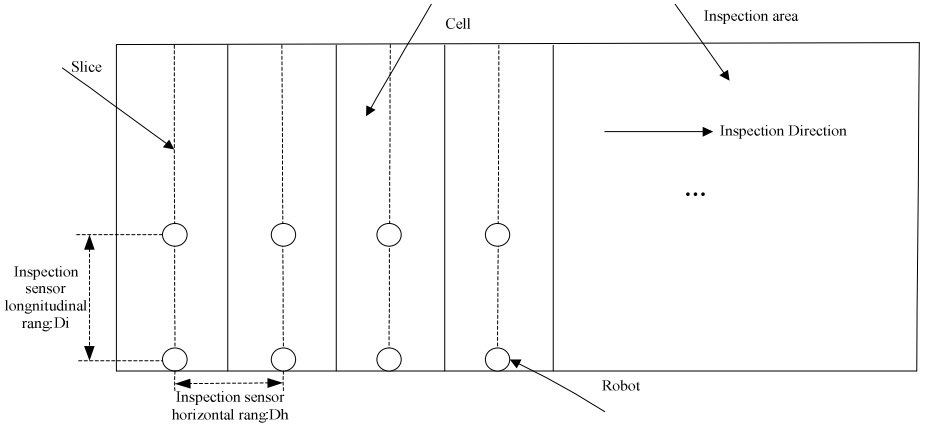


Fig. 3. Illustration of the terms borrowed and All robots start in horizontal formation

We can describe the algorithm run on robot as follows:

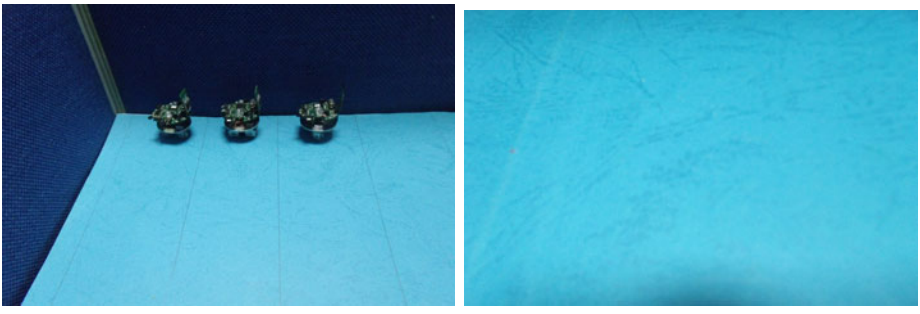
Step1: All robots start in horizontal formation keep a horizontal line formation and remember own localization. Take some photos to carry out vision-based inspection task and transmit the photos to host computer with RF communicate. The distance of between robots is the inspection sensor horizontal rang Dh and some terms used in our inspection algorithm are borrowed as shown in Fig 3.The IR sensors' detected rang Sr is greater than Dh.

Step2: All robots follow Slice movement at the speed of 5.4cm/s until they across the inspection sensor longitudinal rang Di and stop to carry out vision-based inspection to inspect Cell and transmit the photos to host computer. With the microcamera mounted on the basic microrobot whose power consumption accounts for about 60% of that of the microrobot, so the microcamera turn off during movement to save power consumption. The micro robot has global communication, Further more, can communicate with host computer with RF communicate. About this asynchronous pattern which we use, communication cycle is named as time which robot crosses rang Di spends. That is to say, Robots only communicate when they have crossed rang Di, which saves power and decreasing inspection time to realize the max area inspection. Robots modify the position error to keep a horizon line formation and transmit the photos to the host computer by communication. Robots follow the strategy which is mentioned above again and again until their inspection sensor meet the top or bottom boundary.

Step3: When their inspection sensor meet the top or bottom boundary, robots cross n*Dh distance in the horizontal direction keep a horizontal line formation, and then go on step1, step2 and step3 until all area are inspected or the energy is used up.

4 Experiment

The presented inspection coverage algorithm based on asynchronous communication was tested with multiple micro robots. The inspection experiment scene is shown as Fig 4. We do experiment use three robots. The area was completely inspected because the experiment environment which is very small. After experiment, the even power consumption accounts for about 70.6% that of the microrobot, which compare with 100% of only simple back and forth movement without asynchronous communication.



(a) Inspecting robots (b) Inspection image

Fig. 4. Inspection experiment scenes

5 Collusion

The presented inspection coverage algorithm based on asynchronous communication which is fit to multiple micro robots without considering obstacles environment which is our next step work. The algorithm can covers the max area and conserving energy, which was tested by experiment.

Acknowledgments. This work is supported by National Natural Science Foundation of China (Nos. 61175100).

References

1. Wong, E., Litt, J.S.: Autonomous Multi-agent Robotics for Inspection and Repair of Propulsion Systems. In: Proc. AIAA 1st Intelligent Systems Technical Conf. AIAA 2004-6364, Chicago, IL (2004)
2. Takeda, M., Namura, K., Nakamura, K., Shibaike, N., Haga, T., Takada, H.: Development of Chain-type Micromachine for Inspection of Outer Tube Surfaces (basic performance of the 1st prototype). In: Proc. IEEE Int. Conf. Micro Electro Mech. Syst., pp. 805–810. IEEE (2000)
3. Zhang, D., Li, Z., Ling, M., Chen, J.: A Mobile Self-reconfigurable Microrobot with Power and Communication Relays. International Journal of Advanced Robotic Systems 7, 48–57 (2010)
4. Li, Z., Chen, J., Ge, X., Zhang, C.: Novel Microstepping Control Methods for Electromagnetic Micromotors with Star-connected winding. Sens. Actuators A Phy. 134(2), 513–518 (2007)
5. Choset, H., Pignon, P.: Coverage Path Planning: The Boustrophedon Cellular Decomposition. In: Proceeding of the International Conference on Field and Service Robotics, Canberra, Australia (1997)

Adaptive Backstepping Control for Rotate Control of Wind Power with Hydraulic Transmission

Hongbin Wang, Wenzeng Gao, and Shaochan Feng

Key Lab of Industrial Computer Control Engineering of Heibei Province,
Yanshan University, Qinhuangdao 066004, China
hb_wang@ysu.edu.cn, {gaowz823,beautifulfeng520}@163.com

Abstract. For the rotor speed control of the grid-connecting of wind turbines with hydraulic transmission, the system dynamics model is built. To address the difficulty in accurate control of rotor speed for external disturbance and uncertain internal disturbances while connecting to the grid, an adaptive Backstepping controller was designed. The convergence was proved with Lyapunov stability theory. The condition of a constant rotor speed for connecting to grid was meted. The simulation results show that the systerm provides a good dynamic response and strong robustness when system parameters and load torque disturbances appeared.

Keywords: Wind turbines, Hydraulic transmission, Rotor speed control, Adaptive Backstepping.

1 Introduction

Wind power is one of the greatest potential renewable energy sources, people pay more attention to wind power recently. The wind turbine type and grid method were investigated in [1]. At present, doubly-fed induction generator (DIFG) and direct-driven permanent magnet synchronous generators (PMSG) are the most successful applications in the field of wind power. The DIFG and no-load grid-connecting of DIFG were researched in [2-4]. However, the development of DIFG is restricted by the heavy gear box, for gear box is easy to overload and damage. Although the gear box is not needed in the PMSG, it is needed to adopt the whole power converter, Meanwhile, the generator is big and expensive, thus the system cost is very high [5]. A composite transmission with hydraulic and machinery for VSCF wind power generation system was invented in [6], but the transmission is mainly conducted by mechanical components.

Recently a new wind turbine with hydraulic transmission is proposed. The principle diagram is illustrated in figure 1. The hydraulic pump was droved by the wind rotor, the high pressure hydraulic oil which was pushed out from the hydraulic pump drives the hydraulic motor. The hydraulic motor has variable displacement volume, and by which to change the rotor speed. Hydraulic motor and synchronous generator are linked with the coupling. The pressure of hydraulic oil becomes more stable for the configured energy storage devices. Hence it is easy to connect to the

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 581–588.
springerlink.com

electrical networks with synchronous generator, and the grid-connect problem will be solved. Expensive wind power converter and heavy gear box used in traditional wind turbine technology are eliminated in the system, which lead up to a lower costs, the system is characterized by relocating the majority of the critical components from the top nacelle in the top of the turbine down to a power unit at the ground foundation. It reduces the top weight significantly and simplifies service access. Therefore the advantage will be particularly obvious in offshore wind power generation system.

Since the synchronous generator was applied in this wind power system, so the constant rotor speed is required when connecting to grid (speed fluctuation less than± 1%). It is very difficult to control the hydraulic motor speed at a constant value in the condition of random wind speed and uncertainty parameters internal system. To solve the problem, the system dynamics model was built and an adaptive Backstepping controller was designed. The nesting problem caused by the mutual inclusion of parameter adaptive law and control input was solved effectively by using the improved Backstepping with selecting a proper Lyapunov function. The convergence was proved with Lyapunov stability theory. Steady-state tracking and rapid response capability were improved effectively. The condition of a constant rotor speed for connecting to the grid was fulfilled.

2 System Mathematical Model and Problem Description

The principle diagram of rotor speed control of hydraulic motor system is illustrated in figure 2. The displacement volume of hydraulic motor is controlled by valve-controlled cylinder drive system. The auxiliary pump provided the oil of control components. The high pressure hydraulic oil for hydraulic motor was in a constant pressure range as to the configured energy storage devices.

Hydraulic motor torque balance equation as follow

$$Pq_m = J\frac{d\omega}{dt} + B_t\omega + T_g \tag{1}$$

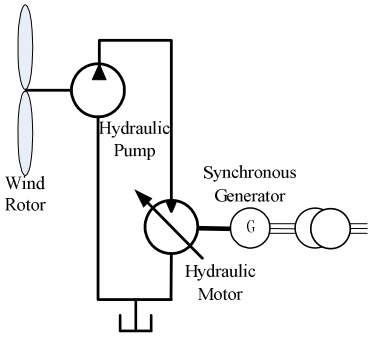


Fig. 1. Schematic diagram of wind turbine with hydraulic transmission

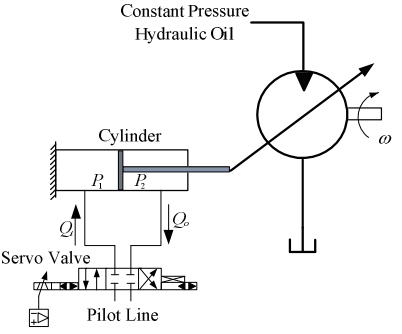


Fig. 2. Schematic diagram of rotor speed control of hydraulic motor

where P - constant pressure hydraulic oil, q_m - displacement volume of hydraulic motor, J - rotational inertia of the hydraulic motor, ω -angular velocity of hydraulic motor, B_t -viscous damper coefficient of hydraulic motor, T_g - friction torque of hydraulic motor and no-load generator.

The displacement volume of hydraulic motor is defined as follow

$$q_m = \frac{q_{m\max} - q_{m\min}}{x_{p\max}} x_p \tag{2}$$

where $q_{m\max}$ - maximal displacement volume of hydraulic motor, $q_{m\min}$ - minimum displacement volume of hydraulic motor, $x_{p\max}$ - maximum hydraulic cylinder piston, x_p - actual hydraulic cylinder piston.

Ignoring the nonlinear of electric-hydraulic servo valve, the servo amplifier and servo valve equivalent as proportion, the equation of servo valve can be written as

$$Q_L = K_{qsv} u - K_c P_L \tag{3}$$

where $Q_L = Q_i - Q_o$; K_{qsv} is the product of servo amplifier and servo valve gain, u - controller output, K_c - gain of control oil pressure, P_L - load pressure of cylinder, $P_L = P_1 - P_2$.

The equation of hydraulic cylinder is given by

$$Q_L = A_p \dot{x}_p + C_t P_L + \frac{V}{4\beta_e} \dot{P}_L \tag{4}$$

where A_p - effective area of hydraulic cylinder pistion, C_t - leakage coefficient of hydraulic cylinder, V - total volume of hydraulic control system, β_e - volume elastic modulus of oil.

The piston force balance equation of hydraulic cylinder is given by

$$A_p P_L = M\ddot{x}_p + B_p \dot{x}_p + K x_p + F_L \tag{5}$$

where M - total quality of hydraulic cylinder piston and load, B_p - viscous damping coefficient of hydraulic cylinder piston and load, K - motor swashplate reset spring elasticity coefficient, F_L - load force on the cylinder.

System states take as

$$[x_1 \ x_2 \ x_3 \ x_4]^T = [\omega \ x_p \ \dot{x}_p \ P_L]^T \tag{6}$$

The system state equations abstract from (1)-(6) are

$$\begin{cases} \dot{x}_1 = (1/\alpha_1)x_2 - T_m - \alpha_2 x_1 \\ \dot{x}_2 = x_3 \\ \dot{x}_3 = -\beta_1 x_2 - \beta_2 x_3 + (1/\beta_3)x_4 \\ \dot{x}_4 = -\gamma_1 x_3 - \gamma_2 x_4 + (1/b)u \\ y = x_1 \end{cases} \tag{7}$$

where

$\alpha_1 = J \cdot x_{p\max} / (P(q_{m\max} - q_{m\min}))$; $T_m = (Pq_{m\min} - Tg)/J$; $\alpha_2 = B_t / J$; $\beta_1 = K/M$;

$\beta_2 = B_p / M$; $\beta_3 = M/A_p$; $d = F_L / M$; $\gamma_1 = 4\beta_e A_p / V$; $\gamma_2 = 4\beta_e (K_c + C_t)/V$;

$b = V/(4\beta_e K_{qsv})$

where u - control input signal of the system; y - output of the system.

All the states of the system can be measured. As to the random wind speed and uncertainty internal system parameter, the constant oil pressure, viscous damper coefficient of Hydraulic motor and friction torque hydraulic motor and no-load generator may change uncertainly.

The disturbances descript as

$P = P_0 + \Delta P$, $T_g = T_0 + \Delta T$, $B_t = B_{t0} + \Delta B_t$.

The control objective is utilizing proper controller, and the output of the system is in the stable range (1500rpm±1%) which the grid-connecting required.

3 Design of Adaptive Backstepping Controller and Stability Analysis

3.1 Design of the Adaptive Backstepping Controller

Step 1: x_d is defined as the desired rotor speed, the error is $e_1 = x_1 - x_d$

So the derivation of e_1 is

$$\dot{e}_1 = \dot{x}_1 - \dot{x}_d = (1/\alpha_1)x_2 - T_m - \alpha_2 x_1 - \dot{x}_d .$$

Avoiding the nesting problem caused by the mutual inclusion of parameter adaptive law and control input, proper coefficients are added to e_1 and e_3 when selecting the Lyapunov function candidate for V_1, V_3 .

Select the first Lyapunov function candidate as $V_1 = \frac{1}{2}\alpha_1 e_1^2$.

The derivation of V_1 is

$$\dot{V}_1 = e_1(x_2 - \alpha_1 T_m - \alpha_1 \alpha_2 x_1 - \alpha_1 \dot{x}_d) \tag{8}$$

Eq.(8) can be rewritten as

$$\dot{V}_1 = e_1(x_2 - \alpha_3 - \alpha_4 x_1) \tag{9}$$

where $\alpha_3 = \alpha_1 T_m$; $\alpha_4 = \alpha_1 \alpha_2$

Take virtual controller

$$x_{2d} = \hat{\alpha}_3 + \hat{\alpha}_4 x_1 - K_1 e_1 \tag{10}$$

where $\hat{\alpha}_3$, $\hat{\alpha}_4$ are the estimated value of α_3, α_4.

Substituting Eq. (9) into Eq. (8) results in

$$\dot{V}_1 = e_1(-\tilde{\alpha}_3 - \tilde{\alpha}_4 x_1) - K_1 e_1^2 + e_1 e_2 \tag{11}$$

where $\tilde{\alpha}_3 = \alpha_3 - \hat{\alpha}_3$; $\tilde{\alpha}_4 = \alpha_4 - \hat{\alpha}_4$

Step 2: According to Eq. (10), the position error is $e_2 = x_2 - x_{2d}$.

The derivation of e_2 is $\dot{e}_2$, $\dot{e}_2 = \dot{x}_2 - \dot{x}_{2d} = x_3 - \dot{x}_{2d}$.

Select the second Lyapunov function candidate as

$$V_2 = V_1 + \frac{1}{2}e_2^2 + \frac{1}{2}\rho_1 \tilde{\alpha}_3^2 + \frac{1}{2}\rho_2 \tilde{\alpha}_4^2$$

The derivation of V_2 is

$$\dot{V}_2 = \dot{V}_1 + e_2\dot{e}_2 - \rho_1\tilde{\alpha}_3\dot{\hat{\alpha}}_3 - \rho_2\tilde{\alpha}_4\dot{\hat{\alpha}}_4 = -e_1\tilde{\alpha}_3 - e_1\tilde{\alpha}_4 x_1 - K_1 e_1^2 + e_2(e_1 + x_3 - \dot{x}_{2d}) - \rho_1\tilde{\alpha}_3\dot{\hat{\alpha}}_3 - \rho_2\tilde{\alpha}_4\dot{\hat{\alpha}}_4 \tag{12}$$

Take new virtual controller

$$x_{3d} = \dot{x}_{2d} - K_2 e_2 - e_1 \tag{13}$$

Then a new error is defined as $e_3 = x_3 - x_{3d}$.

Substituting Eq. (13) into Eq. (12) results in

$$\dot{V}_2 = -K_1 e_1^2 - K_2 e_2^2 + e_2 e_3 \tag{14}$$

Let the disturbances adaptive updating law be $\dot{\hat{\alpha}}_3 = -\frac{1}{\rho_1}e_1, \dot{\hat{\alpha}}_4 = -\frac{1}{\rho_2}e_1 x_1$.

Step 3: The derivation of e_3 is

$$\dot{e}_3 = \dot{x}_3 - \dot{x}_{3d} = -\beta_1 x_2 - \beta_2 x_3 + \frac{1}{\beta_3}x_4 - d - \dot{x}_{3d} \tag{15}$$

Select the third Lyapunov function candidate as $V_3 = V_2 + \frac{1}{2}\beta_3 e_3^2$.

The time derivative of V_3 along the solution of (14) and (15) is

$$\dot{V}_3 = \dot{V}_2 + \beta_3 e_3 \dot{e}_3 = -K_1 e_1^2 - K_2 e_2^2 + e_3(e_2 - \beta_1\beta_3 x_2 - \beta_2\beta_3 x_3 + x_4 - \beta_3 d - \beta_3 \dot{x}_{3d}) \tag{16}$$

Take the fourth virtual controller

$$x_{4d} = -e_2 + \beta_1\beta_3 x_2 + \beta_2\beta_3 x_3 + \beta_3 d + \beta_3 \dot{x}_{3d} - K_3 e_3 \tag{17}$$

Then the error is defined as $e_4 = x_4 - x_{4d}$.

Substituting Eq. (17) into Eq. (16) results in

$\dot{V}_3 = -K_1 e_1^2 - K_2 e_2^2 - K_3 e_3^2 + e_4 e_3$

Step 4: The derivation of e_4 is:

$$\dot{e}_4 = \dot{x}_4 - \dot{x}_{4d} = -\gamma_1 x_3 - \gamma_2 x_4 + (1/b)u - \dot{x}_{4d} \tag{18}$$

Select the fourth Lyapunov function candidate as: $V_4 = V_3 + \frac{1}{2} b e_4^2$.

Then the time derivative of V_4 along the solution of (16) and (18) is

$$\dot{V}_4 = -K_1 e_1^2 - K_2 e_2^2 - K_3 e_3^2 + e_4(e_3 - b\gamma_1 x_3 - b\gamma_2 x_4 + u - b\dot{x}_{4d}) \tag{19}$$

Choose the control u as

$$u = -e_3 + b\gamma_1 x_3 + b\gamma_2 x_4 + b\dot{x}_{4d} - K_4 e_4 \tag{20}$$

3.2 System Stability Analysis

First, proving the speed tracking error $e_1 \to 0$, substituting Eq.(20) into Eq.(19) results in

$$\dot{V}_4 = -K_1 e_1^2 - K_2 e_2^2 - K_3 e_3^2 - K_4 e_4^2 \le 0 \tag{21}$$

Then we obtain: $\dot{V}_4 \le -K_1 e_1^2$

$$\int_0^\infty K_1 e_1^2 dt \le V_4(e_1(0), e_2(0), e_3(0), e_4(0)) - V_4(e_1(t), e_2(t), e_3(t), e_4(t)) \tag{22}$$

As e_1, e_2, e_3, e_4, $\hat{\alpha}_3$ and $\hat{\alpha}_4$ are bounded, we can obtain V_4 is bounded , by the theorem of Barbalat[9] , we can obtain: $\lim_{t\to\infty} e_1 = 0$.

Similarly, $\lim_{t\to\infty} e_2 = 0$, $\lim_{t\to\infty} e_3 = 0$, $\lim_{t\to\infty} e_4 = 0$, therefore, the system is globally asymptotically stable.

4 Simulations

The main nominal parameter-values in the simulation as follow

$q_{m\max} = 1.05\times10^{-4}(\mathrm{m}^3/\mathrm{rad})$; $P_0 = 0.5\times10^6(\mathrm{Pa})$; $q_{m\min} = 3.1\times10^{-5}(\mathrm{m}^3/\mathrm{rad})$;

$x_{p\max} = 0.05(\mathrm{m})$; $K = 6230(\mathrm{N/m})$; $J = 2.26(\mathrm{kg\cdot m^2})$; $B_{t0} = 0.26(\mathrm{N\cdot m\cdot s/rad})$;

$T_0 = 10(\mathrm{N\cdot m})$; $K_{qsv} = 1.25\times10^{-5}$; $A_p = 1.05\times10^{-3}(\mathrm{m}^2)$; $V = 5.0\times10^{-4}(\mathrm{m}^3)$;

$B_p = 18000(\mathrm{N\cdot s/m})$; $M = 2.0(\mathrm{kg})$; $\beta_e = 7\times10^8(\mathrm{Pa})$.

The disturbances about P, B_t and T_g depict as follow

$$\Delta P = 0.1\times P_0 \times \sin(2t)); \ \Delta B_t = 0.05\times B_{t0} \times \sin(2t)); \ \Delta T = 0.05\times T_0 \times \sin(2t)).$$

According to the actual conditions, the initial speed before grid-connecting is different, Three initial speeds (0rpm, 300rpm, 600rpm) are taken into consideration. The obtained simulation results based on matlab technique are shown in Fig. 3~6,

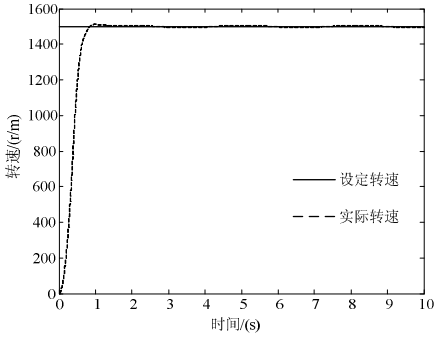


Fig. 3. Rotor speed response curve starting with zero

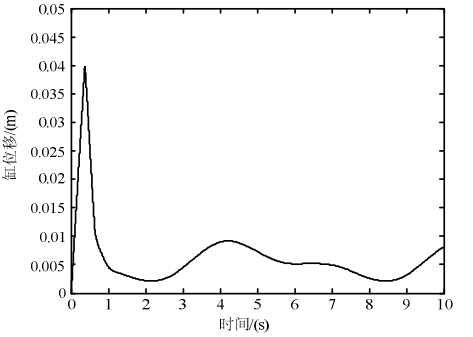


Fig. 4. Position of hydraulic cylinder response curve starting with zero

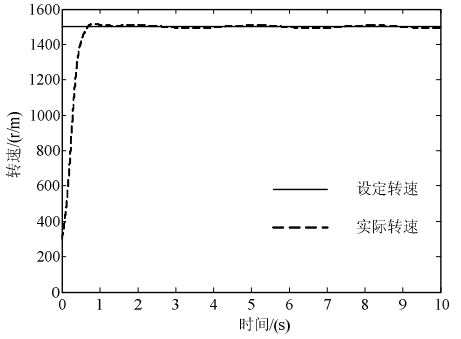


Fig. 5. Rotor speed response curve starting with 300rpm

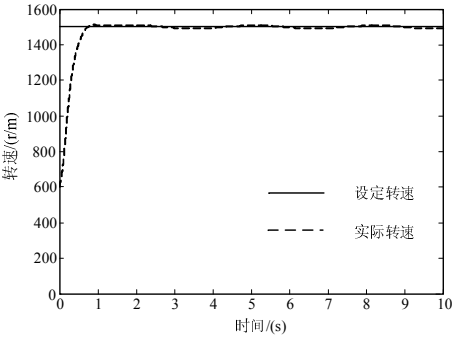


Fig. 6. Rotor speed response curve starting with 600rpm

It is illustrated from the above figures that the algorithm proposed in this paper has perfect control effects against certain disturbance and internal uncertain disturbances. the fluctuation range of speed is ±9rpm(less than the required ±15rpm), and the system can connect to the grid with different initial speed. Output of the controller is steeply, and the chattering of hydraulic cylinder and speed can be essentially avoided.

5 Conclusions

In this paper, the system dynamics model is built, and an adaptive Backstepping control approach is proposed for the rotor speed control of grid-connecting of the wind turbines with hydraulic transmission. The nesting problem caused by the mutual inclusion of parameter adaptive law and control input is solved effectively by using

the improved Backstepping with selecting a proper Lyapunov function. The system with algorithm proposed in this paper provides good dynamic response and strong robustness when external certain and internal uncertain disturbances appeared in the system. It also has perfect control effects on different startup conditions while connecting to the grid.

References

1. He, D., Liu, Y., Wang, Y.: Study of the Shunt-connected Wind Power Generation System. High Voltage Engineering 34(1), 142–147 (2008)
2. Tapia, A., Tapia, G., Ostolaza, J., et al.: Modeling and Control of a Wind Turbine Drive Doubly Fed Induction Generator. IEEE Transactions on Energy Conversion 18(2), 194–204 (2003)
3. Liu, Q., He, Y., Zhang, J.: Operation Control and Modeling-Simulation of AC-excited Variable-speed Constant-frequency (AEVSCF) Wind Power Generator. Proceedings of the Chinese Society for Electrical Engineering 26(5), 43–50 (2006)
4. Mei, F., Pal, B.: Model Analysis of Grid-connected Doubly Fed Induction Generatiors. IEEE Transactions on Energy Conversion 22(3), 728–736 (2007)
5. Zeng, X., Zhang, H.: Design of High-efficiency Converters for Large Direct-drive Wind Turbines. Proceedings of the Chinese Society for Electrical Engineering 30(30), 15–21 (2010)
6. Zhao, K., Liu, W.: A Variable Speed Pitch Regulated Wind Power Generator Based on Hydraulic Composite Transmissions. Chinese Hydraulics & Penumatics 3, 53–55 (2008)
7. Sun, L., Zhao, J.: A Novel Adaptive Backstepping Design of Turbine Main Steam Valve Control. Journal of Control Theory and Application 8(4), 425–428 (2010)
8. Karimi, A., Feliachi, A.: Decentralized Adaptive Backsteping Control of Electric Power Systems. Electric Power Systems Research 78(3), 484–493 (2008)
9. Slotine, J.J.E., Li, W.: Applied Nonlinear Control. Prentice Hall, Engle-wood Cliffs (1991)
10. Zheng, J., Feng, Y., Zheng, X.: Adaptive Backstepping-based Terminal-sliding-mode Control for Uncertain Nonlinear Systems. Journal of Control Theory and Application 26(4), 410–414 (2009)

Appendix: Authors Information

Hongbin Wang was born in 1966. He received Ph.D. degree in Control Theory and Control Engineering from Yanshan University in 2005. His research interest covers process automation, adaptive control, robot control technology, rubust control.

Wenzeng Gao was born in 1985. He is pursing her master degree in the Control Theory and Control Engineering at Yanshan University. His research interest covers adaptive control, process automation.

Study of the Influence of Delay Errors to Combining Performance in Antenna Arraying*

De-qing Kong and Xin-ying Zhu

National Astronomical Observatory, Chinese Academy of Sciences,
Beijing 100012, China
kdq@bao.ac.cn

Abstract. The estimation accuracy of antenna delay is one of the most important parameters for arraying combining performance in deep space network. The estimation of combining loss is presented, according to the probability distribution of delay errors. The impact of delay error to signal waveform and decode is analyzed. Theoretical analysis and simulation results show that the standard deviation of delay should be less than 0.09 times code width for uniform array, to ensure the combining loss caused by the delay error less than 0.1dB. The impact of delay errors to combined signal waveform can be equivalent to Gaussian filtering, if there are enough antennas. Delay errors can also cause intersymbol interference.

Keywords: Antenna arraying, Delay error, Signal combining, Deep space exploration.

1 Introduction

As the signals from the deep-space spacecrafts become weaker and weaker, the need arises to compensate for the reduction in signal-to-noise ratio (SNR) [1]. With maximum antenna apertures and lower receiver noise temperatures pushed to their limits, one effective method for improving the effective SNR is to combine the signals from several antennas. Arraying holds many tantalizing possibilities: better performance, increased operational robustness, implementation cost saving, more programmatic flexibility, and broader support to the science community [2, 3].

The output of an array is a weighted sum of the input signals applied to the combiner. The residual delay estimation accuracy and the phase difference between the signals has a direct impact on the combine performance, and the higher the code rate, the higher the required delay accuracy. With the development of deep space exploration, the demand for downlink code rate is growing rapidly. Currently, the maximum bit rate of the Deep Space Network can reach 20Mbps in Mars exploration (From the Earth 0.6Au), and may be up to 400Mbps (X-band) and 1.2Gbps (Ka-band) in 2020 [4]. Such a high bit rate requires high precision delay. It is necessary to study the influence of delay errors to combining performance in antenna arraying.

* Supported by the Natural Science Foundation of China(10903016).

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 589–595.
springerlink.com

The influence of phase difference estimation errors to combining performance is studied in the references 5 and 6, but the influence of delay errors is not analyzed. The delay errors can directly reduce the combining SNR and also affect the communication signal waveform, which affecting the signal demodulation performance, especially in the case of high bit rate. The estimation of combining loss is presented, according to the probability distribution of delay errors. The impact of delay error to signal waveform and decode is analyzed.

2 Full Spectrum Combining(FSC)[2]

In FSC, the intermediate frequency (IF) signals from each antenna are transmitted to the combining site, where they are combined. To ensure coherence, the signals must be delayed and phase adjusted prior to combining. An estimate of the correct delay and phase normally is accomplished by correlating the signal streams.

The primary advantage of FSC is that it can utilize the spectral characteristics of the signal source but does not crucially depend on them. FSC can be used when the carrier is too weak to track or is not possible to track with a single antenna. In this case, the gross relative delays and phases between antennas are determined a priori from geometry calculations. Then the residual relative delays and phases are determined by cross-correlation of the signals from each antenna. These delays and phases are used to correct the antenna IF signals, and then they are combined.

Obviously, the delay errors can directly reduce the SNR, and the required precision is a function of the bandwidth of the signal. We assume that the antennas have identical receivers, centered at a frequency f_0, and have bandwidth Δf. If there is an error in the compensation of the geometric delay, the signals will lose some coherence. The requirement for coherence over the band becomes

$$\Delta f \Delta \tau \ll 1 \tag{1}$$

where Δf is in cycles and $\Delta \tau$ is in seconds.

3 The Influence of Delay Errors to Combined Signal Power

Assuming the satellite downlink signal is narrowband Gaussian signal, the signal received by the i th antenna is

$$y_i(t) = s_i(t) + n_i(t) \tag{2}$$

where $s_i(t)$ is the received satellite signal, $n_i(t)$ is the noise. Assuming $s_i(t)$ is narrowband Gaussian white noise, and

$$s_i(t) = \alpha_i c(t-\tau_i) \mathrm{e}^{\mathrm{j}\left[2\pi f_0(t-\tau_i)+\phi_i(t-\tau_i)\right]} \tag{3}$$

where α_i is the amplitude of the i th antenna. We assume α_i is a slowly varying function of time, which can be considered constant in a short time. The carrier

frequency in this expression is represented by f_0, the phase of the i th antenna is represented by $\phi_i(t)$, and the delay of the signal received by the i th antenna relative to the phase center is represented by τ_i. $c(t)$ is the satellite downlink data code after the root raised cosine filtering, and the expression is

$$c(t) = h(t) * c'(t) \tag{4}$$

where $h(t)$ is the root raised cosine filter, $c'(t)$ is the code signal function, and the expression is

$$c'(t) = \sum_{n=-\infty}^{\infty} a(nT_c) q(t - nT_c - t_0) \tag{5}$$

where $q(t)$ is a rectangular window function, $a(nT_c)$ is binary sequence, and the expressions are

$$q(t) = \begin{cases} 1 & |t| \le T_c/2 \\ 0 & |t| > T_c/2 \end{cases}, \quad a(nT_c) = \pm 1 \tag{6}$$

where T_c is the signal code width, α is the filter roll-off factor.

For FSC method, using the SIMPLE algorithm [2], the weighted sum of the total L antennas is

$$y_\Sigma(t) = \sum_{i=1}^{L} \gamma_i y_i(t + \hat{\tau}_i) \tag{7}$$

where γ_i is the weight of the i th antenna, which used for phase compensation. For the convenience of analyzing the influence of delay errors to combining performance, assuming the phase difference is fully compensated by the weight, from Eq. (2) the combined signal becomes

$$y_\Sigma(t) = \sum_{i=1}^{L} \gamma_i \alpha_i c(t + \Delta\tau_i) + \sum_{i=1}^{L} \gamma_i n_i(t + \hat{\tau}_i) \tag{8}$$

where $\Delta\tau_i$ is the delay estimation error, that is $\Delta\tau_i = \hat{\tau}_i - \tau_i$; γ_i is the modulus of γ_i, that is $\gamma_i = |\gamma_i|$.

From Eq. (8), The total signal power at the output of the combiner conditioned on residual delays is given by

$$P_{s_\Sigma} = \left\{ \left[s_\Sigma(t) \middle| \Delta\tau_i \right] \right\}^2 = \sum_{i=1}^{L} \sum_{j=1}^{L} \gamma_i \gamma_j \alpha_i \alpha_j \overline{\left[R_{cij}(\Delta\tau_i, \Delta\tau_j) \middle| \Delta\tau_i, \Delta\tau_j \right]} \tag{9}$$

where $\overline{\left[R_{cij}(\Delta\tau_i,\Delta\tau_j)\middle|\Delta\tau_i,\Delta\tau_j\right]}$ is the statistical average of the cross-correlation function $R_{cij}(\Delta\tau_i,\Delta\tau_j)$ relative to the random variable $\Delta\tau_i$ and $\Delta\tau_j$, and the expressions of $s_\Sigma(t)$ and $R_{cij}(\Delta\tau_i,\Delta\tau_j)$ become

$$s_\Sigma(t)=\sum_{i=1}^{L}\gamma_i\alpha_i c(t+\Delta\tau_i) \tag{10}$$

$$R_{cij}(\Delta\tau_i,\Delta\tau_j)=\frac{1}{2T}\int_{-T}^{T}c(t+\Delta\tau_i)c(t+\Delta\tau_j)dt \tag{11}$$

Assuming the code $c(t)$ is Gaussian random process, and the delay estimate of any antenna is independent and Gaussian distributed with the standard deviation σ_{τ_i}, then the following expressions can be gotten that

$$R_{cij}(\Delta\tau_i,\Delta\tau_j)=R_{cij}(\Delta\tau_i-\Delta\tau_j) \tag{12}$$

$$\overline{\left[R_{cij}(\Delta\tau_i-\Delta\tau_j)\middle|\Delta\tau_i-\Delta\tau_j\right]}_{i\neq j}=\left(\frac{1}{2\pi\sigma_{ij}^2}\right)^{1/2}\int_{-\infty}^{\infty}R_{cij}(\gamma)\mathrm{e}^{\left(-\frac{\gamma^2}{2\sigma_{ij}^2}\right)}d\gamma \tag{13}$$

where $\sigma_{ij}=\sqrt{\sigma_{\tau_i}^2+\sigma_{\tau_j}^2}$.

From Eq. (4) and Eq. (11), the expression of R_{cij} becomes

$$R_{cij}(\tau)=T_c\mathbb{F}^{-1}\left[H(f)Sa^2(\pi fT_c)\right] \tag{14}$$

For ease of calculation, Eq. (14) is approximated by the following Gaussian function

$$\tilde{R}_{cij}(\tau)\approx A\mathrm{e}^{\left(-\tau^2/2\sigma_{T_c}^2\right)} \tag{15}$$

where the value of parameter σ_{T_c} can be obtained by fitting in the range of $\pm T_c$, which is calculated as $\sigma_{T_c}\approx 0.58T_c$, then from Eq. (13) the statistical average of the cross-correlation function becomes

$$\overline{\left[R_{cij}(\Delta\tau_i-\Delta\tau_j)\middle|\Delta\tau_i-\Delta\tau_j\right]}_{i\neq j}\approx\left(\frac{1}{2\pi\sigma_{ij}^2}\right)^{1/2}\int_{-\infty}^{\infty}\mathrm{e}^{\left(-\frac{x^2}{2\sigma_{ij}^2}-\frac{x^2}{2\sigma_{T_c}^2}\right)}dx=\frac{1}{\sqrt{k_{ij}^2+1}} \tag{16}$$

where $k_{ij}=\dfrac{\sigma_{ij}}{\sigma_{T_c}}$, from Eq. (9) the power of the combined signal becomes

$$P_{s_\Sigma}=\sum_{i=1}^{L}\gamma_i^2\alpha_i^2+\sum_{i=1}^{L}\sum_{j=1,j\neq i}^{L}\gamma_i\gamma_j\alpha_i\alpha_j\overline{R_{cij}}\approx\sum_{i=1}^{L}\gamma_i^2\alpha_i^2+\sum_{i=1}^{L}\sum_{j=1,j\neq 1}^{L}\gamma_i\gamma_j\alpha_i\alpha_j\Big/\sqrt{k_{ij}^2+1} \tag{17}$$

Assuming all the antennas are equal in size and performance, $\alpha_i = \alpha_0$ and $\gamma_i = 1$, then the power of the combined signal becomes

$$P_{s_\Sigma} = \alpha_0^2 \left[L + L(L-1) \Big/ \sqrt{k_0^2 + 1} \right] \tag{18}$$

Ideally, $k_0 = 0$ and the power of the combined signal is $L^2 \alpha_0^2$. Assuming the combining loss caused by the delay errors is

$$D_s = 10 \log \left(P_{s_\Sigma} \Big/ L^2 \alpha_0^2 \right) \tag{19}$$

From Eq. (18), the combining loss of uniform array becomes

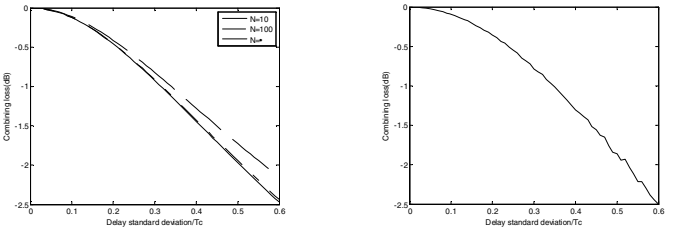
$$D_s = 10 \log \left[1/L + (1 - 1/L) \Big/ \sqrt{k_0^2 + 1} \right] \tag{20}$$

In particular, when the antenna number tends to infinity, the combining loss becomes

$$\lim_{L \to \infty} D_s = -5 \log \left(k_0^2 + 1 \right) \tag{21}$$

Fig. 1(a) shows the combining loss calculated by Eq. (21) in the case of different delay standard deviation, where the number of array antennas is respectively 10, 100 and infinity. It can be seen from Fig. 1(a) the number of antenna has some influence on the combining loss, the greater the number of antennas the greater the combining loss, and eventually approaches to the solid line showed in Fig. 1(a). Calculation results show that the standard deviation of delay should be less than 0.09 times code width(T_c), to ensure the combining loss caused by the delay error less than 0.1dB.

In order to verify the correctness of the above derivation process, the simulation of the relationship between combining loss and delay standard deviation is presented using Monte Carlo method. The simulation experiment is repeated 100 times, and the results are showed in Fig. 1(b). It can be seen that the simulation results in Fig. 1(b) are basically consistent with the theoretical results in Fig. 1(a).



(a) (b)

Fig. 1. Combining loss. (a) Results computed by Eq. (21), (b) Simulation results (100 antennas)

4 The Impact of Delay Error to Signal Waveform

Assuming the array is composed by a large number of antennas, which are all equal in size and performance, from Eq. (8) the statistical average on delay errors of combined signal can be expressed as

$$\begin{aligned}\mu_{\Sigma}(t) &= E\left[s_{\Sigma}(t,\Delta\tau_1,\Delta\tau_2,\ldots,\Delta\tau_L)\right] \\ &= \sum_{i=1}^{L}\gamma_i\alpha_i\left(\frac{1}{2\pi\sigma_{\tau_i}^2}\right)^{1/2}\int_{-\infty}^{\infty}c(t+\Delta\tau_i)\mathrm{e}^{\left(-\frac{\Delta\tau_i^2}{2\sigma_{\tau_i}^2}\right)}d\Delta\tau_i = \sum_{i=1}^{L}\gamma_i\alpha_i\left(h*c'*g_i\right)(t)\end{aligned} \tag{22}$$

where $g_i(t)$ is Gaussian function with the standard deviation σ_{τ_i}, namely

$$g_i(t) = \left(\frac{1}{2\pi\sigma_{\tau_i}^2}\right)^{1/2}\int_{-\infty}^{\infty}\mathrm{e}^{\left(-\frac{\gamma^2}{2\sigma_{\tau_i}^2}\right)}d\gamma \tag{23}$$

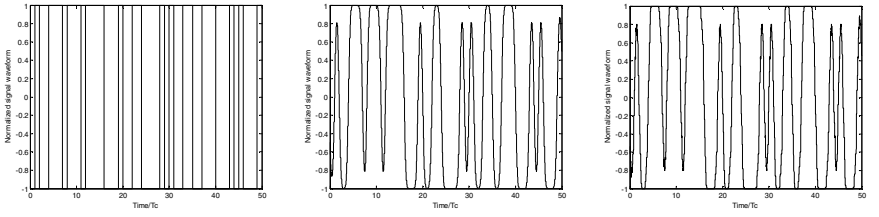
When the array is uniform, from Eq. (22) the following expression can be obtained as

$$\mu_{\Sigma}(t) = L\left(h*c'*g\right)(t) \tag{24}$$

namely the signal waveform is equivalent to

$$\tilde{c}(t) = \left(c'*g\right)(t) \tag{25}$$

It can be seen from Eq. (25) that the combined signal waveform is equivalent to filtered by a Gaussian filter which the parameter of standard deviation is σ_{τ}, due to the impact of delay errors.



(a) (b) (c)

Fig. 2. The impact of delay errors to combined signal waveform. (a) Original signal waveform, (b) Combined signal waveform calculated by Eq. (25), (c) Combined signal waveform of simulation results (100 antennas).

In order to analyze the impact of delay errors to combined signal waveform, calculation and simulation results are performed, and the results are shown in Fig. 2. The waveform of original signal which is randomly selected is shown in Fig. 2(a), the

combined signal waveform calculated by Eq. (25) is shown in Fig. 2(b), and the combined signal waveform of simulation results is shown in Fig. 2(c). The delay standard deviation is taken as σ_{τ_i} in Fig. 2(b) and (c), and the number of antennas is 100 in Fig. 2(c). It can be seen the waveform of simulation are basically consistent with the theoretical results, and the delay errors will cause waveform distortion. Because the Gaussian filter doesn't meet the Nyquist criteria, the delay errors can also cause intersymbol interference.

5 Conclusions

For FSC combining method and SIMPLE algorithm, the influence of delay errors to combining performance in antenna arraying have been studied. The estimation of combining loss has been presented, according to the probability distribution of delay errors. The impact of delay error to signal waveform and decode has been analyzed.

Theoretical analysis and simulation results show that the standard deviation of delay should be less than 0.09 times code width for uniform array, to ensure the combining loss caused by the delay error less than 0.1dB. Delay errors have an impact on the combined signal waveform. When the array is composed by a large number of antennas, the impact of delay errors to combined signal waveform can be equivalent to Gaussian filtering. Because the Gaussian filter doesn't meet the Nyquist criteria, delay errors can cause intersymbol interference.

In the future, the influence of delay errors to deep space ranging in antenna arraying should be studies.

References

1. Zhang, N.-T., Li, H., Zhang, Q.-Y.: Thought and Developing Trend in Deep Space Exploration and Communication. Journal of Astronautics 28(4), 787–793 (2007) (in Chinese)
2. Rogstad, D.H., Mileant, A., Pham, T.T.: Antenna Arraying Techniques in the Deep Space Network. John Wiley and Sons, Hoboken (2003)
3. Li, H.-T., Li, Y.-H., Kuang, N.-X.: Antenna Array Forming Technology in Deep Space Exploration. Journal of Spacecraft TT & C Technology 23(4), 57–60 (2004) (in Chinese)
4. Cesarone, R.J., Abraham, D.S., Shambayati, S., Rush, J.: Deep-space optical communications. In: IEEE Space Optical Systems and Applications (ICSOS), Santa Monica, CA, May 11-13 (2011)
5. Rogstad, D.H.: The SUMPLE Algorithm for Aligning Arrays of Receiving Radio Antennas: Coherence Achieved with Less Hardware and Lower Combining Los. Inter Planetary Network (IPN) Progress Report 42-162, Jet Propulsion Laboratory, Pasadena, California (August 15, 2005)
6. Kong, D.-Q., Shi, H.-L.: The Study of Weights Performance and Combining Loss of SUMPLE Algorithm for Non-uniform Antenna Arraying. Journal of Astronautics 30(5), 1941–1946 (2009) (in Chinese)

Research on the Matching and Optimizing Technology in the Combustion Process of Low Speed Diesel Engines*

Chen Guojin[1], Liu Zhongmin[1], Liu Tingting[1], Yuan Guangjie[1], Su Shaohui[1], and Cao Yijiang[2]

[1] Zhejiang Provincial Key Laboratory of Ship and Port Machinery Equipment Technology, Hangzhou, Zhejiang, 310018
chenguojin@163.com
[2] Zhongji Hitachi Zosen Diesel Engine Co., Ltd., Zhoushan, Zhejiang, 316012

Abstract. Regarding the power, the efficiency and the emission of the low speed diesel engine, this paper has studied the matching and optimizing technique in the diesel engine combustion process emphatically, proposed the electrically controlled fuel injecting system with the high-pressured common rail as well as the humid air motor method and the system plan. The practice proved that the fuel injecting system with the high-pressured common rail and the humid air motor system can improve the combustion process most efficiently, thus enhance the diesel engine's performance, and reduce the pollutant discharge.

Keywords: Matching Technology, Optimizing Technology, Combustion Process, Low Speed Diesel Engine.

1 Introduction

The low-speed two-stroke marine diesel engine is composed by the fixed part, the moving part, the distribution system, the fuel system, the lubrication system, the cooling system, the starting system, the speed regulation device, the reversing arrangement, and the supercharging system. As a result of its good power performance, the fuel efficiency and the reliability, the modern merchant ship takes generally it to directly drive the propeller as a main propulsion system. Its performance has decided the entire ship's power performance and economical performance.

Internationally the low speed diesel engine's brand is mainly MAN B&W and Wartsila. The MAN's intellectualized ME series of the low speed diesel engines, can control the air starting valve, control the starting, advancing and reversing process, replace the speed governor, and implement the optimized control to the low speed auxiliary air compressor, the feed timing and rule, the exhaust gate and the spraying rate of the lubricating oil on the cylinder wall and so on. They may achieve a lower oil consumption rate and the performance optimization in the entire operating mode

* This work was supported by the National Science and Technology Ministry's projects (No.2010DFB70750 and No.2009GJ20029) and the Zhejiang provincial science and technology project (No.2009C01018).

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 597–602.
springerlink.com

and can satisfy the various requirements. For example, the optimized formation of the injection pressure and rule, the lower NOx discharge and the smokeless characteristic, the operating pattern transferred easily to a low oil consumption or a low NOx, the mechanism and system's simplification, the heat load and balance control in the various cylinders, the lowest stable speed, the best accelerating, advancing and reversing characteristic and the emergency stop performance, the optimized control to the cylinder wall lubrication, as well as easy adaptation to the operating characteristic optimization of various ships. Wartsila Corporation expanded the RT-flex serial products with the electrically controlled fuel injecting system of the high-pressured common rail on the RTA series' foundation, and on the applying foundation in the small cylinder diameter RT-flex 50 and RT-flex60C, developed the larger cylinder diameter RT-flex84T-D and RTflex96C with the electrically controlled fuel injecting system of the high-pressured common rail Ⅳ.

2 Optimization of Diesel Engine's Combustion Process

In the low speed diesel engine design, the combustion process optimization mainly solves the carbon smoke and the NOx emission. Regarding the high-power marine diesel engine's emission, what most mainly must be solved is the carbon smoke discharge in the slow-speed and the low operating mode. Firstly, in the entire operating mode, the black smoke that the naked eyes can see should not be produced including the starting mode. The emission value of the marine diesel engine is easy to be satisfied according to the IMO's NOx stipulation. Figure 1 is a group of curves. They are obtained on the experimental contrast in the Wartsila46 medium-speed engine. As can be seen from the figure, when the electrically controlled fuel injecting system with the high-pressured common rail is used, in the entire operating condition, the smoke is below FSN0.15. That is, the smoke color does not be seen, and is below the guarantee FSN0.4. But with the conventional oil supply system, for the entire operating mode the smoke surpasses the guarantee value in the load below 45%. But the biggest value is FSN0.9, also not too high. If the supplement air measure is taken in the low operating mode, the FSN value also possibly is below the guarantee value. If the requested NOx is below 70% of the IMO standard value, the FSN value has surpassed the guarantee value in the load below 60%.

The high-power low-speed diesel engine uses the scheme of the electrically controlled fuel injecting system with the high-pressured common rail, as shown in Figure 2. The two-cylinder common rail made the high-power low-speed diesel engine small in structure against the engine with the high-pressured oil track. The high pressure oil circuit controlled by a high speed magnetic valve can avoid the low operational reliability which governing directly a fuel injector needle brings.

In the electrically controlled fuel injecting system with the high-pressured common rail, the high-pressured oil feed pump provides the pressured oil to the common rail, but the injection discharge and injection timing is controlled by the high speed magnetic valve in the fuel injector. The electronic control unit (ECU) of the fuel injection system acts according to the operator's order and the operating condition of the diesel engine, and controls the injection start point and the injection end point as well as the injection times by controlling the fuel pressure and the magnetic valve's operating time in the common rail. Thus it can be seen, the electrically controlled fuel

injecting system with the high-pressured common rail may control independently the system injection pressure, the injection timing, the injection discharge and the injection rule. In comparison with the mechanical type fuel injection system, its superiority lies in:

2.1 Realizing the High-Pressure Injection

One of the most remarkable characteristics for the fuel injecting system with the high-pressured common rail is to realize the high-pressure injection. The high injection pressure may guarantee the good fuel atomization, thus can reduce the PM, CO, and HC emission. The injection pressure of the mechanical type fuel injection system is adjusted through the speed and the injection discharge. Because the mechanical type cannot guarantee the high-pressure injection, it causes easily the bad atomization, thus makes the PM, CO, and HC emission to increase. For example the diesel engine produces easily two times of injection under the high speed and the high load. The fuel oil in the two times of injection comes into the cylinder by the very low pressure, and causes the fume discharge to increase.

2.2 Guaranteeing the Accurate Injection Timing

To a great extent the injection timing is to affect the pollutant emission. Postponing the injection can cause the NOx emission to reduce, but possibly causes the CO and PM emission to increase. The injection timing also is very sensitive to the HC emission, and has the best value. The injection timing above or below the best value will possibly increase the HC emission. Therefore, choosing an accurate value of the injection timing is very important. It may enable the emission to achieve a best compromised value. However, the mechanical type fuel injection system adjusts the injection timing by the injection pump and the timing advance unit. Its setting range and the precision has been restricted. In the fuel injecting system with the high-pressured common rail, the injection timing is controlled by the ECU and the fuel injector. According to the electronic control signal from the ECU, the fuel injector takes the high-pressured fuel oil in the common rail to inject into the engine's combustion chamber by the best injection timing, the injection discharge, the injection rate and the atomization condition.

2.3 Implementing the Optimized Control of the Fuel Injection Rate

In the fuel injecting system with the high-pressured common rail, the injection start point and the duration time are decided by the command pulse, and they have nothing to do with the speed and the load. Therefore the injection time may be controlled freely. In addition the injection pressure may be controlled flexibly. So the ideal premix combustion can be formed. The pre-injecting start point, the main injecting start point and the injecting end point may be adjusted freely. The pressure in the pre-injecting section may be selected freely. The pressure rise rate after the main injecting start is suitable. The oil can be broken fast. The peak pressure of the injection may be chosen flexibly. The full combustion of the fuel oil may reduce the CO, HC, and PM emission. But the delay period reduced as far as possibly can lower the NOx emission.

2.4 Setting the Parameter Freely and Flexibly

In the fuel injecting system with the high-pressured common rail, the Injection nozzle parameter and the other various parameters may be set freely. Each parameter may satisfy the corresponding request independently. The aspect in satisfying the request is not restrained, the flexibility is good. These are the essential conditions to atomize the fuel oil good and reduce the emission.

Matching reasonably the high-pressure fuel injection system and controlling the parameters of the fuel injection pressure and the injection time, are important to improve the internal combustion process, optimize the diesel engine's power, efficiency and emission.

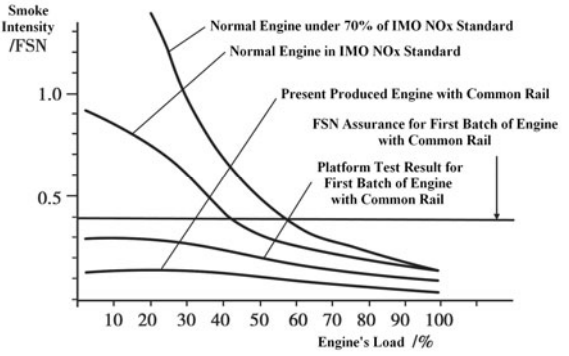


Fig. 1. The influence of the electrically controlled high-pressured common rail to the smoke intensity

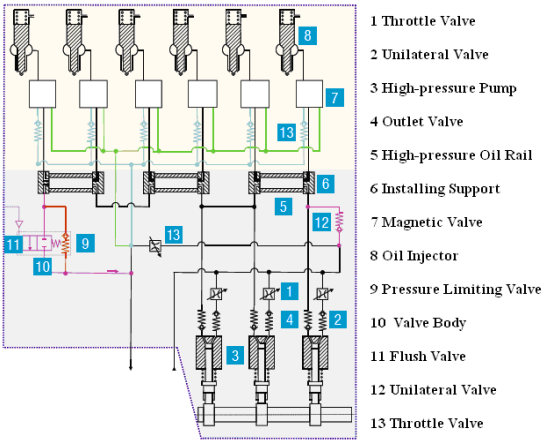


Fig. 2. The scheme of the electrically controlled fuel injecting system with the high-pressured common rail

3 Humid Air Motor

In order to adapt to the more stricter request compared to the existing IMO's emission laws and regulations, as well as the local codes and regulations like Sweden and Norway, besides using the electrically controlled fuel injecting system, in the low-speed engine the other measures also considered. For example, MAN B&W uses the EGR method and the humid air motor (HAM) method. The water-in-oil emulsion is used in the 5 MAN B&W low-speed two-stroke diesel engines for the power plant. If water is sprayed in the air cylinders directly, the more refitment must be made in the structure, moreover the massive fresh water must be used. But with the HAM system, the water is injected into the high temperature air after the supercharger, and it can also play the cooling role. If the air temperature after cooling is 70℃ about, the HAM system may replace completely the intercooler.

With the HAM system, the diesel engine itself does not need to refit, moreover may use the sea water. Therefore it is a good choice. The HAM system was once experimented on the 4T50ME-X testing engine. The NOx discharge drops by 25%~50%, the oil consumption rate increases by 1.0%~1.5%, the CO discharge increases some, the PM and HC discharge changes a little. Using the EGR in the two-stroke diesel engine may realize the cylinder's EGR through adjusting the exhaust timing. Moreover, MAN B&W uses the high-pressured EGR system, namely an air compressor presses the exhaust gas into a sea water bathing pool to purify and cool the exhaust gas. The experiment proved that the NOx discharge may drop by 30%~50%. The EGR and the HAM system further experimented in 2004 on the solid ship. Wartsila has also done the corresponding experimental study. The method combining the direct water injection (DWI) in the cylinder with the EGR in the cylinder is called WaCoReG (Water Cooled Residual Gas). Testing in the 4RT-flex58T-B using the above method, may make the NOx discharge lower than 50% of the limitation value by the IMO laws and regulations. In order to further improve the cylinder wall lubrication, lengthen the time between overhauls (TBO) and cut the lubricating oil consumption rate, Wartsila uses the anti-polishing rings in the two-stroke low-speed engine.

4 Conclusion

Matching and optimizing the low-speed diesel engine's combustion process has decided the engine's power, efficiency and emission. But the combustion process's optimization and match has more to do with the fuel injection rule, the intake swirl control, the Valve timing adjusting, the combustion chamber structure, and the operating condition. This article proposed the electrically controlled fuel injecting system with the high-pressured common rail as well as the humid air motor method and the system plan. They can improve the low-speed diesel engine's combustion process, thus enhance the diesel engine's performance, and reduce the pollutant discharge.

References

1. Zhang, X.-W., Man, C.-Z., Tang, Y.-B., Zhao, Y.-C.: Overseas Development of Electronic Controlled Common Rail Fuel Injection System for Low Speed Marine Diesel Engines. Internal Combustion Engines (5), 1–2 (2010)
2. Zeng, H., Zhang, J.-D., Lin, Y.-J., Gan, H.-B.: Performance Prediction of Marine Low-speed Diesel Engine under Variable Operating Condition. Navigation of China 32, 32–35 (2009)
3. Zhang, X.-G.: Study on the Effects of the Injection Strategy on the Combustion and Emission of a Diesel with Electrical Control System. M.S. thesis, Dalian Univ. of Technology, Dalian, China (2010)
4. Liu, Q.-A.: Numerical Simulation and Optimization of Working Process for Large-Scale Low-Speed Two-Stroke Marine Diesel Engine. M.S. thesis, Dalian Maritime Univ., Dalian, China (2009)

Modeling of Virtual Assembly and Disassembly Process Using Colored Petri Nets

Wu Di-xiao[1], Xu Xing-hua[2], and Shang Jie[2]

[1] Department of Weaponry Engineering,
Naval University of Engineering, Wuhan, Hubei, China
[2] Department of Computer, Naval University of Engineering,
Wuhan, Hubei, China
wdx201101@sina.co, wtosj@sina.com

Abstract. To ensure the validity in virtual assembly/disassembly training, the assembly and disassembly process should not be restricted to a sole static procedure, and the training progress should be carried out freely according to the operations of a trainee. This paper presents a novel approach to construct the process model by using colored Petri nets. The model can represent flexible operation sequences, and include tool limitations. The modeling is validated with an example.

Keywords: Virtual assembly/disassembly training, colored Petri net, flexible operation sequences.

1 Introduction

Virtual assembly and disassembly training (VADT) system, which uses the virtual reality and computer simulation technology, provides the virtual environment of training for the person who use and maintenance of the facilities. It can help the trainee fully understand the construction of the facilities by virtual assembly and disassembly operation, and get fully control of operation method and maintenance technology. In order to ensure the real training effect, virtual assembly and disassembly training process should have the sense of reality and be effective, which need not only vivid romance effect and good interactive method, but also a reasonable virtual assembly and disassemble process model in order to describe the logical sequence of assembly and disassemble completely and clearly.

There are many effective methods of assembly and disassemble process modeling, such as, linear sequences, tree structure[1]~[3], directed graph[4]~[6],, AND/OR graph, Ontology Language and Petri net and so on. Those method describe the assembly and disassemble process from different angles, and Some of them can describe many possible assembly and disassemble process of the same one facility.

Petri net is a kind of model which can describe of vary time and sequence among different states in system, and it is a general tool for describing the behavior of simultaneous discrete events.

In this paper, a modeling of flexible assembly/ disassembly processes with colored Petri nets is proposed, which focuses on the following issues:

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 603–609.
springerlink.com

--The trainee should be allowed to choose assembly or disassembly operations on any possible part in every step.

--The requirements of tool in the process of assembly or disassembly are under the consideration, and so are the restriction conditions of tool resources.

--The model is very flexible, in which all the possible operations and variation in assembly or disassembly process are considered.

2 Colored Petri Nets Model of Assembly and Disassembly Process

In this section, assembly and disassembly processes modeling are represented by using CPN. The part and tools are generalized to places.

2.1 Definition of Colored Petri Net

Jensen defined a non-hierarchical a CPN model as a nine tuple: $CPN = (\Sigma, P, T, A, N, C, G, E, IN)$. Σ is a finite set comprising finite and non-empty color sets. P is a finite set of places, and T is a finite set of transitions. A is a finite set of arcs such that: $P \bigcap T = P \bigcap A = T \bigcap A = \Phi$. N: $A \to P \times T \cup T \times P$ is a node function. C: $P \to \Sigma$ is a color function. G: $T \to Expressions$ is a guard function such that: $\forall t \in T : [Type(G(t)) = Bool \wedge Type(Var(G(t))) \subseteq \Sigma]$. E: $A \to Expressions$ is an arc expression such that $\forall a \in A : [Type(E(a)) = C(p(a))_{MS} \wedge Type(Var(E(a))) \subseteq \Sigma$. IN: $P \to Expressions$ is an initialization function such that: $\forall p \in P : [Type(IN(p)) = C(p)_{MS} \wedge Var(IN(p)) = \Phi]$.

In the above equation, $Type(v)$ denotes the type of the variable v. $Type(expr)$ denotes the type of the expression $expr$. $Var(expr)$ denotes the set of variables in an expression $expr$. $p(a)$ denotes the place of $N(a)$

2.2 Modeling of Assembly and Disassembly Process Using Colored Petri Nets

The specific modeling method of each part in this nine tuple is given as following:

Place set *P*. The set of places is described as $P = P_{part} \cup p_{tool}$, and $P_{part} \cap p_{tool} = \Phi$. P_{part} represents the set of the equipment parts, and P_{tool} represents the set of the tools.

color set Σ and mapping C. The set of color is described as $\forall p \in P_{part}$: $C(p) = \{Assembled, Disassembled\}$, $\forall p \in P_{tool}$: $C(p) = \{Available, Unavailable\}$. The set Σ is described as { $\{Assembled, Disassembled\}$, $\{Available, Unavailable\}$ }.

Transitions set *T*. The operation of assembled or disassembled are belongs to two categories: assembled operation and disassembled operation. The set of transitions is described as $T = T_{assemble} \cup T_{disassemble}$. $T_{assemble}$ represents the set of the assembly

operations. $T_{disassemble}$ represents the set of the disassembly operations, and $T_{assemble} \cap T_{disassemble} = \Phi$.

Directed arc set *A* and mapping *N*. Each transition should have corresponding a four tuple: $< type, p_{target}, P_{dependence}, P_{tool} >$, in which $type$ denotes the type of transition, namely disassemble or assemble. p_{target} denotes the objects of disassemble or assemble, namely the corresponding place. $P_{dependence}$ denotes the set of other part place which is dependent on during the assemble and disassemble process. P_{tool} denotes the set of the tool place needed during the assemble and disassemble process.

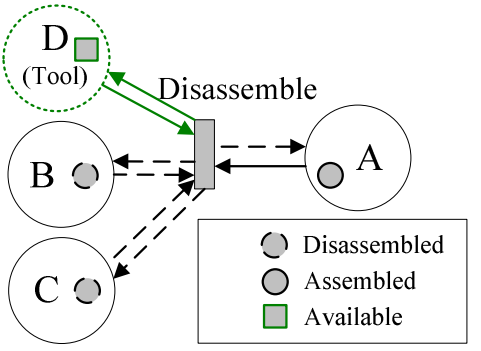


Fig. 1. Petri net diagram of assemble and disassemble process

Each transition has a corresponding four tuple, according which the set A can be constructed. The concrete steps is shown as follows:

Step 1: Traverse set *T*, and enter into the processing circle for each transition t_i (step 2 to step 5).

Step2: Construct a four tuple: $Op_i = < type_i, p_{target\_i}, P_{dependence\_i}, P_{tool\_i} >$ according to the description of transition t_i.

Step 3: Add arc a_i and a_i' to *A, and add the* enum mapping relation $N(a_i) = (p_{target\_i}, t_i)$, $N(a_i') = (t_i, p_{target\_i})$ to *N*.

Step 4: Traverse the set $P_{dependence\_i}$ *in the four tuple* Op_i; If $p_{ij} \in P_{dependence\_i}$, *then* add arc a_{ij} and a_{ij}' to *A, and add the* enum mapping relation $N(a_{ij}) = (p_{ij} \bullet t_i)$, $N(a_i') = (t_i, p_{ij})$ to *N*.

Step 5: Traverse the set P_{tool_i} *in the four tuple* Op_i;•If $p_{tool\_ik} \in P_{tool\_i}$, *then* add arc a_{tool_ik} and a_{tool_ik}' to *A*, and add the enum mapping relation $(a_{tool\_ik}) = (p_{tool\_ik}, t_i)$, $N(a_{tool\_ik}') = (t_i, p_{tool\_ik})$ to *N*.

Step 6: Return to step 1 and begin a new cycle.

After the above 5 steps, directed arc set *A* and mapping *N* can be gotton.

Mapping *E.* During the Traversal of the directed arc set *A* in turn, each dircted arc a_i can be treated under two considerations: (1) If $p(a_i) \in P_{part}$, then add enum mapping relation $E(a_i)$=$PState_i$ to *E,* in which $PState_i$ is a variable to describe the state of part and its definition domain is $\{Assembled, Disassembled\}$; (2) If $p(a_i) \in P_{tool}$,then add enum mapping relation $E(a_i)$=$TState_i$ in which $TState_i$ is a variable to describe the state of tool and its definition domain is $\{Available, Unavailable\}$.

Mapping *G.* If $type_i$ is *Disassemble* , then add enum mapping relation $G(t_i)$=$(p_{target\_i}$=$Assembled) \wedge (\forall p \in P_{dependence\_i}, AT(p) = true) \wedge (\forall p \in P_{tool\_i}, VT(p) = true)$ to *G;* otherwise, namely when $type_i$ is *Assemble,* add enum mapping relation $G(t_i)$=$(p_{target\_i}$=$Disassembled) \wedge (\forall p \in P_{dependence\_i}, DT(p) = true) \wedge (\forall p \in P_{tool\_i}, VT(p) = true)$ to *G*.

Mapping *IN* of inertial state of description place. Firstly, during the Traversal of each place p_i in P_{part}, the enum item can be added to mapping *IN* according the inertial state of p_i. If the inertial state is assemble, then add $IN(p_i)$= *Assembled* ; If the inertial state is disassemble, then add $IN(p_i)$= *Disassembled* . Secondly, during the Traversal of each place p_{tool_j} in P_{part}, add $IN(p_{tool\_j})$= *Available* .

2.3 Virtual Assemble and Disassemble of Petri Net Control

As the model used in this paper is an interactive virtual assemble and disassemble of Petri net control, and the operation of customer is discrete and uncertain, and the transition of state condition happens during the operation of customer, so the fire time should be controlled by system. The fire sequence in the paper is as following:

Step 1: Wait the input of customer by blocking. If there is an input, then enter step 2.

Step 2: Obtain the operation parts of customer and the operation types (assemble or disassemble), and get the related transition *t.* XX

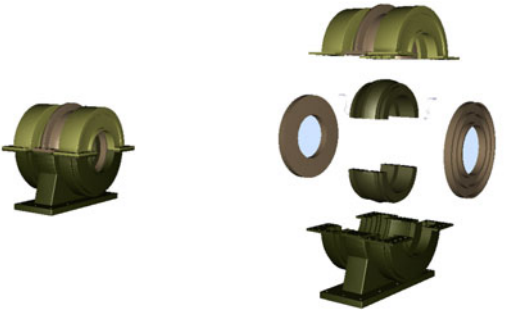
Step 3: Judge if *G*(*t*) is true, namely if the variables related to some places satisfy the fire condition of *t*. If it is true, then enter into step 4, or jump to step 5.

Step 4: When the transition *t* is on fire, and the present part is disassembled in virtual environment, return to step 1.

Step 5: remain the states of Petri net, and present the reason why the fire can not start (such as: should firstly disassemble some parts, and so on), and return to step 1.

3 An Instance

A set of prototype system is developed by the programming language C++, and the graphic library OpenGL 2.0. In order to testify the algorithm, a test was done on the 3-D model of some facility(See Fig. 2) described by VRML 2.0. This facility includes 10 parts, such as: the oil receiver A and B, oil ring A and B, Shell, lid(cap), cover A and B, upper bearing, and lower bearing. The assemble and disassemble relation and inertial states description of each part is shown in table1.



a view point of assemble **b** view point of explore

Fig. 2. Instance of process model of virtual assemble and disassemble

Table 1. Assemble and disassemble relation and inertial states description of each part

part	dependent relation of assemble and disassemble	dependent relation of assemble	necessary tool	inertial states
oil receiver A	Oil ring A	upper bearing	no	installed
oil receiver B	Oil ring B	lower bearing	no	installed
Oil ring A	cover A, Shell	oil receiver A, lower bearing	no	installed
Oil ring B	cover B, Shell	oil receiver B, lower bearing	no	installed
Shell	no	Oil ring A, Oil ring B	Tool A	installed
lid(cap)	no	upper bearing	Tool A	installed
cover A	no	Oil ring A	Tool A	installed
cover B	no	Oil ring B	Tool A	installed
upper bearing	oil receiver A, oil receiver B, lid(cap)	no	Tool B	installed
lower bearing	Oil ring A, Oil ring B	no	Tool B	installed

In table 1, dependent relation of assemble and disassemble denotes that some other parts should be disassembled at first before this part is disassembled; dependent relation of assemble denotes that some other parts should be assembled at first before this part cab be assembled. Necessary tool refers to the necessary tool when this part is disassembled; inertial states refer to the default assemble state of the part.

A virtual assembly/disassembly model described by a colored Petri net (see Fig. 3) is developed by the above mentioned method, among which there are 12places(2 of them are tool place) and 20 transitions (10 of them are disassemble, 10 of them are assemble). And all the parts are assembled and all the tool are available by default set.

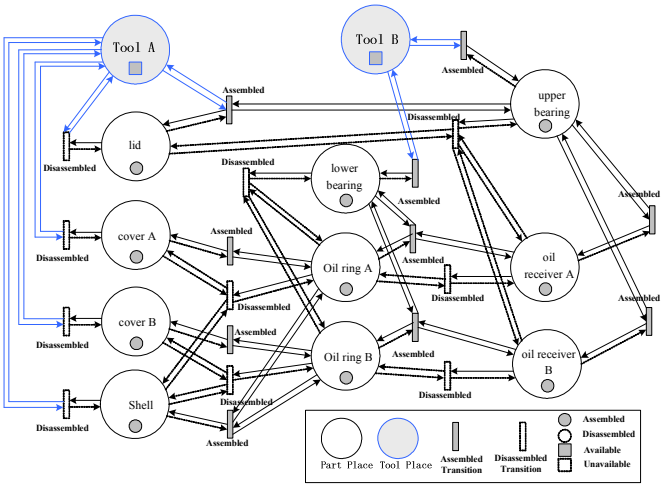


Fig. 3. Petri net model of virtual assemble and disassemble process of some a facility

As shown in fig.3, all the states of assemble and disassemble of this facility is given, and they can change freely by customers. During the maintenance training, the assemble and disassemble are unified to one process, and the model can flexible describe the assemble and disassemble process of local several parts, and can give hints and operation suggestions to those operations that can not carried out because the restrict can not be satisfied.

4 Summary

This paper presents a novel approach for the modeling of virtual assemble and disassemble process. In this model, the dependent relation of virtual assemble and disassemble is transformed to colored Petri net model, and the fire time of transition is decided by the customer's operation. It includes all the possible conditions which conform to the rule description, and can generate the assemble and disassemble sequence flexible. An instance testifies the feasibility of this approach.

References

1. Chen, X., Xu, N., Li, Y., Virtual Environment, A.: for Collaborative Assembly. Proceedings of the Second International Conference on Embedded Software and Systems on pp. 441
2. Yuan, X., Yang, S.X.: Intelligent Interface Design in Virtual Assembly. In: Proceedings of 2001 IEEE International Symposium on Computational Intelligence in Robotics and Automation, Canada, pp. 131–136 (2001)
3. Mok, S.M., Wu, C.-H., Lee, D.T.: A Hierarchical Workcell Model for Intelligent Assembly and Disassembly. In: Computational Intelligence in Robotics and Automation, pp. 125–130 (1999)

4. Kim, H., Lee, J.-G., Lee, S.-S., Park, J.H.: A Simulation-based Shipbuilding System for Evaluation of Validity in Design and Manufacturing. In: IEEE International Conference on Systems, Man and Cybernetics, Washington DC, pp. 522–529 (2003)
5. Zheng, X., Sun, G., Wang, S.: Researches on Virtual Prototyping Presentation Model Based on Directed Acyclic Graph. In: Shen, W.-M., Chao, K.-M., Lin, Z., Barthès, J.-P.A., James, A. (eds.) CSCWD 2005. LNCS, vol. 3865, pp. 633–637. Springer, Heidelberg (2006)
6. Gadh, R., Srinivasan, H., Nuggehalli, S.: Virtual Disassembly - A Software Tool For Developing Product Dismantling And Maintenance Systems. In: 1998 Proceedings Annual Reliability and Maintainability Symposium, Anaheim, pp. 120–125 (1998)

The Reduction Method Using the Property of Attribute Set to Divide Object Sky

Jianguo Tang[1] and Fang Tang[1,2]

[1] School of Physics and Electronic Engineering,
Chongqing Three Gorges University,
Chongqing, China
[2] Department of Electronic and Computer Engineering,
Hongkong University of Science and Technology,
Hongkong
sxxytjg@x263.net, icdetf@ust.hk

Abstract. Traditional attribute set reduction method of information system needs to use distinguish matrix, and which is also required generally to a variety of improved methods reported up to now. But the computation workload is very large when the number of dimensions of information system is big. We studied the property of the information system's attribute set to divide object sky and delivered some useful results. Based on it, a reduction method is proposed. By such an approach, the distinguish matrix is not required. As a result, not only the computation of the reduction process is simplified, but also the required memory resources are relaxed. The effectiveness of the method proposed by this paper is verified by an example.

Keywords: Attribute set, divide, object sky, reduction, distinguish matrix.

1 Introduction

The attribute set reduction of information system is one of the main contents of the rough set theory [1]. As a hot research area in the application of the rough set theory, the attribute set reduction aims to eliminate the margin attributes in a given information system and finally a system only with necessary attributes is obtained. For a low dimensional information system, it is not difficult to solve the attribute set reduction. However, it is proved to be a NP hard for a higher dimensional system and as a result, the computational workload becomes unaffordable. Most of researches on reduction aim to reduce the computational workload, on this aspect, the distinguish matrix, proposed by Skowron et al., is a very important result [2], and a variety of improved reduction methods reported up to now generally are based on the distinguish matrix [3][4][5]. Even though the computational workload using those improved reduction methods have been reduced largely, however, the computational workload is still heavy for a high dimensional information system and huge memory is required to store those distinguish matrix elements. Therefore, it is interesting to find an attribute set reduction algorithm with a lower computational workload.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 611–616.
springerlink.com

2 Problem Description

An information system can be described as:

$$S = \langle U, R, V, f \rangle \tag{1}$$

U is the object sky including all the objects in the system. R is the attribute set. $V = \{V_r \mid r \in R\}$ and f is a information function, where $f : U \times R \to V$, by which the values of every object $u_i \in U$ are assigned.

The definition of reduction in rough set theory can be described as: An attribute subset, which belongs to an information system, can keep the discernibility relation of origin information system and if one of elements in this subset is removed, this subset can not keep the discernibility relation. Such attribute subset is called as a reduction set of the attribute set. Therefore, based on this definition above, the key of reduction is how to verify the system's discernibility. One of the widely used methods to verify the system's discernibility is using distinguish matrix, Therefore, many algorithms reported currently are based on distinguish matrix. For an n-dimension information system, the dimension of its distinguish matrix is n×n. Obviously, if n has a large value, the computational workload is huge to construct the distinguish matrix as well as large memory is required. For example, if n equals to 1000 for a information system, there are 1 million elements in its distinguish matrix. Although symmetry can save half of those resources, the requirement is still tough. Furthermore, in the reduction process, the distinguish matrix need to be reconstructed by many times generally, in order to judge that if a selected subset is a reduction set. Current research mainly forces on finding out the hidden property of the distinguish matrix and then reduce the computational workload of reduction. However, those efforts are all based on the distinguish matrix. In this paper, we propose a reduction method without using distinguish matrix, while achieving low memory resources and low computational workload requirement.

3 Main Results

An information system can be divided into n discernable sub-sky by the different combinations with all of the attribute elements. Thus, every sub-sky can exclusively point to one attribute values combination. In attribute set, it may not to be indispensable that all of those attribute elements and as a result, some margin elements exist. Reduction aims just to remove those margin attributes. That is, when some attribute elements are reduced, the object sky can still be divided into n discernable sub-sky by the reduced attribute set. Such a process to reduce the margin attributes is called reduction. In order to convenient to state, some symbols are defined as follows:

For $B \subseteq R$, U_B is defined as a sky division using attribute set B. $|U_B|$ is the amount of the subsets which make up U_B. Then based on the conception above, a definition of discernibility ratio is given as follows:

Definition 1: For an information system S, $\lambda_B = \frac{|U_B|}{|U|}$ indicates the discernibility ratio of S to attribute set $B(B \subseteq R)$.

Based on Definition 1, the main results are given as follows:

Theorem 1: For an information system S, if exists $B \subseteq R$, satisfy that $\lambda_B = \lambda_R$ and $\forall r_i \in B$, where $\lambda_A < \lambda_R$, then B is called as a reduction of R.

Remark 1: Theorem 1 is no need to be proved because it is just another station for the classical definitions of reduction using the discernibility ratio proposed in this paper.

Remark 2: Attribute set B divides the object sky into the non-overlapped sub skies each other with a number of $|U_B|$. A sub sky is named accurate sub sky if it includes only one object. Otherwise, it is named rough sub sky. Obviously, if $|U_B| < |U_R|$, at least one rough sub sky is existed. If $|U_B| = |U_R|$, the entire sub skies in U_B are the accurate sub sky.

Remark 3: If some rough sub skies in U_B can be divided into smaller sub skies when inserting an extra attribute into B, then this added attribute can increase the discernibility of B; otherwise this added attribute cannot increase the discernibility of B.

According to Definition 1, Theorem 1 and remarks mentioned above, the proposed reduction method can be described as follows:

Step 1: select an attribute b_1 from R which includes a maximum number of attribute values and name it as $B = \{b_1\}$. Let $B = b1$, as the original set of B.

Step 2: select an attribute from R into B, by which the rough sub skies in U_B can be divided into small sub skies as many as possible.

Step 3: compute $|U_B|$, if $|U_B| = |U_R|$, then B is a reduction result and computation is terminated. If $|U_B| < |U_R|$, then return to Step 2.

4 An Example

In this section, the proposed reduction method is demonstrated using the information system given in Table 1.

Table 1. A given information system

U	r1	r2	r3	r4	r5	r6
u1	0	1	0	1	0	0
u2	0	0	0	1	1	0
u3	0	0	0	0	2	1
u4	0	1	0	0	1	1
u5	0	1	0	0	0	1
u6	1	0	0	0	2	1
u7	1	1	1	1	2	0
u8	1	1	1	0	1	0
u9	1	0	0	0	1	0
u10	1	0	1	1	2	1
u11	1	1	1	1	1	1
u12	2	1	1	0	0	1
u13	2	0	1	0	0	1
u14	2	0	0	1	1	1
u15	2	0	1	1	0	1
u16	2	0	0	0	1	0
u17	3	1	0	1	1	0
u18	3	0	1	0	2	0
u19	3	1	1	0	1	0
u20	3	1	1	0	2	1
u21	3	0	0	1	2	1

Step 1: find out the attribute which contains a maximum number of attribute values. Attribute 1 has 4 values and attribute 5 has 3, while the others have 2. Therefore, attribute 1 is selected to build original set of B.

Step 2: considering the remained attributes in R, the attribute should be figured out which can generate some sub sets as many as possible after inserting into B. Those sub sets can be expressed from Table 1 as follows:

$$U_B(B=\{r_1,r_2\})=\{(u_1,u_4,u_5),(u_2,u_3),(u_6,u_9,u_{10}),(u_7,u_8,u_{11}),$$
$$(u_{13},u_{14},u_{15},u_{16}),(u_{12}),(u_{17},u_{19},u_{20}),(u_{18},u_{21})\}$$
$$U_B(B=\{r_1,r_3\})=\{(u_1,u_2,u_3,u_4,u_5),(u_6,u_9),(u_7,u_8,u_{10},u_{11}),$$
$$(u_{12},u_{13},u_{15}),(u_{14},u_{16}),(u_{17},u_{21}),(u_{18},u_{19},u_{20})\}$$
$$U_B(B=\{r_1,r_4\})=\{(u_1,u_2),(u_3,u_4,u_5),(u_6,u_8,u_9),(u_7,u_{10},u_{11}),$$
$$(u_{12},u_{13},u_{16}),(u_{14},u_{15}),(u_{17},u_{21}),(u_{18},u_{19},u_{20})\}$$
$$U_B(B=\{r_1,r_5\})=\{(u_1,u_5),(u_2,u_4),(u_3),(u_6,u_7,u_{10}),(u_8,u_9,u_{11}),$$
$$(u_{12},u_{13},u_{15}),(u_{14},u_{16}),(u_{17},u_{19}),(u_{18},u_{20},u_{21})\}$$
$$U_B(B=\{r_1,r_6\})=\{(u_1,u_2),(u_3,u_4,u_5,u_6),(u_7,u_8,u_9),$$
$$(u_{10},u_{11},u_{12},u_{13},u_{14},u_{15}),(u_{16},u_{17},u_{18},u_{19}),(u_{20},u_{21})\}$$

Obviously, by adding attribute 5 can obtain maximum number of sub skies (9 sub skies). Then attribute 5 is added into B.

Step 3: since $|U_B| < |U_R|$, then return to step 2 and we can get:

$$U_B(B=\{r_1,r_5,r_2\}) = \{(u_1,u_5),(u_2),(u_4),(u_3),(u_6,u_{10}),(u_7),(u_8,u_{11}),(u_9),(u_{12}),(u_{13},u_{15}),(u_{14},u_{16}),(u_{17},u_{19}),(u_{18},u_{21}),(u_{20})\}$$

$$U_B(B=\{r_1,r_5,r_3\}) = \{(u_1,u_5),(u_2,u_4),(u_3),(u_6),(u_7,u_{10}),(u_8,u_{11}),(u_9),(u_{12},u_{13},u_{15}),(u_{14},u_{16}),(u_{17}),(u_{19}),(u_{18},u_{20}),(u_{21})\}$$

$$U_B(B=\{r_1,r_5,r_4\}) = \{(u_1),(u_5),(u_2),(u_4),(u_3),(u_6),(u_7,u_{10}),(u_8,u_9),(u_{11}),(u_{12},u_{13}),(u_{15}),(u_{14}),(u_{16}),(u_{17}),(u_{19}),(u_{18},u_{20}),(u_{21})\}$$

$$U_B(B=\{r_1,r_5,r_6\}) = \{(u_1),(u_5),(u_2),(u_4),(u_3),(u_6,u_{10}),(u_7),(u_8,u_9),(u_{11}),(u_{12},u_{13},u_{15}),(u_{14}),(u_{16}),(u_{17},u_{19}),(u_{18}),(u_{20},u_{21})\}$$

Similarly, attribute 4 should be added into B and then, $B=\{r_1,r_5,r_4\}$, where $|U_B|=17<|U_R|=21$. The new attribute is needed to be added into B. Up to now, there are 4 rough sub skies in all sub skies generated by B, that is {u8, u9}, {u7, u10}, {u12, u13}, {u18, u20}. To simplify the result, we only need to consider that the divide ability of the remained attributes to the 4 rough sub skies above. Because attribute 2 can divide all 4 rough sub skies into accurate sub skies, which divide ability is bigger, then add it into B. Finally, the result is $B=\{r_1,r_5,r_4,r_2\}$, while $|U_B|=|U_R|$. The reduction algorithm is finished.

It should be mentioned that, compare to conventional reduction method, the distinguish matrix is need to be constructed firstly. In this example, the number of objects is 21, then the number of elements of the distinguish matrix is 21×21 = 441. Consider symmetry of the distinguish matrix and that the elements on main diagonal line are zero, there are 210 elements to be need to compute. Every element of the distinguish matrix is a attribute subset formed by attributes with different attribute value for corresponding two objects, such that, for object 1 and object 3, corresponding element of distinguish matrix is D13 = {r2, r4, r5, r6}, and for object 4 and object 5, it is D45 = {r5}. Obviously, the computation and memory requirements are much higher than the method proposed in this paper.

5 Conclusion

It is always a hot research area to find out an advanced reduction method in rough set theory. Currently, most of the reported algorithm is based on the distinguish matrix, which requires not only heavy computational workload, but also huge memory resources. In this paper, we proposed a reduction method without using distinguish matrix and an example is demonstrated in the following section. Compared to the conventional distinguish matrix based reduction algorithm, the computation and memory requirements are significantly relaxed.

References

1. Tsang, E.C.C., Chen, D., Yeung, D.S., Wang, X., Lee, J.: Attributes Reduction Using Fuzzy Rough Sets. IEEE Transactions on Fuzzy Systems, 1130–1141 (2008)
2. Skoworn, A., Rauszer, C.: The discernibility matrices and functions information systems. In: Slowinski, I. (ed.) Intelligent Decision Support handbook of Applications and Advances of the Rough Sets Theory, pp. 331–362. Kluwer Academic Publisher, Dordrecht (1991)
3. Tao, Z., Liu, Q., Li, W.: Algorithm for attribute reduction based on improved discernibility matrix. In: Chinese Control and Decision Conference (2007)
4. Tang, J., Tan, M.: On finding core and reduction in rough set theory. Control and Decision, 449–452 (2003)
5. Tang, J., Tang, F.: Reduction of Inconsistent Decision Systems. In: 7th World Congress on Intelligent Control and Automation (2008)

Dynamic Recurrent Fuzzy Wavelet Neural Network Blind Equalization Algorithm

Yecai Guo[1], Fang Xu[1], Lihua Wang[2], and Kan Fan[1]

[1] Nanjing University of Information Science and Technology, Nanjing, 210044, Jiangsu, China
[2] Anhui University of Science and Technology, Huainan, 232001, Anhui, China
{guo-yecai,lisa.xf}@163.com

Abstract. For dynamic system, the traditional fuzzy neural network blind equalization algorithm is bad in equalization performance. In order to overcome this shortcoming, a dynamic recurrent fuzzy wavelet neural network blind equalization algorithm is proposed. While combining the wavelet neural network with static fuzzy neural network and making full use of strong reasoning capacity and powerful adaptability of fuzzy neural network, this proposed algorithm adds memory unit between the normalization layer and the fuzzy layer of the fuzzy neural network, and introduces feedback in the fuzzy neural network, in this way, the proposed algorithm does well in dynamic system. Due to take advantage of strong approximation ability of wavelet function, it also embeds wavelet function into the fuzzy neural network and makes the convergence performance improve greatly. Simulation results show that the convergence performance of the proposed algorithm is better than the general fuzzy neural network blind equalization algorithm.

Keywords: Dynamic recurrent fuzzy neural network, wavelet transform, blind equalization algorithm.

1 Introduction

Fuzzy neural network is the integration of fuzzy logic and neural network system, it uses empirical knowledge of fuzzy systems and learning mechanism of fuzzy systems simultaneously[1][2] and is widely used in the study of channel equalization. However, it is difficult to employ the traditional fuzzy neural network with a static network structure for processing the transient information, so it could do nothing in the dynamic system's equalization[3]. Therefore, it is necessary to design feedback fuzzy neural network equalizer[4].Because wavelet has the feature of good time-frequency localization and zoom, neural network has self-learning, adaptability, powerful robustness and generalization ability[5][6], and the wavelet decomposition is similar to single hidden layer neural network. Their combination is called as wavelet neural network[7][8]. Wavelet neural network has feed-forward wavelet neural network and embedded wavelet neural network.

Based on traditional static fuzzy neural network and embedded wavelet neural network, a dynamic recurrent fuzzy wavelet neural network blind equalization

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 617–624.
springerlink.com

algorithm (DRFWNN) is proposed by means of adding memory unit between the normalization layer and fuzzy layer of the network. This proposed algorithm is used to equalize the dynamic system. Meanwhile, in order to work for complex number modulation systems such as PSK and QAM, this proposed algorithm considers the real and imaginary components of signals separately.

2 Dynamic Recurrent Fuzzy Wavelet Neural Network Blind Equalization Algorithm

2.1 Dynamic Recurrent Fuzzy Wavelet Neural Network

The traditional fuzzy neural network is an integration of fuzzy logic system and a static feed-forward neural network, so even if it has the advantages of both, but it can not be well adapted to dynamic mapping. DRFWNN presented in this paper embeds wavelet function with strong approximation ability and adds recursion layer to the network structure. Its structure is shown in Fig.1.

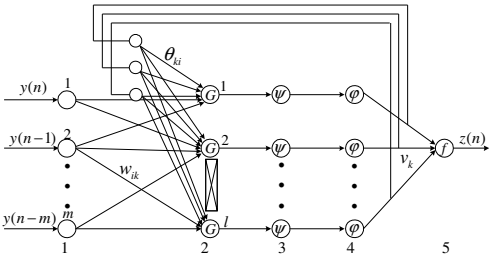


Fig. 1. The Structure of Dynamic Recursive Fuzzy Wavelet Neural Network

In Fig. 1, $y(n)$ is the input of the network, $y_R(n)$ and $y_I(n)$ are respectively its real and imaginary part. w_{ik} is the connection weight between the k th ($k=1,2,\cdots,l$, l is the number of fuzzy node, the below is same) neuron in the second layer and the i th neuron in the first layer, $w_{ik,R}$ and $w_{ik,I}$ are respectively its real and imaginary part. θ_{ki} is recurrent connection weight, $\theta_{ki,R}$ and $\theta_{ki,I}$ are respectively its real and imaginary part. $v_k(n)$ is the connection weight between normalization layer and output layer, $v_{k,R}(n)$ and $v_{k,I}(n)$ are respectively its real and imaginary part. $z(n)$ is the output of the network, $z_R(n)$ and $z_I(n)$ are respectively its real and imaginary part. $I_i^{(t)}$ is the i th ($i=1,2,\cdots,m$, m is the number of input nodes, the below is same) input node's input in the t th ($t=1,2,\cdots,5$, the below is same) layer, $I_{i,R}^{(t)}$ and $I_{i,I}^{(t)}$ are respectively its real and imaginary part . $O_i^{(t)}$ is the i -th input node's output in the t th layer, $O_{i,R}^{(t)}$ and $O_{i,I}^{(t)}$ are respectively its real and imaginary part .

The first layer is its input layer. We have

$$I_i^{(1)}(n) = y_R(n) + j\, y_I(n),\ O_i^{(1)}(n) = I_{i,R}^{(1)}(n) + jI_{i,I}^{(1)}(n) \quad . \tag{1}$$

The second layer is its fuzzy layer. Each neuron in the layer can be as a memory unit and be expressed by membership function. Its input and output are given by

$$\begin{aligned} I_k^{(2)}(n) &= \sum_{i=1}^{m}((w_{ik,R}O_{i,R}^{(1)}(n) - w_{ik,I}O_{i,I}^{(1)}(n)) + \theta_{ki,R}O_{i,R}^{(4)}(n) - \theta_{ki,I}O_{i,I}^{(4)}(n)) \\ &+ j\sum_{i=1}^{m}((w_{ik,R}O_{i,I}^{(1)}(n) + w_{ik,I}O_{i,R}^{(1)}(n)) + \theta_{ki,R}O_{i,I}^{(4)}(n) + \theta_{ki,I}O_{i,R}^{(4)}(n)) \quad . \end{aligned} \tag{2}$$

$$O_k^{(2)}(n) = G(I_{k,R}^{(2)}(n)) + j\, G(I_{k,I}^{(2)}(n)) \quad . \tag{3}$$

where, $G(\cdot)$ is the activation function of the neuron in the second layer, it is also called as Gaussian function, i.e.

$$G(I_{k,R}^{(2)}(n)) = \exp\left\{-\frac{[I_{k,R}^{(2)}(n) - m_k]^2}{[\sigma_k]^2}\right\} \quad . \tag{4}$$

where, m_k is the center of Gaussian membership function, σ_k is the variance of Gaussian membership function.

$G(I_{k,I}^{(2)}(n))$ is obtained by replacing $I_{k,R}^{(2)}(n)$ in Eq.(5) with $I_{k,I}^{(2)}(n)$ (the below is same, as space is limited, the following formula's imaginary part forms are all omitted). In this layer, m_k, σ_k and w_{ik} are adjustable parameters.

The third layer: fuzzy logic layer, the input and output of logic node in this layer are given by

$$I_k^{(3)}(n) = O_k^{(2)}(n) \quad . \tag{5}$$

$$O_k^{(3)}(n) = \psi_{a,b}(I_{k,R}^{(3)}(n)) + j\psi_{a,b}(I_{k,I}^{(3)}(n)) \quad . \tag{6}$$

where, $\psi_{a,b}(\cdot)$ is wavelet transform function, here Marlette wavelet generating function is chosen, i.e.

$$\psi_{a,b}(I_{k,R}^{(3)}(n)) = |a|^{-\frac{1}{2}} \cdot \left(1 - \left(\frac{I_{k,R}^{(3)}(n) - b}{a}\right)^2\right) \cdot e^{-\frac{1}{2}\left(\frac{I_{k,R}^{(3)}(n)-b}{a}\right)^2} \quad . \tag{7}$$

where, b is the shift factor, a is the scale factor.

The aim of the above algorithm is to delete the rule which has the less influence on the output and the corresponding node.

The fourth layer: normalization layer. Its input and output are written as

$$I_k^{(4)}(n) = O_k^{(3)}(n) \quad , \quad O_k^{(4)}(n) = \varphi(I_{k,R}^{(4)}(n)) + j\varphi(I_{k,I}^{(4)}(n)) \tag{8}$$

$$\varphi(I_{k,R}^{(4)}(n)) = \frac{I_{k,R}^{(4)}(n)}{\sum_{k=1}^{l} I_{k,R}^{(4)}(n)} \quad . \tag{9}$$

The fifth layer: output layer. In this layer, defuzzification is carried out mainly through the weighted sum method, namely the linear combination of conclusion drew by each rule in the front layer. The output of this layer and total output are given by

$$\begin{gathered} I^{(5)}(n) = \sum_{k=1}^{l} (v_{k,R} O_{k,R}^{(4)}(n) - v_{k,I} O_{k,I}^{(4)}(n)) + j \sum_{k=1}^{l} (v_{k,R} O_{k,I}^{(4)}(n) + v_{k,I} O_{k,R}^{(4)}(n)), \\ z(n) = f(I_R^{(5)}(n)) + jf(I_I^{(5)}(n)) . \end{gathered} \tag{10}$$

where, $f(\cdot)$ is the transfer function between the input and output of output layer, it controls the output of entire network, i.e.

$$f(I_R^{(5)}(n)) = I_R^{(5)}(n) + \lambda \sin(\pi \cdot I_R^{(5)}(n)) \quad . \tag{11}$$

where, the value of λ influences the pace in which $f(\cdot)$ adjusts output signal. In practical applications, for different signals and channels, the value of λ is different.

Dynamic Recurrent fuzzy wavelet neural network essentially uses the feedback information of wavelet neural network to adjust the various parameters of fuzzy wavelet neural network. Thus it can adjust wavelet neural network dynamically and is suitable for dynamic systems.

2.2 DRFWNN Blind Equalizer

When DRFWNN is introduced into the blind equalizer, we can obtain the dynamic recurrent fuzzy wavelet neural network blind equalization algorithm. The cost function [3] of DRFWNN blind equalization algorithm is defined as

$$J = \frac{1}{2}(|z(n)|^2 - R_{CM})^2 \quad . \tag{12}$$

where, $z(n)$ is output of the DRFWNN, $R_{CM} = E[|a(n)|^4] / E[|a(n)|^2]$, error function $e(n) = |z(n)|^2 - R_{CM}$, $\boldsymbol{a}(n)$ is transmitted signal.

2.3 Dynamic Learning Algorithm of Parameters

In order to minimize the cost function of DRFWNN blind equalization algorithm, the parameters of recurrent fuzzy wavelet neural network are adjusted by using back propagation algorithm and the stochastic gradient descent algorithm. The derivation of iteration formulas are shown as follows. Iteration formula of $v_k(n)$ is given by

$$v_k(n+1) = v_k(n) - \mu_1 \frac{\partial J(n)}{\partial v_k(n)} \quad . \tag{13}$$

$$\frac{\partial J(n)}{\partial v_k(n)} = 2e(n)\left|z_R(n) + jz_I(n)\right|\left(\frac{\partial z(n)}{\partial v_{k,R}(n)} + j\frac{\partial z(n)}{\partial v_{k,I}(n)}\right) \quad . \tag{14}$$

$$\frac{\partial z(n)}{\partial v_{k,R}(n)} = \frac{1}{|z(n)|}[f(I_R^{(5)}(n))f'(I_I^{(5)}(n))Q_{k,R}^{(4)}(n) + f(I_I^{(5)}(n))f'(I_I^{(5)}(n))Q_{k,I}^{(4)}(n)] \tag{15}$$

.

where, μ_1 is the iteration step-size, $f'(\cdot)$ is the derivative, the below is same. Similarly, we can obtain the formula of $\frac{\partial z(n)}{\partial v_{k,I}(n)}$. Substituting the formulas of $\frac{\partial z(n)}{\partial v_{k,R}(n)}$ and $\frac{\partial z(n)}{\partial v_{k,I}(n)}$ into Eq.(16), we can get the iteration formula of $v_k(n)$,i.e.

$$v_k(n+1) = v_k(n) + \mu_1 K(n) O_k^{*(4)}(n) \quad . \tag{16}$$

$$K(n) = -2e(n)[f(I_R^{(5)}(n))f'(I_R^{(5)}(n)) + jf(I_I^{(5)}(n))f'(I_I^{(5)}(n))] \quad . \tag{17}$$

where, $*$ indicates conjugate, the below is same.Iteration formula of $w_{ik}(n)$ is written as

$$w_{ik}(n+1) = w_{ik}(n) + e(n)\mu_2 K_0(n) O_i^{*(1)}(n) \quad . \tag{18}$$

where, μ_2 is the iteration step-size.

$$\begin{aligned} K_0(n) &= \varphi'(I_{k,R}^{(4)}(n))\psi'_{a,b}(I_{k,R}^{(3)}(n))G'(I_{k,R}^{(2)}(n)) \\ &\cdot \mathrm{Re}\{[f(I_R^{(5)}(n))f'(I_R^{(5)}(n)) + jf(I_I^{(5)}(n))f'(I_I^{(5)}(n))]v_k^*(n)\} \\ &\quad + j\varphi'(I_{k,I}^{(4)}(n))\psi'_{a,b}(I_{k,I}^{(3)}(n))G'(I_{k,I}^{(2)}(n)) \\ &\cdot \mathrm{Im}\{[f(I_R^{(5)}(n))f'(I_R^{(5)}(n)) + jf(I_I^{(5)}(n))f'(I_I^{(5)}(n))]v_k^*(n)\} \quad . \end{aligned} \tag{19}$$

Iteration formula of $\theta_{ki}(n)$ is given by

$$\theta_{ki}(n+1) = \theta_{ki}(n) + e(n)\mu_3 K_0(n) O_k^{*(4)}(n) \quad . \tag{20}$$

where, μ_3 is the iteration step, $K_0(n)$ is the same as Eq.(19).

Iteration formula of $m_k(n)$ is written as

$$m_k(n+1) = m_k(n) + e(n)\mu_4 K_1(n) \quad . \tag{21}$$

where, μ_4 is the iteration step-size.

$$K_1(n) = \varphi'(I_{k,R}^{(4)}(n))\psi'_{a,b}(I_{k,R}^{(3)}(n))\frac{\partial G(I_{k,R}^{(2)}(n))}{\partial m_k(n)}$$
$$\cdot \mathrm{Re}\{[f(I_R^{(5)}(n))f'(I_R^{(5)}(n)) + jf(I_I^{(5)}(n))f'(I_I^{(5)}(n))]v_k(n)\}$$
$$+ j\varphi'(I_{k,I}^{(4)}(n))\psi'_{a,b}(I_{k,I}^{(3)}(n))\frac{\partial G(I_{k,I}^{(2)}(n))}{\partial m_k(n)} \tag{22}$$
$$\cdot \mathrm{Im}\{[f(I_R^{(5)}(n))f'(I_R^{(5)}(n)) + jf(I_I^{(5)}(n))f'(I_I^{(5)}(n))]v_k(n)\} \quad .$$

Iteration formula of $\sigma_k(n)$ is written as

$$\sigma_k(n+1) = \sigma_k(n) + e(n)\mu_5 K_2(n) \quad . \tag{23}$$

where, μ_5 is the iteration step-size.

$$K_2(n) = \varphi'(I_{k,R}^{(4)}(n))\psi'_{a,b}(I_{k,R}^{(3)}(n))\frac{\partial G(I_{k,R}^{(2)}(n))}{\partial \sigma_k(n)}$$
$$\cdot \mathrm{Re}\{[f(I_R^{(5)}(n))f'(I_R^{(5)}(n)) + jf(I_I^{(5)}(n))f'(I_I^{(5)}(n))]v_k(n)\}$$
$$+ j\varphi'(I_{k,I}^{(4)}(n))\psi'_{a,b}(I_{k,I}^{(3)}(n))\frac{\partial G(I_{k,I}^{(2)}(n))}{\partial \sigma_k(n)} \tag{24}$$
$$\cdot \mathrm{Im}\{[f(I_R^{(5)}(n))f'(I_R^{(5)}(n)) + jf(I_I^{(5)}(n))f'(I_I^{(5)}(n))]v_k(n)\} \quad .$$

Iteration formulas of $a(n)$ and $b(n)$ are given by

$$a(n+1) = a(n) - \mu_6 \frac{\partial J(n)}{\partial a(n)} \quad , \quad b(n+1) = b(n) - \mu_7 \frac{\partial J(n)}{\partial b(n)} . \tag{25}$$

where, μ_6 and μ_7 are the iteration step-size.

$$\frac{\partial J(n)}{\partial a(n)} = 2e(n)\frac{\partial |z(n)|^2}{\partial a(n)} \quad . \tag{26}$$

$$\frac{\partial |z(n)|^2}{\partial a(n)} = f(I_R^{(5)}(n))f'(I_R^{(5)}(n))\frac{\partial I_R^{(5)}(n)}{\partial a(n)} + f(I_I^{(5)}(n))f'(I_I^{(5)}(n))\frac{\partial I_I^{(5)}(n)}{\partial a(n)} . \tag{27}$$

$$\frac{\partial I_R^{(5)}(n)}{\partial a(n)} = \sum_{k=1}^{I}(v_{k,R}(n)\frac{\partial \varphi(I_{k,R}^{(4)})}{\partial \psi_{a,b}(I_{k,R}^{(3)}(n))}\frac{\partial \psi_{a,b}(I_{k,R}^{(3)}(n))}{\partial a(n)}$$
$$- v_{k,I}(n)\frac{\partial \varphi(I_{k,I}^{(4)})}{\partial \psi_{a,b}(I_{k,I}^{(3)}(n))}\frac{\partial \psi_{a,b}(I_{k,I}^{(3)}(n))}{\partial a(n)}) \quad . \tag{28}$$

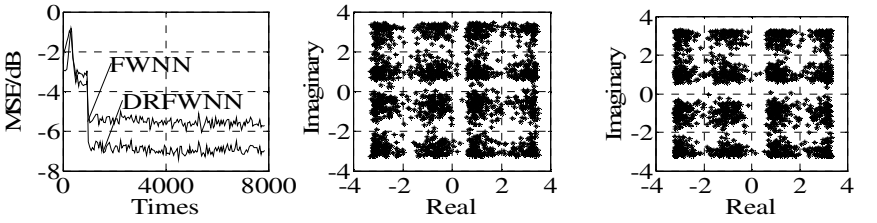
$$\frac{\partial \varphi(I_{k,R}^{(4)})}{\partial \psi_{a,b}(I_{k,R}^{(3)}(n))} = \frac{\sum_{k=1}^{n} I_{k,R}^{(4)}(n) - I_{k,R}^{(4)}(n)}{\left(\sum_{k=1}^{n} I_{k,R}^{(4)}(n)\right)^2} \quad . \tag{29}$$

$$\begin{aligned} \frac{\partial \psi_{a,b}(I_{k,R}^{(3)}(n))}{\partial a(n)} &= -\frac{|a|^{-\frac{3}{2}}}{2}\left(1-\left(\frac{I_{k,R}^{(3)}(n)-b}{a}\right)^2\right) e^{-\frac{1}{2}\left(\frac{I_{k,R}^{(3)}(n)-b}{a}\right)^2} \\ &+ |a|^{-\frac{1}{2}} \frac{(I_{k,R}^{(3)}(n)-b)^2}{a^3}\left(3-\left(\frac{I_{k,R}^{(3)}(n)-b}{a}\right)^2\right) e^{-\frac{1}{2}\left(\frac{I_{k,R}^{(3)}(n)-b}{a}\right)^2} \quad . \end{aligned} \tag{30}$$

The expression of $\partial \varphi(I_{k,I}^{(4)}) / \partial \psi_{a,b}(I_{k,I}^{(3)}(n))$ is obtained by replacing $I_{k,R}^{(4)}(n)$ in Eq.(29) with $I_{k,I}^{(4)}(n)$. In the same way, we can get the expression of $\partial \psi_{a,b}(I_{k,I}^{(3)}(n)) / \partial a(n)$. The iteration formula of a is obtained by substituting Eqs.(26) ~ (30) into Eq.(25). In the same way, we can get the iteration formula of b .

3 Computer Simulation and Analysis

In order to verify the effectiveness of DRFWNN, the simulation experiment is conducted in underwater acoustic channel. And it will be compared with the fuzzy wavelet neural network (FWNN) without recursion layer. Underwater acoustic channel c=zeros(1,1001), $c(1)$=0.076, $c(2)$=0.122, $c(1001)$=1. The transmitted signal is 16QAM, the SNR is 20dB; the length of equalizer is 32. For FWNN equalizer, the weight coefficients are initially set to 0 except that the 8th tap is set to 1, the scale factor and shift factor $a_{FWNN} = 1.09$ and $b_{FWNN} = 0.03$, parameter λ_{FWNN}=8 . For DRFWNN equalizer, the weight coefficients are initially set to 0 except that the 8th tap is set to 1, the scale factor and shift factor $a_{DRFWNN} = 1.05$ and $b_{DRFWNN} = 0.03$, parameter $\lambda_{DRFWNN} = 8$. 800 times Monte Carlo simulation results are shown in Fig.2.



(a) Curves of MSE (b) Output of FWNN (c) Output of DRFWNN

Fig. 2. Simulation Results

Fig.2 indicates that DRFWNN has a drop of 2dB for mean square error(MSE) comparison with FWNN, although the convergence speed of DRFWNN is as quick as FWNN; and the output constellations of DRFWNN is clearer and tighter than FWNN.

4 Conclusions

Dynamic recurrent fuzzy wavelet neural network blind equalization algorithm proposed in this paper uses DRFWNN to replace the traditional blind equalizer. First, the initial center and width of the membership function in fuzzy neural network is determined by adaptive algorithm; then, mapping relation in complex range between the input and output of nonlinear systems is realized by the neuron composed of wavelet bases; next, by adding the recursive layer which can memorize transient information to fuzzy neural network, DRFWNN makes the network have dynamic characteristic and respond dynamic system well; finally, the weights of fuzzy wavelet neural network, shift factor, scale factor and the parameters of the membership function are adjusted on-line. The theoretical analysis of underwater acoustic channel simulation results show that the algorithm has a faster convergence speed and smaller steady-state error and improves the equalization accuracy for the dynamic system.

Acknowledgment. This paper is supported by Specialized Fund for the Author of National Excellent Doctoral Dissertation of China (200753), Natural Science Foundation of Higher Education Institution of Jiangsu Province (08KJB510010) and "the peak of six major talent" cultivate projects of Jiangsu Province (2008 026), Natural Science Foundation of Jiangsu Province (BK 2009410),Natural Science Foundation of Higher Education Institution of Anhui Province (KJ2010A096), Jiangsu Preponderant Discipline “Sensing Networks and Modern Meteorological Equipment”.

References

1. Tan, T., Ren, K., Chen, X., Ma, J.: Fuzzy Neural Network Technology. Journal of Chongqing University of Arts and Sciences (Natural Science Edition) 30(1), 71–74 (2011)
2. Zhang, X.: The Research of Blind Equalization Algorithm based on the Fuzzy Neural Network. Taiyuan University of Technology, Doctor’s Thesis (2008)
3. Zhang, Z.: The Study of Blind Equalization Algorithm based on Recursive Fuzzy Neural Network. Taiyuan University of Technology, Master’s Thesis (2006)
4. Parisi, R., Di Claudio, E.D., Orlandi, G., Rao, B.D.: Fast Adaptive Digital Equalization by Recurrent Neural Networks. IEEE Transactions on Signal Processing 45(11), 2731–2739 (1997)
5. Li, H.: Summary of Research on Fuzzy Neural Network. Journal of Liaoning Institute of Science and Technology 12(2), 15–17 (2010)
6. Xue, Y.: The Research of Multiuser Detection based on Wavelet Neural Network. Taiyuan University of Technology, Master’s Thesis (2007)
7. Xiao, Y., Dong, Y.: A Blind Equalization by Cascaded Hybrid Wavelet Neural Network. Information and Control 38(4), 479–483 (2009)
8. Fan, Y., Sang, Y.: The Research of Nonlinear Control based on Fuzzy Neural Network. In: 2010 International Conference on Electrical and Control Engineering (ICECE), pp. 2417–2420 (2010)

Weighted Multi-Modulus Blind Equalization Algorithm Based on Momentum Term

Shijie Guo[1], Fang Xu[2], Wencai Xu[2], and Kang Fan[1]

[1] College of Physics and Electronic Information,
Huaibei Normal University, Huaibei 235000, China
[2] Nanjing University of Information Science & Technology,
Jiangsu Technology and Engineering Center of Meteorological Sensor Network,
Jiangsu Key Laboratory of Meteorological Observation and Information Processing,
Nanjing 210044, China
{guo-yecai,lisa.xf}@163.com,
523430346@qq.com,
xuwencai0556@126.com

Abstract. In order to overcome the defects of traditional constant modulus blind equalization algorithm in equalizing higher-order QAM(Quadrature Amplitude Modulation) signals, on the basis of combining the momentum term algorithm and the weighted multi-modulus blind equalization algorithm, Momentum term based Weighted Multi-Modulus blind equalization Algorithm(WMMA) is proposed. In the proposed algorithm, weighted multi-modulus blind equalization algorithm is used to adjust the modulus of the cost function adaptively and momentum term is employed to improve the convergence speed. The simulation results in the underwater acoustic channel show that the proposed algorithm has higher convergence speed and accuracy than Multi-Modulus blind equalization Algorithm(MMA) and Constant Modulus blind equalization Algorithm(CMA).

Keywords: blind equalization algorithm, weighted multi-modulus algorithm, momentum term.

1 Introduction

In the fields of wireless communication, multipath fading channel can make the received signals have serious inter-symbol interference, which seriously affects communication efficiency and quality. In order to overcome the influence of the inter-symbol interference on the quality and speed of communication, blind equalization technique without training sequences in the receiver is usually adopted. In blind equalization algorithms, Constant Modulus blind equalization Algorithm (CMA) with simple structure is the most basic and typical and can save the limited bandwidth resources and improve the communication efficiency[1]. However, when it is used to equalize higher-order Quadrature Amplitude Modulation(QAM) signals, but it will result in the larger misadjustment. Multi-Modulus Algorithm(MMA) in equalizing higher-order QAM signals has a good equalization performance comparison with CMA, but it has some disadvantages, such as slow convergence and large mean square error[2]. As shown in[3][4][5][6], Weighted Multi-Modulus Algorithm(WMMA) can

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 625–630.
springerlink.com

obviously improve the convergent rate in equalizing higher-order QAM signals via adjusting the modulus adaptively in the iteration process of weight coefficients.

In this paper, after momentum term[7] is introduced into WMMA, we can get Momentum term based Weighted Multi–Modulus Algorithm (MWMMA). The MWMMA has a fast convergence rate and small mean square error in equalizing higher-order QAM signals. The simulation results in the underwater acoustic channel verify the superiority of the proposed algorithm comparison multi-modulus blind equalization algorithm and constant modulus blind equalization algorithm.

2 Multi-Modulus Algorithm(MMA)

The cost function of MMA is defined as

$$J_{MMA} = \mathrm{E}\{[|z_r(n)|^2 - R_r^2]^2 + [|z_i(n)|^2 - R_i^2]^2\}. \quad (1)$$

According to the gradient descent method, the iteration formula of equalizer weight vector $\boldsymbol{f}(n)$ is given by

$$\boldsymbol{f}(n+1) = \boldsymbol{f}(n) - \mu\{z_r(n)[z_r^2(n) - R_r^2] + j \cdot z_i(n)[z_i^2(n) - R_i^2]\}\boldsymbol{y}^*(n) \quad (2)$$

where μ is iteration step-size, j is imaginary unit, $z_r(n)$ and $z_i(n)$ are in-phase component and quadrature component of equalizer's output signals, respectively. $\boldsymbol{y}(n)$ is the input signal of equalizer. $R_r^2 = \mathrm{E}[|a_r(n)|^4]/\mathrm{E}[|a_r(n)|^2]$, $R_i^2 = \mathrm{E}[|a_i(n)|^4]/\mathrm{E}[|a_i(n)|^2]$. $a_r(n)$ and $a_i(n)$ are the real part and imaginary part of transmitted signal $a(n)$, respectively.

The cost function of CMA is given by

$$J_{CMA} = \mathrm{E}\{[z^2(n) - R_{CMA}^2]^2\} \qquad (R_{CMA}^2 = \mathrm{E}[|a(n)|^4]/\mathrm{E}[|a(n)|^2]). \quad (3)$$

It is found from the Eq.(3) that the CMA only uses the equalizer's output signals, phase fuzzy is existed. In the MMA, the equalizer's output signals are divided into real part and imaginary part, at the same time, there are different modulus values in in-phase component and quadrature component, respectively. As a result, MMA can greatly eliminate the phase fuzziness and improves the equalization performance by using amplitude information and phase information, respectively. However, the error model of MMA doesn't match the square constellation chart of higher-order QAM signals, so the error term in MMA is to increase excess MSE with the order number of the constellations. In order to make full use of the prior information about constellation charts and reduce the steady-state error, it is necessary to select a more suitable error model to match the higher-order QAM constellation charts.

3 Weighted Multi-Modulus Algorithm(WMMA)

In order to further reduce the steady-state error, WMMA[3] is appropriate for equalizing higher-order QAM signals, the cost function of the WMMA is defined as

$$J_{\text{WMMA}} = \text{E}\{[z_r^2(n) - |\hat{z}_r(n)|^{\lambda_r} R_{\lambda_r}^2]^2 + [z_i^2(n) - |\hat{z}_i^2(n)|^{\lambda_i} R_{\lambda_i}^2]^2\}. \tag{4}$$

where weight coefficients $\lambda_r, \lambda_i \in [0,2]$, $\hat{z}_r(n)$ and $\hat{z}_i(n)$ are the real part and imaginary part of $\hat{z}(n)$, $\hat{z}(n)$ is the decision value of equalizer's output signal $z(n)$, $R_{\lambda_r}^2 = \text{E}[|a_r(n)|^4]/\text{E}[|a_i(n)|^{2+\lambda_r}]$, $R_{\lambda_i}^2 = \text{E}[|a_i(n)|^4]/\text{E}[|a_i(n)|^{2+\lambda_i}]$,then the iteration equation of equalizer's weight vector can be written as

$$\begin{aligned} \boldsymbol{f}(n+1) = \boldsymbol{f}(n) - \mu\{ & z_r(n)[z_r^2(n) - |\hat{z}_r(n)|^{\lambda_r} R_{\lambda_r}^2] + \\ & jz_i(n)[z_i^2(n) - |\hat{z}_i(n)|^{\lambda_i} R_{\lambda_i}^2]\}\boldsymbol{y}^*(n). \end{aligned} \tag{5}$$

Clearly, the value of decisive symbol is entirely dependent on the selected constellation charts. For square constellation charts, the source symbols are distributed in in-phase and orthogonal directions with equal probability, so let that $\lambda_r = \lambda_i = \lambda$, $R_{\lambda_r} = R_{\lambda_i} = R_\lambda$. Then Eq.(4) and Eq.(5) can be rewritten as

$$J_{\text{WMMA}} = \text{E}[(z_r^2(n) - |\hat{z}_r(n)|^{\lambda} R_{\lambda}^2)^2 + (z_i^2(n) - |\hat{z}_i^2(n)|^{\lambda} R_{\lambda}^2)^2]. \tag{6}$$

$$\begin{aligned} \boldsymbol{f}(n+1) = \boldsymbol{f}(n) - \mu\{ & z_r(n)[z_r^2(n) - |\hat{z}_r(n)|^{\lambda} R_{\lambda}^2] + \\ & j \cdot z_i(n)[z_i^2(n) - |\hat{z}_i(n)|^{\lambda} R_{\lambda}^2]\}\boldsymbol{y}^*(n). \end{aligned} \tag{7}$$

There is a fundamental distinction between MMA and WMMA. MMA's modulus is just a constant about the statistical characteristics of the source signals, whereas the modulus of the WMMA is related to many factors, such as the statistical characteristics of source signals, weighted factor λ and outputs of decision device, especially the modulus value can be adjusted adaptively during the iteration process.

4 Momentum Weighted Multi–Modulus Algorithm (MWMMA)

Convergence speed is the most important performance index in all performance indexes of blind algorithms, the real-time of the algorithm gets strong when the convergent speed becomes fast. So in order to enhance the real-time of algorithm, momentum term is introduced into the WMMA, Momentum term based Weighted Multi–Modulus Algorithm (MWMMA) is proposed. The principle of proposed algorithm is shown in Fig.1.

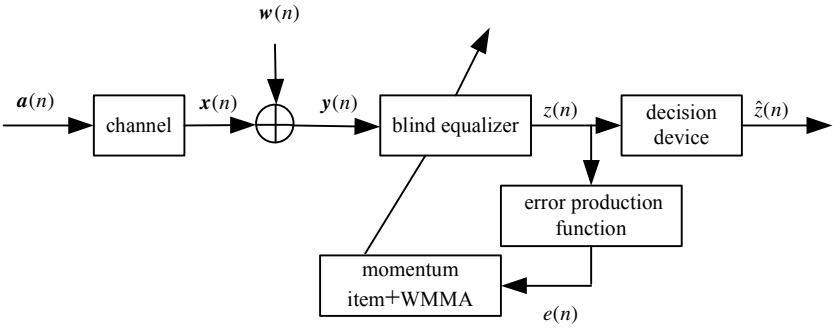


Fig. 1. Principle diagram of MWMMA

From Fig.1, the weight vectors of equalizer can be written as

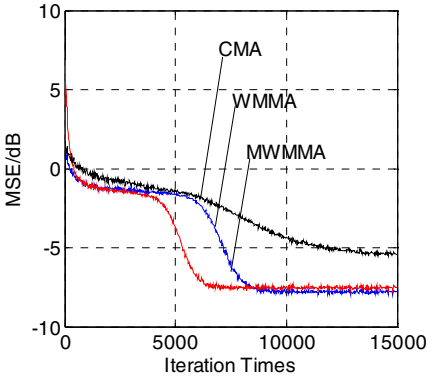
$$\boldsymbol{f}(n+1) = \boldsymbol{f}(n) + \mu \boldsymbol{y}^*(n)e(n) + \alpha_M[\boldsymbol{f}(n) - \boldsymbol{f}(n-1)]. \tag{8}$$

where $e(n) = \{z_r(n)[z_r^2(n) - |\hat{z}_r(n)|^\lambda R_\lambda^2] + jz_i(n)[z_i^2(n) - |\hat{z}_i(n)|^\lambda R_\lambda^2]\}y^*(n)$, α_M is momentum factor and $0 < \alpha_M < 1$. For traditional CMA, the faster is the convergence rate, the larger is the steady-state error. This is a pair of contradiction. While momentum term can improve the convergence speed, at the same time, steady-state error is not changed obviously.

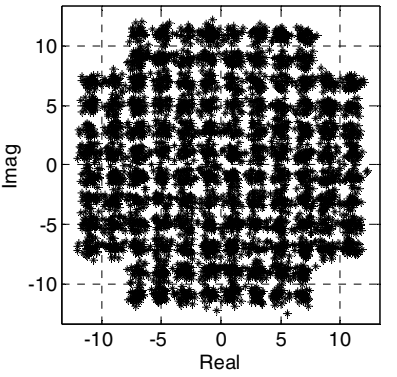
5 Simulation Tests

In order to verify the performance of the MWMMA, WMMA and CMA, respectively, simulation tests were carried out. In the tests, underwater acoustic channel $\boldsymbol{c} = [0.3132, -0.1040, 0.8908, 0.3134]$, signal to noise ratio(SNR) was 30dB, weighted coefficient $\lambda = 1.2$, momentum factor $\alpha_M = 0.2$, the length of equalizer was 16, the center-tap was initialized as 1, the step-size of the MWMMA , WM- MA and CMA, respectively , was 0.0000003,0.0000003, and 0.0000001, 128QAM signals were transmitted. The simulation results were shown in Fig.2.

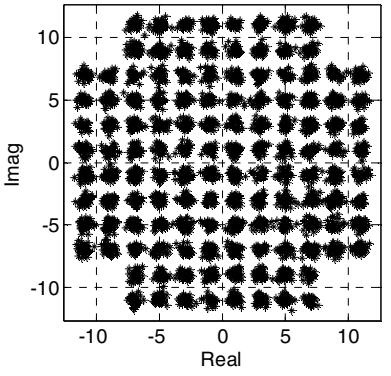
Fig.2(a) shows that the MWMMA has an improvement of about 2000 steps and 7500 steps for convergence rate comparison with the WMMA and CMA. The MWMMA has a drop of 2.5dB for mean square error (MSE) comparison with the CMA and has slightly different with the WMMA. In addition, from Fig.s2(b)～2(d) , we can know that the output constellations of MWMMA are cleares and most focus in all algorithms.



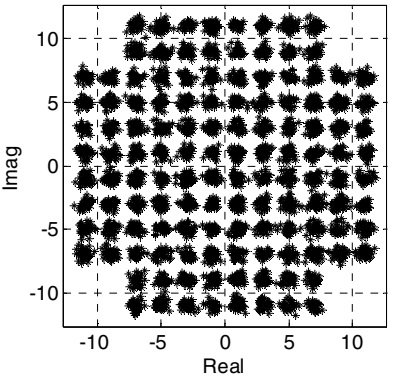
(a) Curves of mean square error(MSE)



(b) Output constellations of CMA



(c) Output constellations of WMMA



(d) Output constellations of MWMMA

Fig. 2. Simulation results

6 Conclusions

When the traditional constant modulus blind equalization algorithm(CMA) are used to equalize higher-order QAM signals, it often produces phase fuzzy and has poor convergent performance. Traditional CMA is not suitable for equalizing higher-order QAM signals. In this paper, momentum term based weighted multi-modulus algorithm is proposed. The proposed algorithm adapts weighted multi-modulus blind equalization algorithm to adjust the modulus of the cost function adaptively in order to reduce the steady-state error and uses momentum term to improve the convergent speed. Theoretical analyses and simulation results show the proposed MWMMA algorithm can greatly improve the convergence rate and reduce the steady mean square error.

Acknowledgment. This paper is supported by Specialized Fund for the Author of National Excellent Doctoral Dissertation of China (200753), Natural Science

Foundation of Higher Education Institution of Jiangsu Province (08KJB510010) and "the peak of six major talent" cultivate projects of Jiangsu Province (2008 026), Natural Science Foundation of Jiangsu Province (BK 2009410),Natural Science Foundation of Higher Education Institution of Anhui Province (KJ2010A096), Jiangsu Preponderant Discipline “Sensing Networks and Modern Meteorological Equipment”.

References

1. Guo, Y.-C.: Adaptive Blind Equalization Technology. Press of Hefei University of Technology (2007) (in Chinese)
2. Abrar, S., Nandi, A.K.: An adaptive constant modulus blind equalization algorithm and its stochastic stability analysis. IEEE Transaction on Digital Object Identifier 17, 55–58 (2010)
3. Wang, Z., Qin, K.: Decision feedback equalizer based on multi-modulus algorithm. Video Engineering 33, 19–20 (2009)
4. Xu, X., Dai, X., Xu, P.: Weighted multi-modulus blind equalization algorithm for high-order QAM signals. Journal of Electronics & Information Technology 29, 1352–1355 (2007) (in Chinese)
5. Guo, Y.-C., Zhang, Y.-P.: Dual-mode multi-modulus blind equalization algorithm for high-order QAM Signals. Journal of System Simulation 20, 1423–1426 (2008) (in Chinese)
6. Rao, W., Yuan, K.-M., Guo, Y.-C., Yang, C.: A simple constant modulus algorithm for blind equalizer suitable for 16-QAM signals. In: International Conference on Signal Processing Proceedings, pp. 1963–1966 (2008)
7. Han, Y., Guo, Y., Li, B., Zhou, Q.: Momentum term and orthogonal wavelet- based blind equalization algorithm. Journal of System Simulation 20, 1559–1562 (2008) (in Chinese)

Stability of Stationary φ-Mixing and β-Mixing Processes for Replacement Case

Ke Luo[1], Fen Zhou[2], and Wei Gao[3,4]

[1] Shaoyang University Library, Shaoyang University, Shaoyang, Hunan, China
[2] Department of Computer Science, Zhaotong Teacher's College, Zhaotong, Yunnan China
[3] Department of Information, Yunnan Normal University, Kunming, Yunnan, China
[4] Department of Mathematics, Soochow University, Suzhou, Jiangsu, China
luoke00@qq.com, Zhoufen_85@163.com, gaowei@ynnu.edu.cn

Abstract. This paper studies the scenario where the observations are drawn from a stationary φ-mixing sequence, a widely adopted assumption in the study of non-i.i.d. processes that implies dependence between observations weakening over time. In this paper, the stability of φ-mixing and β-mixing processes in replacement case is investigated, notions of stability for φ-mixing and β-mixing processes in replacement case is defined, and the generalization bounds of stable φ-mixing and β-mixing processes are obtained.

Keywords: learning in non-i.i.d. scenarios, strong stability, weak stability, mixing distribution, generalization bounds.

1 Introduction

The generalization properties of ranking algorithms are central focuses in their research. Most generalization bounds in some learning algorithm are based on some measures of the complexity of the hypothesis used like VC-dimension, covering number, Rademacher complexity and so on. However, the notion of algorithmic stability can be used to derive bounds that tailored to specific learning algorithms and exploit their particular properties. A ranking algorithm is called stable if for a wild change of samples, the ranking function doesn't change too much. Some theory analysis can be found in [1-5].

As in much of learning theory, existing stability analyses and bounds apply only in the scenario where the samples are independently and identically distributed (i.i.d.). In many machine learning applications, this assumption, however, does not hold; in fact, the i.i.d. assumption is not tested or derived from any data analysis. The observations received by the learning algorithm often have some inherent temporal dependence. This is clear in system diagnosis or time series prediction problems. Clearly, prices of different stocks on the same day, or of the same stock on different days, may be dependent. But, a less apparent time dependency may affect data sampled in many other tasks as well.

This paper studies the scenario where the observations are drawn from a stationary φ-mixing and β-mixing sequence, a widely adopted assumption in the

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 631–637.
springerlink.com

study of non-i.i.d. processes that implies a dependence between observations weakening over time . The difference from [6] is that: we consider stability under replacement case in our paper.

2 Setting and Definitions

A sequence of random variables $Z=\{Z_t\}_{t=-\infty}^{\infty}$ is said to be stationary if for any t and non-negative integers m and k, the random vectors $(Z_t,\ldots,Z_{t+m})$ and $(Z_{t+k}, \ldots,Z_{t+m+k})$ have the same distribution. Thus, the index t or time, does not affect the distribution of a variable Z_t in a stationary sequence. This does not imply independence however. In particular, for $i< j<k$, $\Pr[Z_j \mid Z_i]$ may not equal $\Pr[Z_k \mid Z_i]$, that is, conditional probabilities may vary at different points in time.

Let $Z=\{Z_t\}_{t=-\infty}^{\infty}$ be a stationary sequence of random variables. For any $i, j\in Z \cup \{-\infty,+\infty\}$, let σ_i^j denote the σ-algebra generated by the random variables Z_k, $i\le k\le j$. Then, forany positive integer k, the φ-mixing and β-mixing coefficients of the stochastic process Z are defined as

$$\varphi(k)=\sup_{n,A\in\sigma_{n+k}^{\infty},B\in\sigma_{-\infty}^{n}}\left|\Pr[A|B]-\Pr[A]\right|,$$

$$\beta(k)=\sup_{n}\mathrm{E}_{B\in\sigma_{-\infty}^{n}}[\sup_{A\in\sigma_{n+k}^{\infty}}\left|\Pr[A|B]-\Pr[A]\right|].$$

Z is said to be φ-mixing (β-mixing) if $\varphi(k)\to 0$ (resp. $\beta(k)\to 0$) as $k\to\infty$. It is said to be algebraically φ-mixing (algebraically β-mixing) if there exist real numbers $\varphi_0>0$ (resp. $\beta_0>0$) and $r>0$ such that $\varphi(k)\le\varphi_0/k^r$ (resp. $\beta(k)\le\beta_0/k^r$) for all k, exponentially mixing if there exist real numbers $\varphi_0>0$ (resp. $\beta_0>0$), $\varphi_1>0$ (resp. $\beta_1>0$) and $r>0$ such that $\varphi(k)\le\varphi_0\exp(-\varphi_1 k^r)$ (resp. $\beta(k)\le\beta_0\exp(-\beta_1 k^r)$) for all k. $\varphi(k)$ and $\beta(k)$ measure the dependence of an event on those that occurred more than k units of time in the past. β-mixing is a weaker assumption than φ-mixing and thus covers a more general non-i.i.d. scenario.

Consider the familiar supervised learning setting where the learning algorithm receives a sample of m labeled points $S=(z_1,\ldots,z_m)=((x_1,y_1),\ldots,(x_m,y_m))\in(X\times Y)^m$, where X is the input space and Y the set of labels ($Y\subseteq R$ in the regression case), both assumed to be measurable. For a fixed learning algorithm, we denote by h_S the hypothesis it returns when trained on the sample S. The error of a hypothesis on a pair $z\in X\times Y$ is measured in terms of a cost function $c: Y\times Y\to R^+$. Thus, $c(h(x), y)$ measures the error of a hypothesis h on a pair (x, y), $c(h(x), y) = (h(x)-y)^2$ in the standard regression cases. We will often use the shorthand $c(h, z) := c(h(x), y)$ for a hypothesis h and $z = (x, y)\in X\times Y$ and will assume that c is upper bounded by a

constant $M>0$. We denote by $\hat{R}(h)$ the empirical error of a hypothes is h for a training sample $S=(z_1, \ldots, z_m)$:

$$\hat{R}(h)=\frac{1}{m}\sum_{i=1}^{m} c(h, z_i).$$

In the standard machine learning scenario, the sample pairs $z_1, \ldots, z_m$ are assumed to be i.i.d.,a restrictive assumption that does not always hold in practice. We will consider here the more general case of dependent samples drawn from a stationary mixing sequence Z over $X \times Y$. As in the i.i.d. case, the objective of the learning algorithm is to select a hypothesis with small error over future samples. But, here, we must distinguish two versions of this problem.

In the most general version, future samples depend on the training sample S and thus the generalization error or true error of the hypothesis h_S trained on S must be measured by its expected error conditioned on the sample S:

$$R(h_S)=\underset{z}{E}[c(h_S, z)\big|S].$$

This is the most realistic setting in this context, which matches time series prediction problems. A somewhat less realistic version is one where the samples are dependent, but the test points are assumed to be independent of the training sample S. The generalization error of the hypothesis h_S trained on S is then:

$$R(h_S)=\underset{z}{E}[c(h_S, z)\big|S]=\underset{z}{E}[c(h_S, z)].$$

[6] defined the uniform stability for φ-mixing processes in the case of deleting one sample and obtained some conclusions relative on uniform stability. In this paper, we define uniform stability in replacement case which are also good measures to show how robust a φ-mixing processes is. We assume S^i be sample set such that differ by replace i-th ($1 \le i \le m$) single point z with $z^{'}$.

Definition 1. A learning algorithm is said to be uniformly $\hat{\beta}$-stable if the hypotheses it returns for any two training samples S and S^i satisfy

$$\forall z \in X \times Y, \ \left|c(h_S, z)-c(h_{S^i}, z)\right| \le \hat{\beta}.$$

The main tool to get generalization bounds for uniform φ-mixing processes is well known McDiarmid inequality as follows.

Lemma 1 [7]. Let $X_1,\ldots, X_N$ be independent random variables, each taking values in a set C. Let $\phi: C^N \to \mathrm{R}$ be such that for each $k \in \{1,\ldots,N\}$, there exist $c_k>0$ such that

$$\sup_{x_1,\cdots,x_N \in C, x_k^{'} \in C} \left|\phi(x_1,\cdots,x_N)-\phi(x_1,\cdots,x_{k-1},x_k^{'},x_{k+1},\cdots,x_N)\right| \le c_k,$$

Then for any $\varepsilon>0$,

$$\mathbf{P}\{\left|\phi(X_1,\cdots,X_N)-\mathbf{E}\{\phi(X_1,\cdots,X_N)\}\right| \ge \varepsilon\} \le 2(e^{-2\varepsilon^2/\sum_{k=1}^{N} c_k^2}).$$

3 φ-Mixing Generalization Bounds

Let $\Phi(S) = R(h_S) - \hat{R}(h)$. We denote by $R(h_S) = \underset{z}{E}[c(h_S, z)|S]$ the expectation in the dependent case and by $\tilde{R}(h_{S_b}) = \underset{\tilde{z}}{E}[c(h_{S_b}, \tilde{z})]$ the expectation where the test points are assumed independent of the training, with S_b denoting a sequence similar to S but with the last b points removed. The block S_b is assumed to have exactly the same distribution as the corresponding block of the same size in S.

Follow the trick used in Lemma 5, Lemma 6 and Lemma 7 of [6], we can get following three results.

Lemma 2. Assume that the learning algorithm is $\hat{\beta}$-stable and that the cost function c is bounded by M. Then, for any sample S of size m drawn from a φ-mixing stationary distribution and for any $b \in \{0, \ldots, m\}$, the following holds:

$$\left|R(h_S) - \tilde{R}(h_{S_b})\right| \leq \frac{b\hat{\beta}}{2} + M\varphi(b).$$

Lemma 3. Let S and S^i be two sequences drawn from a φ-mixing stationary process, and let h_S and h_{S^i} be the hypotheses returned by a $\hat{\beta}$-stable algorithm when trained on each of these samples. Then, for any $i \in \{1, \ldots, m\}$, the following inequality holds:

$$\left|\Phi(S) - \Phi(S^i)\right| \leq (b+2)\hat{\beta} + 2M\varphi(b) + \frac{M}{m}.$$

Lemma 4. Let h_S be the hypothesis returned by a $\hat{\beta}$-stable algorithm trained on a sample S drawn from a β-mixing stationary distribution. Then, for all $b \in [1, m]$, the following inequality holds:

$$\underset{S}{E}[\left|\Phi(S)\right|] \leq (3b+1)\hat{\beta} + 3\beta(b)M.$$

Theorem 5.1 of Kontorovich and Ramanan [8] shows that for a φ-mixing distribution and a 1-Lipschitz function, the constants ci can be bounded as follows in Lemma 1: $c_i \leq 1 + 2\sum_{k=1}^{m-1}\varphi(k)$.

Follow the trick used in Theorem 11 and Theorem 12 of [6], and use the result of Lemma 3 and Lemma 4, we can get our main results as follows.

Theorem 1. Let h_S denote the hypothesis returned by a $\hat{\beta}$-stable algorithm trained on a sample S drawn from a φ-mixing stationary distribution and let c be a measurable non-negative cost function upper bounded by $M>0$, then for any $b \in \{0, \ldots, m\}$ and any $\varepsilon > 0$, the following generalization bound holds:

$$\Pr_S[\left|R(h_S)-\hat{R}(h_S)\right|>\varepsilon+(3b+1)\hat{\beta}+6M\varphi(b)]$$
$$\leq 2\exp\left(\frac{-2\varepsilon^2(1+2\sum_{k=1}^{m-1}\varphi(k))^{-2}}{m((b+2)\hat{\beta}+2M\varphi(b)+M/m)^2}\right).$$

Theorem 2. Let h_S denote the hypothesis returned by a $\hat{\beta}$-stable algorithm trained on a sample S drawn from an algebraically φ- mixing stationary distribution, $\varphi(k)=\varphi_0 k^{-r}$ with $r>1$, and let c be a measurable non-negative cost function upper bounded by $M>0$, then, for any $\varepsilon>0$, the following generalization bound holds:

$$\Pr_S[\left|R(h_S)-\hat{R}(h_S)\right|>\varepsilon+4\hat{\beta}+(r+1)6M\varphi(b)]$$
$$\leq 2\exp\left(\frac{-2\varepsilon^2(1+2\varphi_0 r/(r-1))^{-2}}{m(3\hat{\beta}+(r+1)2M\varphi(b)+M/m)^2}\right),$$

where $b=\left(\frac{\hat{\beta}}{2r\varphi_0 M}\right)^{-1/(r+1)}$.

4 β-Mixing Generalization Bounds

As in the previous section, $\Phi(S)$ is defined by $\Phi(S) = R(h_S) - \hat{R}(h)$. To simplify the presentation, here, we define the generalization error of h_S by $R(h_S)=E_z[c(h_S, z)]$. Thus, test samples are assumed independent of S. Note that for any block of points $Z = z_1, \ldots, z_m$ drawn independently of S, the following equality holds:

$$\underset{Z}{E}\left[\frac{1}{|Z|}\sum_{z\in Z}c(h_S,z)\right]=\frac{1}{k}\sum_{i=1}^{k}\underset{Z}{E}[c(h_S,z_i)]=\frac{1}{k}\sum_{i=1}^{k}\underset{Z_i}{E}[c(h_S,z_i)]=\underset{z}{E}[c(h_S,z)].$$

since, by stationarity, $\underset{Z_i}{E}[c(h_S,z_i)]=\underset{Z_j}{E}[c(h_S,z_j)]$ for all $1\leq i,j\leq k$. Thus, for any such block Z, we can write $R(h_S)=\underset{Z}{E}\left[\frac{1}{|Z|}\sum_{z\in Z}c(h_S,z)\right]$. For convenience, we extend the cost function c to blocks as follows:

$$c(h, Z)=\frac{1}{|Z|}\sum_{z\in Z}c(h_S,z).$$

With this notation, $R(h_S)=\underset{Z}{E}[c(h_S,z)]$ for any block drawn independently of S, regardless of the size of Z.

From a sample S made of a sequence of m points, we construct two sequences of blocks S_a and S_b, each containing μ blocks. Each block in S_a contains a points and each block in S_b contains b points. S_a and S_b form a partitioning of S; for any $a,b \in \{0,\ldots,m\}$ such that $(a+b)\ \mu = m$, they are defined precisely as follows:

$$S_a = (Z_1^{(a)}, \ldots, Z_\mu^{(a)}), \text{ with } Z_i^{(a)} = z_{(i-1)(a+b)+1}, \ldots, z_{(i-1)(a+b)+a}$$

$$S_b = (Z_1^{(b)}, \ldots, Z_\mu^{(b)}), \text{ with } Z_i^{(b)} = z_{(i-1)(a+b)+a+1}, \ldots, z_{(i-1)(a+b)+a+b}$$

For all $i \in \{1, \ldots, \mu\}$. We shall consider similarly sequences of i.i.d. blocks $\tilde{Z}_i^a$ and $\tilde{Z}_i^b$, $i \in \{1, \ldots, \mu\}$, such that the points within each block are drawn according to the same original β-mixing distribution and shall denote by $\tilde{S}_a$ the block sequence $(\tilde{Z}_1^{(a)}, \ldots, \tilde{Z}_\mu^{(a)})$.

Follow the trick used in Lemma 17 and Lemma 18 of [6], we can get following two results.

Lemma 5. Let $\tilde{S}_a$ be an independent block sequence as defined above, then the following bound holds for the expectation of $\left|\Phi(\tilde{S}_a)\right|$:

$$\mathop{\mathrm{E}}_{S_a}[\left|\Phi(\tilde{S}_a)\right|] \le a\hat{\beta}.$$

Lemma 6. Assume a $\hat{\beta}$-stable algorithm. Then, for a sample S drawn from a β-mixing stationary distribution, the following bound holds:

$$\mathop{\mathrm{E}}_{S}[\left|\Phi(S)\right| \ge \varepsilon] \le \Pr_{\tilde{S}_a}[\left|\Phi(\tilde{S}_a)\right| - \mathrm{E}[\left|\Phi(\tilde{S}_a)\right|] \ge \varepsilon_0^{'}] + (\mu - 1)\beta(b),$$

where $\varepsilon_0^{'} = \varepsilon - \frac{\mu b M}{m} - \mu b\hat{\beta} - \mathrm{E}_{\tilde{S}_a^{'}}[\left|\Phi(\tilde{S}_a^{'})\right|]$.

Follow the trick used in Theorem 19 and Corollary 20 of [1], and use the result of Lemma 5 and Lemma 5, we can get our main results as follows.

Theorem 3. Assume a $\hat{\beta}$-stable algorithm and let $\varepsilon^{'}$ denote $\varepsilon - \frac{\mu b M}{m} - \mu b\hat{\beta} - a\hat{\beta}$ as in Lemma 6. Then, for any sample S of size m drawn according to a β-mixing stationary distribution, any choice of the parameters a,b, $\mu > 0$ such that $(a+b)\ \mu = m$, and $\varepsilon \ge 0$ such that $\varepsilon^{'} \ge 0$, the following generalization bound holds:

$$\Pr_S[\left|R(h_S) - \hat{R}(h_S)\right| \ge \varepsilon] \le \exp\left(\frac{-2\varepsilon^{'2} m}{(2a\hat{\beta}m + (a+b)M)^2}\right) + (\mu - 1)\beta(b).$$

Theorem 4. Assume a $\hat{\beta}$ -stable algorithm and let δ' denote $\delta - (\mu - 1)\beta(b)$. Then, for any sample S of size m drawn according to a β -mixing stationary distribution, any choice of the parameters a,b, $\mu > 0$ such that $(a+b)$ $\mu = m$, and $\delta \geq 0$ such that $\delta' \geq 0$, the following generalization bound holds with probability at least $(1-\delta)$:

$$\left|R(h_S) - \hat{R}(h_S)\right| \leq \mu b\left(\frac{M}{m} + \hat{\beta}\right) + a\hat{\beta}\left(M\frac{m}{\mu} + 2a\hat{\beta}m\right)\sqrt{\frac{\log(1/\delta')}{2m}}.$$

Acknowledgments. We would like to thank the anonymous referees for providing us with constructive comments and suggestions.

References

1. Agarwal, S., Niyogi, P.: Stability and Generalization of Bipartite Ranking Algorithms. In: Auer, P., Meir, R. (eds.) COLT 2005. LNCS (LNAI), vol. 3559, pp. 32–47. Springer, Heidelberg (2005)
2. Agarwal, S., Niyogi, P.: Generalization bounds for ranking algorithms via algorithmic stability. Journal of Machine Learning Research 10, 441–474 (2009)
3. Gao, W., Zhang, Y., Liang, L., Xia, Y.: Stability analysis for ranking algorithms. In: 2010 IEEE International Conference on Information Theory and Information Security, pp. 973–976. IEEE Press, Beijing (2010)
4. Kutin, S., Niyogi, P.: Almost-everywhere algorithmic stability and generalization error. In: Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence (2002)
5. Rakhlin, A., Mukherjee, S., Poggio, T.: Stability results in learning theory. Analysis and Applications 3, 397–417 (2005)
6. Mohri, M., Rostamizadeh, A.: Stability Bounds for Stationary φ -mixing and β-mixing Processes. Journal of Machine Learning Research 11, 789–814 (2010)
7. McDiarmid, C.: On the method of bounded differences. In: Surveys in Combinatorics 1989, pp. 148–188. Cambridge University Press (1989)
8. Kontorovich, L., Ramanan, K.: Concentration inequalities for dependent random variables via the martingale method. Annals of Probability 36, 2126–2158 (2008)

A Purely MUX Based High Speed Barrel Shifter VLSI Implementation Using Three Different Logic Design Styles

Abhijit Asati[1] and Chandra Shekhar[2]

[1] Assistant Professor, Electrical & Electronics Engineering Department, BITS, Pilani, India
abhijitmicro@gmail.com
[2] Director, Central Electronics Engineering Research Institute, Pilani, India
chandra@ceeri.ernet.in

Abstract. Barrel shifter is one of the important data path elements and widely used in many key computer operations from address decoding to computer arithmetic, using basic operations like data shifting and rotation. In this paper MUX based barrel shifter circuits are designed and implemented in 0.6µm, N-well CMOS process using three different logic design styles, namely, optimized static CMOS, transmission gate (TG) CMOS and dual rail domino CMOS logic. The proposed barrel shifter architecture implementation shows large reduction in the propagation delay, while keeping the almost similar average power consumption as compared to the implementation by Ramin Rafati [10]. A further inter comparison of implementations using three different logic design styles for various performance and characteristic parameters like circuit delay, average power, maximum instantaneous power, leakage Power, transistor count, layout core area, total routing length and number of via are presented.

Keywords: Barrel shifter, MUX based, data shifting, rotation, left, right.

1 Introduction

Many key computer operations from address generation to arithmetic functions can be performed using key operations like data shifting and rotation. Shifting a single data bit one field at a time can be a slow process; this is where a barrel shifter comes in. A barrel shifter is a combinational logic device/circuit that can shift or rotate a data word by desired number of bits in a single operation. It is another most important block in DSP circuit. It is used for floating-point normalization, word pack/unpack, and field extraction from a bit stream, editing, data modification, and arithmetic manipulation. Barrel shifter can be implemented as a combinational logic circuit, using conventional multiplexes, decoders, and logic gates. Different barrel shifter architectures show tradeoffs between silicon area and speed of operation. A significant reduction in area and power occupied by the barrel shifter circuit is achieved by implementing rightward operations as operations in leftward direction [1], [2]. Some architecture has dedicated blocks for all operations to be performed by a barrel shifter; hence these architectures are faster but consume larger silicon area and power. In this paper we present a MUX based barrel shifter architecture design

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 639–646.
springerlink.com

and its VLSI implementation using three different logic design styles. The VLSI implementations of barrel shifter circuits are done using 0.6µm, N-well CMOS process (SCN_SUBM, lambda=0.3) of MOSIS, in three different logic design styles, namely, optimized static CMOS, TG CMOS and dual rail domino CMOS logic. Section 2 explains the MUX based barrel shifter architecture, section 3 describes 4-bit MUX based barrel shifter architecture design and fill bit logic; section 4 describes design of a 16-bit barrel shifter. Physical implementations and their results are described in section 5 and section 6 concludes the paper.

2 MUX Based Barrel Shifter Architecture

The MUX based barrel shifter architecture are designed using 4:1, 8:1, 16:1, 32:1 and 64:1 multiplexers. The larger size MUX can be designed hierarchically. A 2^n:1 MUX will have 2^n input lines, n select lines and one-output lines. MUX with any n, greater than one can be implemented using 2:1 MUX. One 2^n: 1 MUX requires (2^n-1) 2:1 MUX s and has a delay of 'n' 2:1 MUX. In this architecture n=1, 2, 3, 4, 5, 6 are used in the design 2:1, 4:1, 8:1, 16:1, 32:1 and 64:1 multiplexer. The 2:1 MUX is first designed using three different logic design styles; it is then used to design the 4:1 MUX. The 8:1 MUX is designed using the 4:1 MUX as the basic building block. Similar hierarchy is followed in the design of 16:1, 32:1 and 64:1 multiplexer.

3 4-Bit MUX Based Barrel Shifter Architecture Design and Fill Bit Logic

The table 1 shows the behaviour of 4-bit MUX-based barrel shifter, which utilizes control inputs D for direction, S/R for operation (shift/rotate) and S_1, S_0 for number of bits to be shifted or rotated. Y_0, Y_1, Y_2, Y_3 represent the output bits and I_0, I_1, I_2, I_3 are input bits. The circuit can be used to rotate or shift a 4-bit word in and direction by 0, 1, 2, 3 bits. D='0' means the direction of shift/rotate operation is towards left and D='1' means the direction of shift/rotate operation is towards right. 'F' represents the fill bit, fill bit is '0' for left shift and fill bit is MSB bit for right shift operation. The line S/R='1' for rotate operation and S/R= '0' for shift operation. The bits S_1, S_0 are length selection bits. S_1S_0='00' means length is zero bit, S_1S_0= '01' means length is one bit, S_1S_0= '10' means length is two bits and S_1S_0= '11' means length is three bits. Table 1 explains the various operations performed by 4-bit barrel shifter. As there are four control inputs, we need a 16:1 multiplexer for each output bit. Thus for 4 output bits, we need four 16:1 MUX in the design of a 4-bit barrel shifter, in addition we need a 2:1 MUX for fill bit as shown in figure 1. Thus total 2:1 MUX required in design are 61. Each row of a truth table can be implemented with a dedicated 16:1 MUX circuit, which is designed using 2:1 MUX cell, to obtain final output.

Fill bit are required only in case of shift operation. In left shift operation, the lower significant bits are filled with '0's. While in the right shift operation, in order to preserve the sign of the input number, MSB is sign extended. This is accomplished using a 2:1 MUX. The D='0' represent the left operation for which fill bit is '0' and for D='1' means the right operations for which fill bit is MSB bit of input. The Fill

bits are utilized for shift operation, while it is discarded for rotation operation. Figure 2 shows the schematic of a 4-bit MUX-based barrel shifter, which also includes fill bit logic. The critical delay of the circuit is 5 times the delay of a 2:1 MUX cell, without considering the interconnect delay.

Table 1. Truth-Table for 4-bit barrel shifter operation

Operation	D	S/R	S_1	S_0	Y_3	Y_2	Y_1	Y_0
Arithmetic Shift Left	0	0	0	0	I_3	I_2	I_1	I_0
	0	0	0	1	I_2	I_1	I_0	F
	0	0	1	0	I_1	I_0	F	F
	0	0	1	1	I_0	F	F	F
Rotate left	0	1	0	0	I_3	I_2	I_1	I_0
	0	1	0	1	I_2	I_1	I_0	I_3
	0	1	1	0	I_1	I_0	I_3	I_2
	0	1	1	1	I_0	I_3	I_2	I_1
Arithmetic Shift right	1	0	0	0	I_3	I_2	I_1	I_0
	1	0	0	1	F	I_3	I_2	I_1
	1	0	1	0	F	F	I_3	I_2
	1	0	1	1	F	F	F	I_3
Rotate right	1	1	0	0	I_3	I_2	I_1	I_0
	1	1	0	1	I_0	I_3	I_2	I_1
	1	1	1	0	I_1	I_0	I_3	I_2
	1	1	1	1	I_2	I_1	I_0	I_3

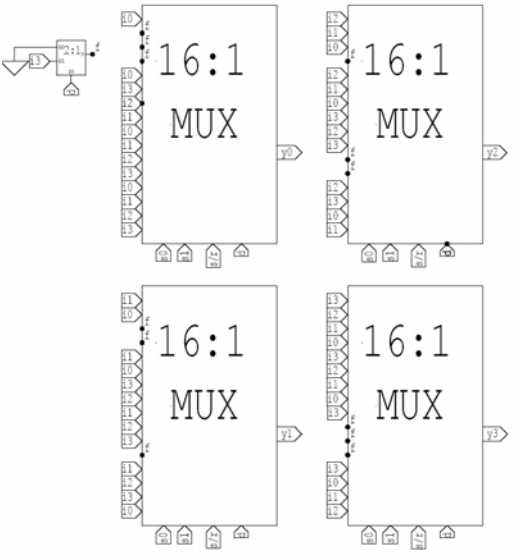


Fig. 1. Schematic diagram of a 4-bit, MUX based barrel shifter

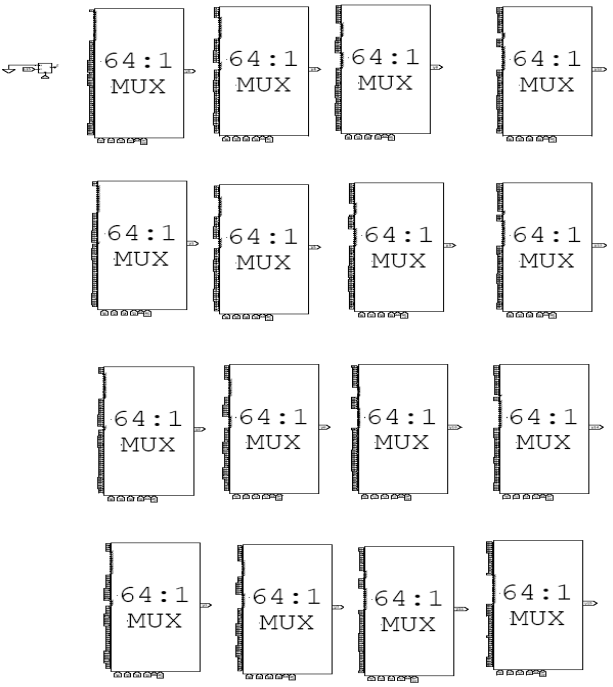


Fig. 2. Schematic diagram of a 16-bit, MUX based barrel shifter

4 Design of a 16-Bit Barrel Shifter

Similar technique is followed in design of 16-bit barrel shifter circuit design. As there are six control inputs, we need a 64:1 multiplexer for each output bit. Thus for 16 output bits, we need 16, 64:1 MUX in the design of a 16-bit barrel shifter, in addition we need a 2:1 MUX for fill bit. Thus total 2:1 MUX required in design are 1009. Since each row of a truth table is implemented using a dedicated 64:1 MUX circuit, therefore, therefore delay of six, 2:1 MUX cell will be required for it; one additional delay will be required for fill bit therefore seven 2:1 MUX cell delay are required to produce final output for 16-bit barrel shifter. The schematic of 16-bit barrel shifter circuit is shown in figure 2. The critical delay of the circuit is 7 times the delay of a 2:1 MUX cell, without considering the interconnect delay. This architecture offers the important architectural properties like regularity, hierarchy, modularity and consistency. The 16-bit barrel shifter design has six external control signals out of which first four bits represent the length of the operation, while fifth and sixth bits represent 'type' and 'direction' of operation respectively. Length bits S [3:0], which allow shifting or rotation of data from 0 to 15 positions. S/R bit decides shifting or rotation (S/R=0 means shifting and S/R=1 means rotation) and D bit decides direction of shift/rotation (D=0 means leftward and D=1 means rightward direction).

5 Physical Implementation and Result

Layout for a 16-bit barrel shifter schematic circuit shown in figure 2 is implemented in 0.6µm, N- in well CMOS process (SCN_SUBM, lambda=0.3) of MOSIS, using three different logic design styles, namely, optimized static CMOS, TG CMOS and dual rail domino CMOS logic.

A schematic library consisting of single functional cell (i.e. 2:1 MUX) is individually defined for optimized static CMOS, TG and dual rail domino logic design styles. Figure 3 and figure 4 shows the 2:1 MUX cell schematic in optimized static CMOS logic and TG logic respectively. A dual-rail domino 2:1 MUX is shown in figure 5. Corresponding to the schematic library, three different physical libraries were designed using optimized static CMOS logic, TG logic and dual rail domino logic design styles using the design principles of [3], [4], [5], [6], [7], [8], [9]. Three different versions of each physical library were developed by respectively sizing the W/L ratios of the NMOS transistor to values of 3, 5 and 7 (W/L values smaller than 3 were also experimented with but not considered further as they resulted in parasitic dominated slower speeds due to weak drives of transistors and were not considered good candidates for high performance). The layout assemblies for the 16-bit barrel shifter were carried out using these cell libraries and automatic place and route tool LEDIT (SPR) from M/s Tanner Research Inc.

The product of average switching energy and circuit delay was then computed for each implementation of the barrel shifter using the same logic design style but utilizing three different physical libraries differing in their transistor sizes as described above. It was noticed that for all the three logic design styles, the physical library utilizing W/L ratio of 3 for NMOS transistor gave the smallest average switching energy-delay product. The generated layouts were simulated after parasitic extraction using circuit simulator, ELDO spice. Supply voltage V_{DD} is kept at 3.3V. The results of VLSI circuit implementation using optimized static CMOS logic design style are then compared with other high speed 16-bit barrel shifter implementation designed using domino and D^3L logic by Ramin Rafati [10] as shown in table 2. Comparing these two-barrel shifter architectures (at 10MHz clock rate) shows that proposed barrel shifter architecture shows almost 50% reduction in propagation delay, while maintain the similar average power consumption. The improvement in delay is obtained due to its shorter gate delay critical path and key architectural properties like regularity, hierarchy, modularity and consistency. These properties improve the wiring delays on critical path in sub micron/deep submicron technologies.

A further inter comparison of implementations using three different logic design styles utilizing the corresponding libraries with NMOS transistor sizing of 3 is shown in table 3 and table 4. The comparison of parameters for three implementations using three different logic design styles carried out for propagation delay, average power (at 10MHz), maximum power and leakage power is shown table 3. Transistor count, core area and total routing length and number of vias are shown in table 4.

Delay comparison for three logic design style shows that implementation using TG logic design style shows reduction in propagation delay by a factor of 1.61 as compared to optimized static CMOS implementation and by a factor of 3.12 as compared to D^3L logic. The reduction in delay for TG logic design style is mainly due to smaller critical path and decrease in routing length. The dual rail domino implementation is much slower as compared to optimized static CMOS implementation due to large increase in routing length (due increased routing complexity), use of simpler logic cells, use of one additional transistor in evaluate path of logic cell and use of weak pull up transistor, which increases the contention current during evaluation. Average power consumption in TG is reduced approximately by a factor of 3.1 as compared to optimized static CMOS implementation and by a factor of 2.3 as compared to D^3L logic due to reduced switching nodes and their capacitance. The dual rail domino implementation show much higher power consumption due to higher switching activity than in equivalent optimized static CMOS gate because all the domino nodes are pre-charged to V_{DD} during each clock cycle. Table 4 shows that the barrel shifter implementation using TG logic design style requires least transistor count, least total routing length and therefore requires smallest core layout area as compared to VLSI implementations using optimized static CMOS and dual rail domino logic design styles.

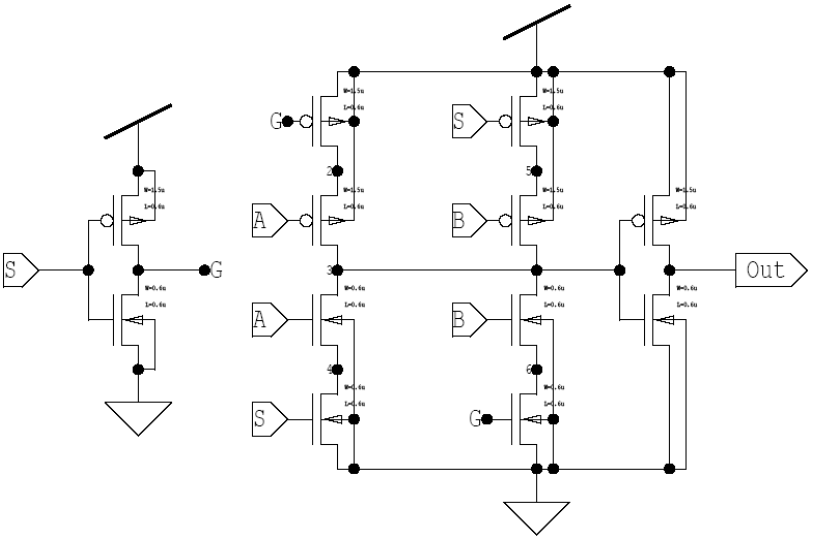


Fig. 3. A schematic of 2:1 MUX cell using optimized static CMOS logic

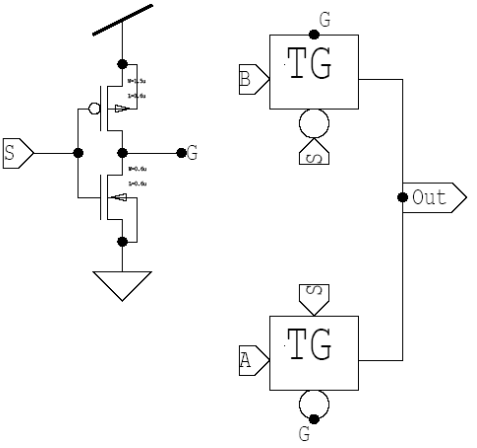


Fig. 4. A schematic of 2:1 MUX cell using TG CMOS logic

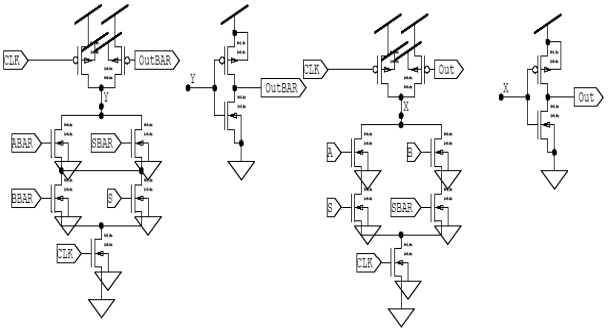


Fig. 5. A schematic of 2:1 MUX cell using dual rail domino CMOS logic

Table 2. Comparison of barrel shifter architectures

Algorithm (technology)	V_{DD} (V)	Propagation delay (τ) ns	Average power (mW)
Proposed - optimized static CMOS (0.6μm)	3.3	1.163	3.59
Reference [10] D^3L (0.6 μm)	5	2.25	2.66
Reference [10] Domino (0.6 μm)	5	2.68	3.23

Table 3. Comparison of performance and power for proposed barrel shifters implemented using three different logic design styles

Logic Design Style	Propagation delay (τ) ns	Average power (mW)	Maximum Power (mW)	Leakage Power (nW)
Optimized static CMOS	1.163	3.59	65.87	65.75
TG CMOS	0.721	1.135	276.26	28.78
Dual-rail domino CMOS	4.28	11.13	1665.51	64.34

Table 4. Comparison of VLSI implementation characteristics for barrel shifters Implemented using three different logic design styles

Logic Design Style	Transistor count	Core area (mm^2)	Total routing length (mm)	Number of Via
Optimized static CMOS	12108	28.34	1901.42	4316
TG CMOS	6054	23.38	1732.41	4459
Dual-rail domino CMOS	18162	106.44	8053.4	10192

6 Conclusion

This paper presents design of 16-bit MUX based high speed barrel shifter circuit architectures and their implementations for three different logic design styles, namely, optimized static CMOS, TG CMOS and dual rail domino CMOS logic; using commercially available 0.6 µm, N-well CMOS process. Comparison of these VLSI implementations shows that the barrel shifter implementation using TG logic design style is fastest and consumes lowest power. The barrel shifter implementation using TG logic implementation is most suitable due to its simpler layout characteristics.

References

1. Tharkan, G.M.: A new design of a fast barrel switch network. IEEE Journal of Solid State Circuits 27(2), 217–221 (1992)
2. Pereira, R.: Fully pipelined TSPC barrel shifter for high speed applications. IEEE Journal of Solid State Circuits 30(6), 686–690 (1995)
3. Anis, M., Allam, M., Elmasry, M.: Impact of technology scaling on CMOS logic styles. IEEE Transaction on Circuits and Systems-II, Analog and Digital Signal Processing 49(8), 577–587 (2002)
4. Kang, S.-M., Leblebici, Y.: CMOS digital integrated circuits, 3rd edn. Tata McGraw-Hill Publishers, New Delhi (2003)
5. Weste, N., Eshraghian, K.: Principles of CMOS VLSI design, 2nd edn. Tata McGraw-Hill Publishers, New Delhi (1998)
6. Rabaey, J.M., Chandrakasan, A., Nikolic, B.: Digital integrated circuits, 2nd edn. Prentice–Hall of India Private Limited (2004)
7. Wang, J.-M., Fang, S.-C.: New efficient designs for XOR and XNOR functions on the transistor level. IEEE Journal of Solid State Circuits 29(7) (July 1994)
8. Srivastava, P., Pua, A., Welch, L.: Issues in the design of domino logic circuits. In: Great Lakes Symposium on VLSI, pp. 108–113 (1998)
9. Sachin, M.Z., Sapatnekar, S.: Dual-monotonic domino gate mapping and optimal output phase assignment of domino logic. In: IEEE International Symposium on Circuits and System, vol. 2, pp. 309–312 (2000)
10. Rafati, R., Fakhraie, S.M.: A 16-bit barrel shifter implemented in data driven dynamic logic (D^3L). IEEE Transaction on Circuits and Systems-I 53(10), 2194–2202 (2006)

Research of an Atypical Unexpected Incident in Telecom Complaint Text for 3G

Yang Wen-Chuan, Chen Ning-Jun, and Duan Xiao-Yan*

Beijing University of Posts and Telecommunication, Beijing 100876, China
vermonthmoon@hotmail.com

Abstract. Accompany with the Telecom new services in 3G, a kind of atypical unexpected incident hides in 3G complaint text. The incident has small probability and atypical attributes, and hard to understand and unobvious by ordinary way. We define this atypical unexpected incident as AUI. This paper does some research for AUI, and it helps to forecast and alert the service problems. We construct the mathematical model for AUI, design the extracting algorithm, and try to verify its efficiency with complaint text. The experimental results map out the correct possibilities for our algorithm.

Keywords: Complaint text, Atypical Unexpected Incident, Association rules, Data Mining.

1 Introduction

Recently, we find a kind of atypical unexpected incident hides in complaint texts. We define this atypical unexpected incident as AUI. The AUI complaint texts have some special attributes, such as low probability, territoriality, outburst, concealment, infectivity, unknown incentives, and unobvious by ordinary way.

For example, some urban companies find that they have to pay heavy cellular roaming charges for customers among different carriers in May. When they urgently shut down some of the prepaid phone card, complaints appears and become the regional focus.

Since the small probabilities for AUI, telecom companies usually skip it or deal it in an ordinary way. This can result in some mass incidents and bring baneful influence to the current services, which lead to massive customer churn in the fierce communication competitions. This paper does some research for *AUI*, and it helps to do business problems pre-warning and forecasting. We construct the mathematical model for AUI, design the extracting algorithm, and try to verify its efficiency with complaint data. The experimental results map out the correct possibilities for our algorithm.

2 Difference between the Typical and the Atypical Incident

The typicality incident appears in high frequency, which is general in all area. AUI, appearing small and regionally, is not easy to find.

* Corresponding author.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 647–652.
springerlink.com

The inducement of the typicality incident is certain and can be solved in a mature way.

The inducement of AUI is uncertain. There is no mature way to solve it.

The typicality incident provides the guidance to solve regular question. And AUI reflects the breakdown of one particular business, which contributes to improve and perfect it.

The probability of the typicality incident is stable, but the appearance of AUI is sudden. The appearance time of AUI is positive related to the development of the business.

Table 1. Difference between the typical and atypical incident

	typical incident	*atypical incident*
probability	*high and stable*	*low and gusty*
frequency	*high, general in all area*	*low, regionalism, hardly found*
precipitating factor	*certain*	*uncertain*
effect	1 *the regular feedback of business operation situation*	1 *the expression of a breakdown belonging to one particular business*
	2 *the guidance to solve regular question.*	2 *contribute to improve and perfect the business*
Solution	*mature*	*immature*

3 Cyclic Variation Rate(Abbr. for Rate)

Rate for *AUI*：Ratio of increment $m_{A[q]}$ and $m_{A[q],t-1}$, see below：

$$Rate=\frac{m_{A,t}-m_{A,t-1}}{m_{A,t-1}},t\in T \tag{1}$$

Here, *t* is Time scale. The minimal *Rate* is $Rate_{min}$.

Theorem 1 **Singleton Increase**：

Let $Complain_{AUI[q]}=\{Complain_1,\ldots,Complain_i\}$ is the AUI Complain Set，q is AUI Number, $1\leq q\leq m$. Assume $m_{A[q]}$ for $Complain_{AUI}$, and $m_{A[q],t}$ for t-day($t\in T$)，$t_1,t_2\in T$，if $t_1<t_2$, then:

$$m_{A[q],t1}<m_{A[q],t2} \tag{2}$$

Prove: It can be proved by the definition and attribution of *AUI* and *Rate*.

Anyway, under circumstance $t_1<t_3$ if we have $m_{A[q],t1}\approx m_{A[q],t2}$，it is not a *AUI*.

Theorem 2 Set Increase:

Let $AUISet_t = \{Complain_{AUI[1]},\ldots,Complain_{AUI[q]}\}$ is the t-day AUIs Set, $1 \le q \le m$. Assume $t_1, t_2 \in T$, if $t_1 < t_2$, then :

$$m_{AUISet_{t1},t1} < m_{AUISet_{t2},t2} \tag{3}$$

Prove: It can be proved *Theorem 1*.

$$m_{A[u],t1} < m_{A[u],t2},\ 1 \le u \le q$$

$$\therefore \sum_{i=1}^{q} m_{A[i],t1} < \sum_{i=1}^{q} m_{A[i],t2}$$

Based on the inner connection between $AUISet_t$ and its m, we can deduce the following equation :

$$m_{AUISet_t,t} = \sum_{i=1}^{q} m_{A[i],t}$$

$$\therefore m_{AUISet_{t1},t1} < m_{AUISet_2,t2}$$

When we do cyclic variation analysis among daily reports, if a suspect incident coincides the *Theorem* 1 and 2, it is identified as *AUI*.

4 Mathematical Model

4.1 State Base Transition

Based on the mathematic model for AUI, we analyze the Complains with multi-dimensional association rule, and obtain the corresponding candidate SFeature. We can estimate the Cyclic Variation Rate, and find the suspect AUI sets. These candidate sets are the foundation and data source for further extraction and analysis of AUI.

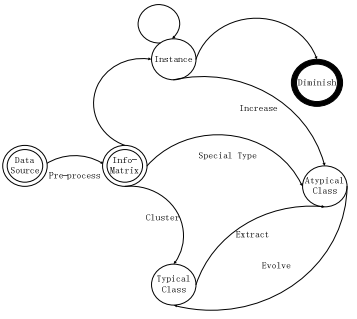


Fig. 2. Verification and classification diagram for the *Complain*

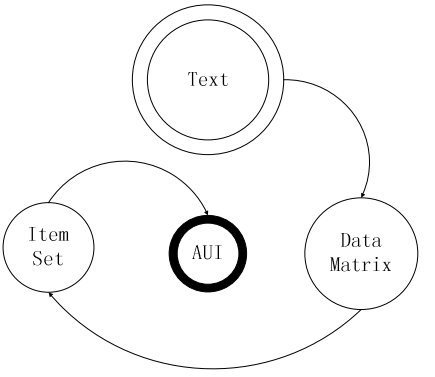


Fig. 3. Diagram of transition among different *Complain* state

The state transition for incident verification and classification is shown as Fig1.

Figure2 shows the data state transition among different Complain state from Group to AUI. The capacity of data is gradatim diminished.

4.2 Algorithm with Multi-dimensional Association Rule

The following is an ***algorithm*** of mining $S_{Feature}$ with multi-dimensional association rule based on *Feature Item Set*.

Input: *GcF*, $Complain_i$ and its CF_i $(1 \le i \le m)$,*Vector-Axis Matrix*, *minsup*, *maxsup*, *minconf*, $Rate_{min}$, and K ;

Output: Set of *Multi-dimensional association rule (Q), Set of* $S_{Feature}$ *(R). R* is corresponding with *Q*

Steps for mining

Step1 $R=\varnothing$, $Q=\varnothing$。

Step2 Input the *GcF*, and construct the $Space_{CF}$ through $Mapping_{CF}$.

Step3 Input the *Vector-Axis Matrix*, and use the Genetic Algorithm to select the *K-clustering* centers, which are the terminal points of *Vectors* belonging to the *Matrix* in $Space_{CF}$.

Step4 With the *K-clustering* centers, we use the *K-means algorithm* to find *K sets*(Set_p, $1 \le p \le K$) in *Vector-Axis Matrix*.

Step5 for (int p=1;p<=K;p++){

/* By using FP-Growth algorithm, we try to find all candidate multi-dimensional association rules in Set p which fit for $minsup \le Support \le maxsup$ & $Confidence \ge minconf$. Then we add them into the candidate set L_C . C is the amount of candidate association rules. The number for *Vectors* which fit for the *No.c* rule is noted as m_c, $1 \le c \le C$ */

Set_p——FP-growth *mining*（L_c、*minsup*、*maxsup*、*minconf*）;

/*Put the l_c with the maximal m_c in L_C to Q* /

If (m_c>=$m_{anything\ else}$ in L_C) $Q = Q+l_c$;

/*According to the $Mapping_{CF}$= {$Item_1 \leftrightarrow x_1$,..., $Item_j \leftrightarrow x_j$}, translate the prior and rear of l_c to *Item Set*, $1 \le j \le n_f$ * /

If(l_c=($A' \Rightarrow B'$)&& A' =[x_{k1}, x_{kn}]&& B' =[x_{l1}, x_{ln}]){//$1 \le k1,kn \le n_f$, $1 \le l1,lm \le n_f$

A=[$Item_{k1}$, $Item_{kn}$] ;

B=[$Item_{l1}$, $Item_{lm}$] ;

$R = R + (A \cup B)$;

}

}

/* constitute the association rule Q from R * /

Step6 for (int ct=1; ct<= *R.length*; ct++){

/* Calculate the corresponding $m_{A[}r_{s]}$ of the $S_{Feature}$ $r_s(1 \le s \le R.length)$ in R * /

int $m_{A[}r_{s]}$ =0;

for (int ct1=1;ct1<=q.length;ct1++){

/*$1 \le q \le K$, q is the clustering class for r_s, and calculate the similarity of r_s and *Vector* in q */

$\overline{Vector_{ct1}} = (\overline{x_1}, \overline{x_2}, ..., \overline{x_{n_f}})$ // from Formula26

If $\left(SIM(\overline{Vector_{ct1}}, r_s) = \frac{\overline{Vector_{ct1}} \bullet (r_s)}{\left|\overline{Vector_{ct1}}\right| \left|(r_s)\right|} = 1 \right)$ $m_{A[}r_{s]}$ ++ ;

}

$r_s = r_s$: $m_{A[}r_{s]}$;

}

Update R ;

Step7 Export one-day Q and R.

Step8 *For R*, Q, R_t and Q_t in time variant t, $t \in [t_{d1}, t_{dx}] \in T$. We choose the union set of R_t and put them into $Candidate_K$, and each one is noted as c_k. Its corresponding rule is put into $Qcandidate_K$, and noted as q_k, $K=Candidate.length$, $1 \le k \le K$.

Step9 $R=\varnothing$, $Q=\varnothing$。

Step10 for (int k=1;k<=K;K++){

int *amount* = 0 ;

for (int *data*=2;*data*<=x;*data*++){

If $(Rate \ge Rate_{\min})$ *amount*++ ; // calculation for *Rate* see *Lemma 5*

}

If (*amount* == x-1) {

$R = R + c_k$;

$Q = Q + q_k$;

}

}

Step11 Export the phases Q and R.

5 Conclusions

This paper does some research on a kind of atypical unexpected incident hidesin telecom complaint text. With the dynamic base state method, we accurately describe the state evolution for AUI. We construct the mathematical model for AUI, design the extracting algorithm. The research for AUI in telecom complaint text is useful. It helps the telecom company improve the quality of service, avoid the customer churn and forecast the regional unexpected incident.

We will do some further research on the following topics as below. The first t hing is to optimize the de-noising algorithm. And, we hope to get the feedback fr om the end users in application.

References

1. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large database. In: Proc. of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207–216. ACM Press, New York (1993)
2. Knowles, J.: Feature subset selection in unsupervised learning via multiobjective optimization. Int'l J.Computational Intelligence Research 2(3), 217–238 (2006)
3. Saeys, Y., Abeel, T., Van de Peer, Y.: Robust Feature Selection Using Ensemble Feature Selection Techniques. In: Daelemans, W., Goethals, B., Morik, K. (eds.) ECML PKDD 2008, Part II. LNCS (LNAI), vol. 5212, pp. 313–325. Springer, Heidelberg (2008)
4. Dash, M., Liu, H.: Handling large unsupervised data via dimensionality reduction [3]; Zhang, L.-X.,Wang, J.-Q., Zhao, Y.-N., et al.: Feature selection in machine learning. Computer Science 31(11) 180–184 (2004)
5. Hand, D., Mannila, H., Smyth, P.: Principles of data mining. The MIT Press (2001)
6. Jouve, P.-E., Nicoloyannis, N.: A Filter Feature Selection Method for Clustering. In: Hacid, M.-S., Murray, N.V., Raś, Z.W., Tsumoto, S. (eds.) ISMIS 2005. LNCS (LNAI), vol. 3488, pp. 583–593. Springer, Heidelberg (2005)
7. Silva, H., Fred, A.: Feature Subspace Ensembles: A Parallel Classifier Combination Scheme Using Feature Selection. In: Haindl, M., Kittler, J., Roli, F. (eds.) MCS 2007. LNCS, vol. 4472, pp. 261–270. Springer, Heidelberg (2007)
8. Opitz, D.: Feature selection for ensembles. In: 16th National Conference on Artificial Intelligence, pp. 379–384. AAAI (1999)
9. Ren, J., Qiu, Z., Fan, W., Cheng, H., Yu, P.S.: Forward Semi-supervised Feature Selection. In: Washio, T., Suzuki, E., Ting, K.M., Inokuchi, A. (eds.) PAKDD 2008. LNCS (LNAI), vol. 5012, pp. 970–976. Springer, Heidelberg (2008)
10. Skurichina, M., Duin, R.P.W.: Combining Feature Subsets in Feature Selection. In: Oza, N.C., Polikar, R., Kittler, J., Roli, F. (eds.) MCS 2005. LNCS, vol. 3541, pp. 165–175. Springer, Heidelberg (2005)

Research of a Multi-agent Algorithm to Process the Evolutive Information

Wenchuan Yang[1], Bei Jia[2], and Xiaoyan Duan[1]

[1] Beijing University of Post & Telecommunication, Beijing, 100876, P.R. China
[2] Xi'an Communication Institute, Shaanxi, Xi'an, 710106, P.R. China
yangwenchuan@bupt.edu.cn, {jiabeihn,vermonthmoon}@hotmail.com

Abstract. During the data processing based on population data warehouse, a large scale of evolutive information was produced accompany the data mining and knowledge discovery, which is very important to the final rule. This paper introduces the research of a multi-agent algorithm to process the evolutive information in data mining on population data warehousing, also gives the mathematic model and corresponding rule for it, which include data extraction and check-up, data normalization, and a mechanism to support market analysis and forecast. We also summarize some basic requirements and data model constructing for applying data warehouse in population fields.

Keywords: Data Mining, Evolutive Information, Multi-agent System.

1 Introduction

As defined by W. H. Inmon[1], the Data Warehouse is a subject-oriented, integrated, non-volatile, time- accumulate data set, which is fit for decision. The Statistic System is an important national department, which is responsible for the macro-decision of the government and analyzing of predict to provide the data support. These statistic data have variety sources, including the professional data, detailed statistic data, and local data about the economic circumstance. In Statistic Data Warehouse (SDW), we use the Data Warehouse technique to resolve these problems.

The statistics system is changed every year, and the data continuously increased time by time. This induces the inconsistent of the data contents. The Statistic System requires the analysis and forecast system is independent with the business data.

Data Mining is a subsystem of the SDW system. Data Mining can provide the analysis of the macro economy, the estimate of the market development trend, and various forms of the display, provide the support for decision.

In the researching and application of the SDW Data mining subsystem, we focused on analyzing the characteristic of the statistic data during data mining, and we use the data warehouse technique to character the analysis statistic data. In the process of the Data Mining and knowledge discovering, we adjusted the algorithm method in order to discover lots of factors. These factors infect the knowledge-discovering schedule. We implement a specific statistic application and acquire some important knowledge. In the researching and application process of the SDW subsystem, we discovered lots

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 653–660.
springerlink.com

of information, which have the temporal characteristic to accompany the Data Mining processing.

2 The Multi-agent Data Warehouse

The Multi-agent Data Warehouse [2] consists mainly of: ETL (Extract, Translate, Loading) module; Data warehouse internal organization management module; OLAP and data query module; decision analysis and system method base module.

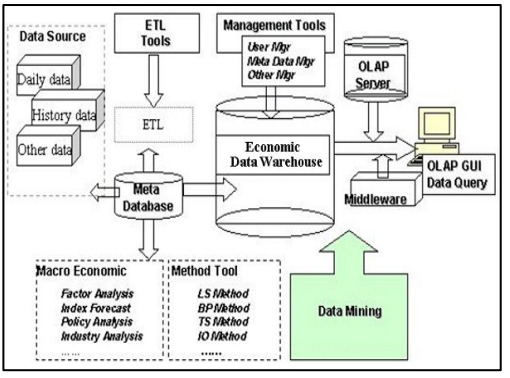
The ETL module is responsible for organizes to convert the warehouse data from the source database, completing the data collection and refreshment. In order to assure the data consistency, we need to use knowledge and rule to verify raw data. Only corrected data can be pumped into corresponding position in data warehouse according to time and subject.

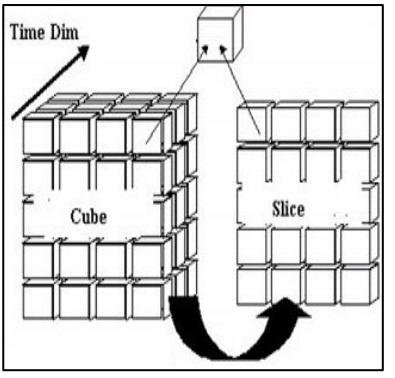
The data warehouse module is main consists of Meta database and modules which is responsible for internal control and management, together with some data mart. The information in data mart is organized as star-schema model and special subject-oriented, which consist of fact table (to store agent information data) and dimension table (to store dimension data).

The information in data mart makes the subject more definitude and analysis more specialty. Meta database is a special base to store data that is responsible for controlling other data.

OLAP GUI, which is directly used by users of Decision Support System. Its main function is to transfer the relation database table (inside data warehouse, especially some structured storage in data mart) to multi-dimension data cube [3] according to the user decision subject. It also presents the user with OLAP GUI, which is a slice from the multi-dimension cube, to do data query and presentation.

The System architecture is shown as figure 1.





3 Mathematics Model

Data Mining is a focused on finding out useful information from various data with none-trivial methods. Data Mining itself is a hybrid research subject, and is influenced by many subjects. It also depended on the method or special application of data mining. The relationship between Data Mining and Data Warehouse is lose-coupling.

Data Warehouse provides the data source for Data Mining[4], while Data Mining find out knowledge form Data Warehouse. However, both of them can be independent to each other. With the subjects in Data Mining, the designers have introduced their knowledge and model into the Data Warehouse.

The Data Warehouse increases to a certain size as time goes by. OLAP play an important role to validate the existing model with the various data provided by the Warehouse. And Data Mining is focused on provide the leaders with interrelated knowledge.

We will process lots of original and middle data besides some commonly special knowledge during the knowledge discovering. In our daily work, we find that: The mid data is important in knowledge discovering when properly used and processed. Now we will define the evolutive information first, and then define and introduce evolutive information processing[5].

When processing the statistic data, we define the various mid-information with temporal character ,such as evolutive information. The mathematic description of evolutive information is as following:

Definition1. Evolutive Information (EI): EI is a small cuboid in the evolutes space (S), which is uniquely defined by time scale, evolution function etc dimensions.

The n+1dimension evolutive information space is defined as $S=G*R$.

R stands for evolutes domain. Domain $R=R1\times R2\times\ldots\times Rn$, where each Ri $(1\le i\le n)$means the index item in the statistic.

G is defined as evolutes interval G=*[g1,gr]*,where gi(i∈natural number)stands for the minimum time that the system can identify(called time scale in this paper).

The time scale means the span in the time dimension during the evolutive processing.

Definition2. Potential. Potential is a non-negative real function. It maps each space for cuboid or cuboid set to a non-negative[6] real number. Potential use δ or ρ denote, the potential is directly called as function.

Definition3. Cluster. Cluster is a maximal undirected complete set. The maximal complete set means all the nodes in the set are connected. As for cluster C, the cuboid set, which is consisted of cuboid in the cluster, is denoted with V_C.

The probability function δ_C and utility function ρ_C in cluster C are respectively described as:

$$\delta_C = \prod_{X\in V_C} P(X \mid pa(X)),\ \rho_C = \sum_{X_i\in V_C} \rho_{X_i} \tag{1}$$

One cluster without variable, its junction probability function is unit function, its junction utility function is null.

Definition4. Evolutive function is defined as the global function which is mapped form n+l dimension evolutive Information domain *S* to domain *R*, and marked as *F(s)*(the elaborate definition and mathematic description will be given later).

Evolutive Information is vibrant during Data Mining, which is represented as the changing processing of scale data in the time dimension. The evolutive information processing accompanied with various temporal character. It will be great application values for discovering new knowledge that analysis these mid-processing evolutive information instantly, and selecting the best to the modified parameter in Data Mining. Therefore, the processing of evolutive information becomes an important investigated issue in the Data Mining in Data Warehouse.

Referring to the temporal character of evolutive Information and the special demand of Statistic System, we provide an additional method that describes how to slices the evolutive Information in the time dimension.

Definition5. Evolution time slice is defined as (g,s1,s2,…,sn)∈R, where the time scale g∈G, and n dimension (s1,s2,…,sn)∈R.

The global function mapping from n + 1 dimension evolutive information dimension S to evolutes index domain R is defined as evolutive function. If and only if the following conditions are satisfied, that means when the time scale increases[7], the entropy(ξ) in the evolution time slice co-increase.

The strict mathematic description of evolutive function is as follow:

For any g1,g2∈G and g1<g2, the entropyξ of the contain evolution time slice presents increase by degree, that is ,

$$\xi 1 \leq \xi 2 . \xi i = F(gi, P). \tag{2}$$

P can take any value in evolutes index domain, that is, P=(s1,s2,…,sn)∈R.

If g1＜g2 containsξ1＝ξ2, we defined F as static evolutive function.

The static evolutive function can describe the stable phenomenon of the statistic data in a certain area. That means, in the processing of static evolutive function, the evolutive information obtained at some moment will remain stable in the subsequent process.

4 Data Processing

The processing of evolutive information is to settle[8] and handle all the evolutive information in the evolutes space, and instantly modify the Data Mining method with this information, hence accumulate the process of discovering knowledge correctly.

There are lots of index data in Data Warehouse Statistic, for a period of time, the key point of the evolutes space S. Evolutive information process actualizing is to slice the evolutive process in to partial evolution process with slicing-space-domain method, and to complete the global evolutes process distributed. Its corresponding formalized describe as follows:

If exist a limited division of evolutive information space S, satisfied $S=S_1 \cup S_2 \cup \ldots \cup S_k$, then we called evolutive function F the partial evolutive function, the evolutes process of which is called evolution processing.

We will propose and proof the integrity of the partial evolution in the following.

Theorem of the integrity for partial evolution: If we can divide the global space S into limited, equal and partial evolution space, we can get the complete and overall evolution space by processing of each partial evolution space.

Proof: according to the definition of partial evolutive function , S can be decomposed into finite subclass: $S_1,S_2,\ldots,S_k$. Since each Si fits for function F with monotonous and addition. In each Si, we embed an evolutive information EIj ($1 \leq j \leq n$). According to the choice axiom, we can choose the minimum one from the $EI_1, EI_2, \ldots, EI_j$. Let's suppose it is Eim. Then we select it as the basis unit to re-divide the set which will cover S. As for the smallest property of Eim, we know the partial evolution F is complete.

Through the theorem above, it can induce monotonous and can adding of the partial evolution, and induce the actualizing method of evolutive information processing too.

During the research and implementation of knowledge discovering in SDW, we try to find the character of the statistic data and the special demand of the government decision-making.

Also we adopt the evolutive information processing method in the Data Mining System over Data Warehouse. The main work is to divide the evolutes space into equal partial domain with Object Oriented method.

Each domain is an EI cuboid. With the coordination of multiple partial evolutive information space through messages, we implement a distributed parallel evolutive algorithm. Based on the monotonous and addition of evolutive information, we only reserve increment informationΔEI. This improves the calculating efficiency during cumulating knowledge synchronously. It also provides the system with the capability of processing the various evolutive informations over statistic data.

During the research and implementation of Statistic Data Warehouse, we get the notable and successful result with the implementation of static evolutive function in statistic. In a certain range, lots of statistic data is in stable status. This phenomenon is very suitable to be processed with the static evolutive function. For example, the area of some district, the nation component, the fund in some companies, and so on. This information does little contribution to knowledge discovering, which is something like the constant variant in the advanced program. This information can be eliminated from data source by modifying parameter and arithmetic during the Data Mining.

5 A Trivial Algorithm

Anyway, when we process data mining over data warehouse, we need carefully consider about the evolutive information. Following is the trivial algorithm for slice to process initial cube at a temporal time.

Trivial Algorithm 1. The structure of the slice, which is transformed from the cuboid, can be define as $S=(\alpha,\beta)$, α is a cluster set. The two clusters in α are connected by cuboid in β. Assume any pair of neighbor clusters C_i and C_j, $C_i \in \alpha, C_j \in \alpha, S_k \in \beta$. S_k is the split dimension between C_i and C_j, so we have

$$V_{S_k} = V_{C_i} \cap V_{C_j} \tag{3}$$

The consistent of slice S stands for that: in cluster S_k ,we have：

$$\sum_{V_{C_i} \setminus V_{S_k}} \delta_{C_i} = \sum_{V_{C_j} \setminus V_{S_k}} \delta_{C_j} \tag{4}$$

For each cluster C of α and each cluster S for β, we have respectively probability function δ_C, δ_S and utility function ρ_C, ρ_S. The junction probability function δ_C and junction utility function ψ for the slice S are ：

$$\delta_T = \frac{\prod_{C\in\Gamma}\delta_C}{\prod_{S\in\Delta}\delta_S} \quad \text{and} \quad \rho = \sum_{C\in\Gamma}\rho_C \tag{5}$$

Trivial Algorithm 2： Cluster splitting and average algorithm[9]. For one possible split set CUT(S) in slice $S=(\alpha,\beta)$, assume u and v are cuboid set . Let's combine them to an ordered 2-tuple {u,v}, then the element in CUT(S) is consist of these 2-tuple , each element is fully belong to one cluster ：

$$\mathrm{CUT}(T) = \{\{u,v\} \mid \exists C \in \alpha.\{u,v\} \subseteq V_C\}. \tag{6}$$

The element in CUT(S) set is produced by the element number in cluster of the slice S, under the direction of strategic relevance and weakly relevance.

One executive splitting $\{u,v\} \in \mathrm{CUT}(T)$ in $S=(\alpha,\beta)$, which removes one cluster C $\{u,v\} \subseteq V_C$ from S, but add two cluster C_u and C_v. So we have $V_{C_u} = V_C \setminus \{v\}$ and $V_{C_v} = V_C \setminus \{u\}$, also we can add one new split S, with $V_S = V_C \setminus \{u,v\}$. Based on strategic relevance, this makes the position of C_u and C_v in slice S fulfilling partial order relationship, and connected through S[10]. If $u \in V_D$, let's move the cuboid in splitting cluster of D, which is neighbor to C, to C_u. Also do same processing to C_v. This is also called the average algorithm.

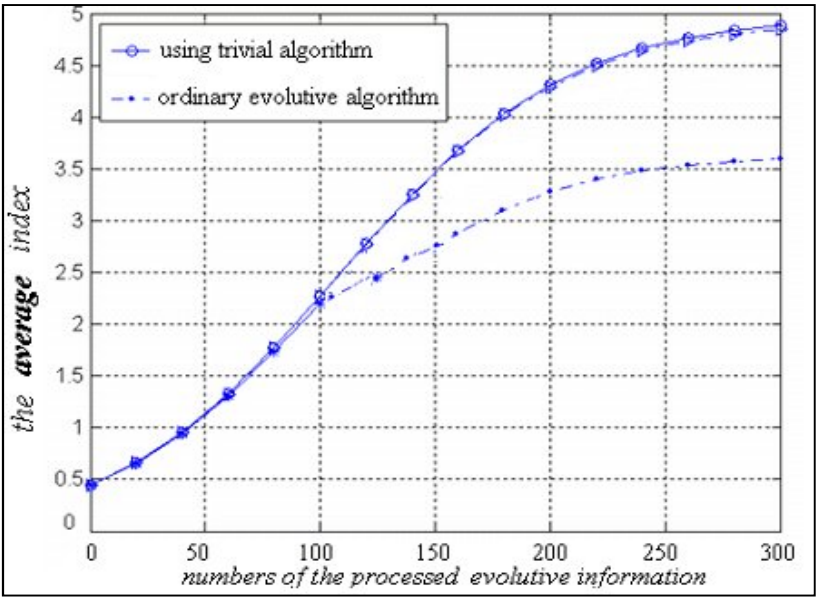


Fig. 2. Performance between the two Evolutive Algorithms

The probability function and utility function for new cluster are respectively

$$\phi_{C_u} = average(C, v, \delta), \quad \rho_{C_u} = average(C, v, \rho) \tag{7}$$

and

$$\delta_{C_v} = average(C, u, \delta), \quad \rho_{C_u} = average(C, u, \rho). \tag{8}$$

The probability function and utility function for new split cluster are respectively

$$\delta_S = average(C, (v, u), \delta), \quad \rho_S = average(C, (u, v), \rho). \tag{9}$$

The slice, which we get through the above process, is called as splitting slice.

Fig2 is the comparison of the performance between the ordinary and trivial evolutive algorithm[11]. We can find that, the performance of the processing with ordinary algorithm is almost the same, but it is greatly improved by using the trivial algorithm.

6 Conclusions

The evolutive information with temporal character, which is concomitant with Data Mining process, has great application values for adjusting Data Mining method[12] and accelerating discovering knowledge. With precision mathematic description and

the proof, this paper proposes the actualized rules for the evolution process, and we also prove its validation in the Data Mining over SDW.

References

1. Inmon, W.H.: Building the Data Warehouse, 3rd edn. John Wiley (2001)
2. Samos, J., Saltor, F., Sistac, J., Bardés, A.: Database Architecture for Data Warehousing: An Evolutionary Approach. In: Quirchmayr, G., Bench-Capon, T.J.M., Schweighofer, E. (eds.) DEXA 1998. LNCS, vol. 1460, pp. 746–756. Springer, Heidelberg (1998)
3. Inmon, W.H.: Information architecture for the 90's: legacy systems, operational data store, data warehouse. PRISM Tech. Topic 1(13) (2000)
4. Inmon, W.H.: The operational data store. PRISM Tech. Topic, 1(17) (2003)
5. Valides-Perez, P.: Principles of human-computer collaboration for knowledge-discovery in science. Artificial Intelligence 107, 335–346 (2001)
6. Widrow, B., Rumelhart, D.E., Lehr, M.A.: Neural networks: Application in industry, business and science. Communication of ACM 37, 93–105 (2004)
7. Wang, R., Storey, V., Firth, C.: A framework for analysis of data quality research. IEEE Trans. Knowledge and Data Engineering 7, 623–640 (2005)
8. Birman, A.: Computing Approximate Congestion Probabilities for a Class of All-Optical Networks. IEEE Journal on Selected Areas in Communications (2006)
9. Chung, S., Kashper, A., Ross, K.W.: Computing approximate congestion probabilities for large loss networks with state-dependent routing. IEEE/ACM Trans. Networking (2003)
10. Sridharan, A.: Congestion in All-Optical Networks. IEEE/ACM Transactions on Networking 12(2) (April 2007)
11. Shortle, J.F.: Dynamic Call-Congestion Algorithms for Telecommunications Networks. IEEE Transactions on Communications 51(5) (May 2008)
12. Romanow, A.: Dynamics of TCP traffic over ATM networks

An Algorithm for Sparse Reconstruction Based on the Relevance Vector Machine

Tianbao Dong, Jingshu Yang, Yingke Lei

Lab 702, Electronic Engineering Institute,
230037 Hefei, China
dtb_1@163.com

Abstract. Sparse reconstruction based on the relevance vector machine (RVM) in underdetermined blind source separation (UBSS) was discussed in the paper. The Laplace Priors were modeled using a hierarchical form in the paper. The posterior over the estimated parameters was formulated based on Bayesian principle. Based on the hierarchical priors, an iterative algorithm for sparse reconstruction was derived using MAP criterion. Then a fast iterative algorithm was derived through the analysis of the characteristic of the likelihood. Finally, the fast iterative algorithm was tested by the simulations.

Keywords: The relevance vector machine, Bayesian learning, Underdetermined blind source separation, Sparse reconstruction.

1 Introduction

Underdetermined blind source separation (UBSS) is a topic in recent years. The instantaneous linear model of UBSS is

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \tag{1}$$

Where $\mathbf{x}(t) = [x_1(t), x_2(t), ..., x_M(t)]^T$ is a known $M \times 1$ observation vector at time instant t, $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_N]$ is an unknown $M \times N$ mixing matrix, $\mathbf{s}(t) = [s_1(t), s_2(t), ..., s_N(t)]^T$ is the source vector at time instant t. $\mathbf{n}(t)$ is the vector of white Gaussian noises. The purpose of UBSS is to obtain the mixing matrix and sources from the observations.

Sparse component analysis (SCA) is one of approaches to solving UBSS problem, which utilizes the sparsity of sources in time domain or in transforming domain. SCA usually consists of two steps. First, the mixing matrix is estimated from the observations. Then the sources are reconstructed in the second step. The mixing matrix is usually estimated using clustering method. Sources are estimated mainly based on optimization and matching pursuit, which can be categorized into three groups: greedy algorithms for minimizing l_0-norm, linear programming algorithms for minimizing l_1-norm and statistical optimization algorithms.

The relevance vector machine (RVM) is proposed by Michael E. Tipping in [1], which solves the problem of supervised learning. The paper focuses on the problem of sparse signal reconstruction and implements the sparse signal reconstruction based on RVM, assuming that the mixing matrix has already been estimated.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 661–669.
springerlink.com

2 The Principle of the Relevance Vector Machine

The original purpose of RVM is to solve the problems of regression and classification in supervised learning. Given a set of examples of input vectors and corresponding targets $\left\{(\mathbf{x}_i, t_i) \mid i = 1, \cdots, N; \mathbf{x}_i \in R^n\right\}$, RVM can be used to solve regression problem if the targets are real values, or be used to solve classification problem if the targets are class labels. Similar to Support Vector Machine, RVM also makes predictions based on a function of the form

$$y(\mathbf{x}) = \sum_{n=1}^{N} w_n K(\mathbf{x}, \mathbf{x}_n) + w_0 \tag{2}$$

where $\{w_n\}$ are the model weights and $K(\mathbf{x}, \mathbf{x}_n)$ is a kernel function. The model for RVM is

$$\mathbf{t} = y(\mathbf{x}) + \boldsymbol{\varepsilon} \tag{3}$$

where $\boldsymbol{\varepsilon}$ are independent mean-zero Gaussian noises. RVM performs 'learning' using a particular form of Gaussian parameter prior based on Bayesian learning, which obtains parameter values through maximizing the marginal likelihood of the data with respect to hyperparameters.

3 Algorithm for Sparse Signal Reconstruction

We observe that model (1) has similar forms to model (3), and the kernel function of RVM corresponds to the mixing matrix in UBSS. The sparsity of model (3) is determined by the weights, so RVM can be utilized to solve the problem of UBSS.

3.1 Hierarchical Sparsity Prior

Gaussian prior is adopted for RVM and sparsity of weights can be obtained through Bayesian learning. In [2], Student-t distribution prior is used to model sparsity of weights in a hierarchical form, which leads to less sparse solution than RVM. Laplace prior is usually adopted In SCA, which leads to less sparse solution than student-t prior. Laplace prior is also can be obtained in a hierarchical form. We assume that sources in (1) are independent and a Gaussian prior over the sources is given

$$p(\mathbf{s} \mid \boldsymbol{\alpha}) = \prod_{i=1}^{N} N(0, \alpha_i) \tag{4}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_N)$ with α_i the variance of the ith source. An exponential prior is considered over $\boldsymbol{\alpha}$

$$p(\boldsymbol{\alpha} \mid \lambda) = \prod_{i=1}^{N} \Gamma(1, \lambda/2) = \prod_{i=1}^{N} \frac{\lambda}{2} e^{-\frac{\lambda \alpha_i}{2}}, \quad \alpha_i \geq 0, \lambda \geq 0 \tag{5}$$

Then prior for **s** can be reformulated by means of integrating out $\boldsymbol{\alpha}$ [3]

$$\begin{aligned} p(\mathbf{s} \mid \lambda) &= \int_0^\infty p(\mathbf{s} \mid \boldsymbol{\alpha})\, p(\boldsymbol{\alpha} \mid \lambda) d\boldsymbol{\alpha} \\ &= \prod_{i=1}^{N} \frac{\sqrt{\lambda}}{2} e^{-\sqrt{\lambda}|s_i|} \frac{\sqrt{\lambda}}{2} e^{-\sqrt{\lambda}|s_i|} \end{aligned} \tag{6}$$

We observe that Laplace prior is obtained from (6).

For convenience, similar to RVM, σ^2 represents the variance of additive noise. Let $\beta^{-1} = \sigma^2$, prior for β is as follow,

$$p(\beta \mid a, b) = \Gamma(a, b) = \frac{b^a}{\Gamma(a)} \beta^{a-1} e^{-b\beta} \tag{7}$$

where $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$.

3.2 Basic Algorithm for Sparse Signal Reconstruction

The noise in (1) is assumed to be independent and Gaussian with mean-zero value. The likelihood of observations can be written as

$$p(\mathbf{x} \mid \mathbf{s}, \beta) = (2\pi\beta^{-1})^{-\frac{M}{2}} \mathrm{e}^{\{-\frac{1}{2\beta^{-1}}(\mathbf{x}-\mathbf{As})^T(\mathbf{x}-\mathbf{As})\}} \tag{8}$$

where β is defined in section 3.1 and *M*, **x** **A** and **s** are defined in (1). The prior for s is obtained according to (4) and (5). Posterior is important in Bayesian learning. The posterior to be defined in this paper is $p(\mathbf{s}, \boldsymbol{\alpha}, \lambda, \beta \mid \mathbf{x})$, which can be defined according to Bayesian rule

$$p(\mathbf{s}, \boldsymbol{\alpha}, \lambda, \beta \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \mathbf{s}, \boldsymbol{\alpha}, \lambda, \beta)\, p(\mathbf{s}, \boldsymbol{\alpha}, \lambda, \beta)}{p(\mathbf{x})} \tag{9}$$

We can not compute the posterior $p(\mathbf{s}, \boldsymbol{\alpha}, \lambda, \beta \mid \mathbf{x})$ in (8) directly. So we rewrite (9) as

$$p(\mathbf{s}, \boldsymbol{\alpha}, \lambda, \beta \mid \mathbf{x}) = p(\mathbf{s} \mid \mathbf{x}, \boldsymbol{\alpha}, \lambda, \beta)\, p(\boldsymbol{\alpha}, \lambda, \beta \mid \mathbf{x}) \tag{10}$$

Then posterior for **s** is obtained

$$p(\mathbf{s} \mid \mathbf{x}, \boldsymbol{\alpha}, \lambda, \beta) = \frac{p(\mathbf{x} \mid \mathbf{s}, \boldsymbol{\alpha}, \lambda, \beta)\, p(\mathbf{s} \mid \boldsymbol{\alpha})}{p(\mathbf{x} \mid \boldsymbol{\alpha}, \lambda, \beta)} \tag{11}$$

$$= (2\pi)^{-N/2} \mid \boldsymbol{\Sigma} \mid^{-1/2} \exp\{-\frac{1}{2}(\mathbf{s}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{s}-\boldsymbol{\mu})\} \tag{12}$$

Where

$$\boldsymbol{\Sigma} = (\beta \mathbf{A}^{\mathbf{T}} \mathbf{A} + \boldsymbol{\Lambda})^{-1} \tag{13}$$

$$\boldsymbol{\mu} = \beta \boldsymbol{\Sigma} \mathbf{A}^{\mathbf{T}} \mathbf{x} \tag{14}$$

$\mathbf{\Lambda}$ in (13) is defined as $\mathbf{\Lambda} = diag(1/\alpha_1, 1/\alpha_2, \cdots, 1/\alpha_N)$. We can observe from (12) that if $\boldsymbol{\mu}$ is estimated, and then $\mathbf{s}$ can be obtained. The second item in (10) can be written as

$$p(\boldsymbol{\alpha}, \lambda, \beta \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \boldsymbol{\alpha}, \lambda, \beta) p(\boldsymbol{\alpha} \mid \lambda) p(\lambda) p(\beta)}{p(\mathbf{x})}$$

$$\propto p(\mathbf{x} \mid \boldsymbol{\alpha}, \lambda, \beta) p(\boldsymbol{\alpha} \mid \lambda) p(\lambda) p(\beta) \tag{15}$$

The prior for parameter λ is assumed to be uniform. Parameters $\boldsymbol{\alpha}, \lambda, \beta$ are obtained by maximizing posterior $p(\boldsymbol{\alpha}, \lambda, \beta \mid \mathbf{x})$. Instead of maximizing the posterior directly, we maximize equivalently its logarithm, which results in the following function to be maximized:

$$L = \ln p(x \mid \boldsymbol{\alpha}, \lambda, \beta) + \ln p(\boldsymbol{\alpha} \mid \lambda) + \ln p(\beta) \tag{16}$$

Where $p(x \mid \boldsymbol{\alpha}, \lambda, \beta)$ is calculated from (11):

$$\begin{aligned} p(\mathbf{x} \mid \boldsymbol{\alpha}, \lambda, \beta) &= \int p(\mathbf{x} \mid \mathbf{s}, \boldsymbol{\alpha}, \lambda, \beta) p(\mathbf{s} \mid \boldsymbol{\alpha}) d\mathbf{s} \\ &= (2\pi)^{-M/2} \mid \beta^{-1}\mathbf{I} + \mathbf{A}\mathbf{\Lambda}^{-1}\mathbf{A}^{\mathbf{T}} \mid^{-1/2} e^{\{-\frac{1}{2}\mathbf{x}^T(\beta^{-1}\mathbf{I}+\mathbf{A}\mathbf{\Lambda}^{-1}\mathbf{A}^{\mathbf{T}})^{-1}\mathbf{x}\}} \end{aligned} \tag{17}$$

Let $\mathbf{C} = \beta^{-1}\mathbf{I} + \mathbf{A}\mathbf{\Lambda}^{-1}\mathbf{A}^{\mathbf{T}}$, then

$$L = -\frac{1}{2}\ln \mid \mathbf{C} \mid -\frac{1}{2}\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x} + N\ln\frac{\lambda}{2} - \frac{\lambda}{2}\sum_i \alpha_i + (a-1)\ln\beta - b\beta + const \tag{18}$$

Where *const* is independent of the parameters $\boldsymbol{\alpha}, \lambda, \beta$. Calculating the derivatives of L with respect to α_i, λ, β respectively and setting them equal to zero, we can estimate the parameters:

$$\alpha_i = -\frac{1}{2\lambda} + \sqrt{\frac{1}{4\lambda^2} + \frac{\mu_i^2 + \Sigma_{ii}}{\lambda}} \tag{19}$$

$$\lambda = \frac{2N}{\sum_i \alpha_i} \tag{20}$$

$$\beta = \frac{N + 2a}{\| \mathbf{x} - \mathbf{A}\boldsymbol{\mu} \|^2 + 2b} \tag{21}$$

Where Σ_{ii} is the *i*th diagonal element of Σ. Then the basic iterative reconstruction algorithm consists of (13), (14), (19), (20) and (21). Sources are estimated when the algorithm converges.

3.3 Fast Algorithm for Sparse Signal Reconstruction

The algorithm proposed in the previous section converges slowly for the inverse matrix needs to be calculated. We propose a fast algorithm based on the idea similar to [4]. C in (18) can be rewritten as:

$$\begin{aligned}\mathbf{C} &= \beta^{-1}\mathbf{I} + \sum_{m\neq i}\alpha_m \mathbf{a}_m \mathbf{a}_m^T + \alpha_i \mathbf{a}_i \mathbf{a}_i^T \\ &= \mathbf{C}_{-i} + \alpha_i \mathbf{a}_i \mathbf{a}_i^T\end{aligned} \tag{22}$$

where $\mathbf{C}_{-i}$ denotes that the contribution of the *i*th basis is not included. Using the Woodbury identity in (22) and the determinant identity, we obtain

$$|\mathbf{C}| = |\mathbf{C}_{-i}||1+\alpha_i \mathbf{a}_i^T \mathbf{C}_{-i}^{-1}\mathbf{a}_i| \tag{23}$$

$$\mathbf{C}^{-1} = \mathbf{C}_{-i}^{-1} - \frac{\mathbf{C}_{-i}^{-1}\mathbf{a}_i\mathbf{a}_i^T\mathbf{C}_{-i}^{-1}}{1/\alpha_i + \mathbf{a}_i^T\mathbf{C}_{-i}^{-1}\mathbf{a}_i} \tag{24}$$

Treating L in (18) as a function of $\boldsymbol{\alpha}$ only and considering (23), (24), we obtain

$$\begin{aligned}L(\boldsymbol{\alpha}) &= [-\frac{1}{2}\ln|\mathbf{C}_{-i}| - \frac{1}{2}\mathbf{x}^T\mathbf{C}_{-i}^{-1}\mathbf{x} - \frac{\lambda}{2}\sum_{j\neq i}\alpha_j] + [\frac{1}{2}\ln\frac{1}{1+\alpha_i h_i} + \frac{q_i^2\alpha_i}{2(1+\alpha_i h_i)} - \frac{\lambda\alpha_i}{2}] \\ &= L(\boldsymbol{\alpha}_{-i}) + l(\alpha_i)\end{aligned} \tag{25}$$

where

$$l(\alpha_i) = \frac{1}{2}\ln\frac{1}{1+\alpha_i h_i} + \frac{q_i^2\alpha_i}{2(1+\alpha_i h_i)} - \frac{\lambda\alpha_i}{2}$$

$$h_i \triangleq \mathbf{a}_i^T\mathbf{C}_{-i}^{-1}\mathbf{a}_i \quad , \quad q_i \triangleq \mathbf{a}_i^T\mathbf{C}_{-i}^{-1}\mathbf{x} \tag{26}$$

The derivative of (26) with respect to α_i can be expressed as

$$\frac{\partial L(\boldsymbol{\alpha})}{\partial \alpha_i} = \frac{\partial l(\alpha_i)}{\partial \alpha_i} = -\frac{1}{2}[\frac{(\lambda h_i^2)\alpha_i^2 + (h_i^2 + 2\lambda h_i)\alpha_i + (\lambda + h_i - q_i^2)}{(1+\alpha_i h_i)^2}] \tag{27}$$

$\partial L(\boldsymbol{\alpha})/\partial\alpha_i = 0$ is satisfied at

$$\alpha_i = \frac{-h_i(h_i+2\lambda) \pm h_i\sqrt{\Delta}}{2\lambda h_i^2} \tag{28}$$

where $\Delta = (h_i+2\lambda)^2 - 4\lambda(h_i - q_i^2 + \lambda) > 0$. Note that if $q_i^2 - h_i < \lambda$, then $h_i + 2\lambda > \sqrt{\Delta}$ and both solutions in (28) are negative, which does not satisfy the requirement of α_i ($\alpha_i \geq 0$ for it represents variance). The maximum of $L(\boldsymbol{\alpha})$

occurs at $\alpha_i = 0$ for $\left.\frac{\partial L(\boldsymbol{\alpha})}{\partial \alpha_i}\right|_{\alpha_i=0} < 0$. We also observe that if $q_i^2 - h_i > \lambda$, then there are two real solutions for (28), one is negative and the other is positive. The positive solution is chosen as the final solution. In summary, the maximum of $L(\boldsymbol{\alpha})$ is obtained at

$$\alpha_i = \begin{cases} \dfrac{-h_i(h_i + 2\lambda) + h_i\sqrt{\Delta}}{2\lambda h_i^2}, & \text{if } q_i^2 - h_i > \lambda \\ 0, & else \end{cases} \tag{29}$$

In the case of $\alpha_i = 0$, the value of the corresponding signal is near zero, and the corresponding basis vector can be pruned out from the model. According to (29), sparse signal reconstruction can be implemented as follows:

(a) Assume that Ω is basis vector sets related to non-sparse signals.
(b) Basis vector $\mathbf{a}_i$ is pruned out fromΩand α_i is set equal to zero if $\mathbf{a}_i$ already belongs to Ω and $q_i^2 - h_i \le \lambda$.
(c) Basis vector $\mathbf{a}_i$ is added toΩ if $\mathbf{a}_i$ is not in Ω and $q_i^2 - h_i > \lambda$.
(d) α_i is reestimated if $\mathbf{a}_i$ belongs to Ω and $q_i^2 - h_i > \lambda$.
(e) Repeat (a)-(d) until the algorithm converges.

α_i is estimated if the algorithm converges and then signals are also estimated. For simplification of calculation, we have defined:

$$H_i = \mathbf{a}_i^T \mathbf{C}^{-1} \mathbf{a}_i, \quad Q_i = \mathbf{a}_i^T \mathbf{C}^{-1} \mathbf{x} \tag{30}$$

Then

$$h_i = \frac{\alpha_i H_i}{\alpha_i - H_i}, \quad q_i = \frac{\alpha_i Q_i}{\alpha_i - Q_i} \tag{31}$$

H_i and Q_i can be calculated based on matrix determinant and inverse identities:

$$H_i = \beta \mathbf{a}_i^T \mathbf{a}_i - \beta^2 \mathbf{a}_i^T \mathbf{a} \boldsymbol{\Sigma} \mathbf{a}^T \mathbf{a}_i \tag{32}$$

$$Q_i = \beta \mathbf{a}_i^T \mathbf{x} - \beta^2 \mathbf{a}_i^T \mathbf{a} \boldsymbol{\Sigma} \mathbf{a}^T \mathbf{x} \tag{33}$$

The fast algorithm is summarized below.

A fast algorithm for sparse reconstruction

Initialization : λ, $\{\alpha_i\}$, $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, Ω

While converge condition is not satisfied

Choose a basis vector $\mathbf{a}_i$

if $q_i^2 - h_i > \lambda$ and $\alpha_i = 0$ then

Add $\mathbf{a}_i$ to Ω

else if $q_i^2 - h_i > \lambda$ and $\alpha_i > 0$ then

Reestimate α_i

else if $q_i^2 - h \leq_i \lambda$ and $\alpha_i > 0$ then

Prune $\mathbf{a}_i$ from Ω and set $\alpha_i = 0$

end if

Update $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, h_i, q_i, λ

End while

s = μ

End of the algorithm

4 Experiments

4.1 Experiment 1

In this experiment, we consider a length N=512 signal that contains k=20 spikes created by choosing 20 locations at random and then putting ± 1 at these points. The mixing matrix is constructed by creating a matrix with i.i.d draws of a Gaussian

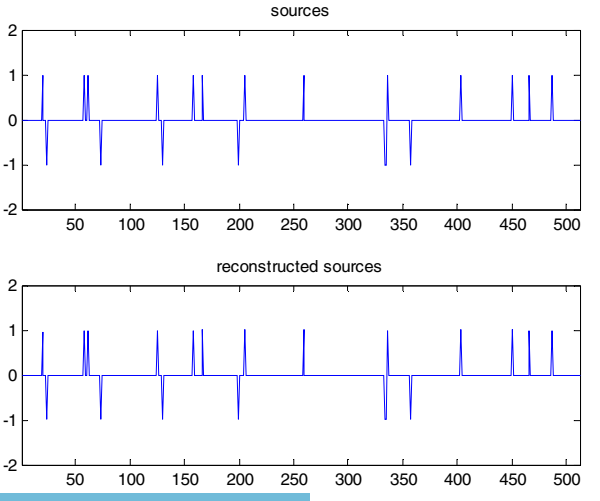


Fig. 1. Simulation for sparse reconstruction using synthetic data

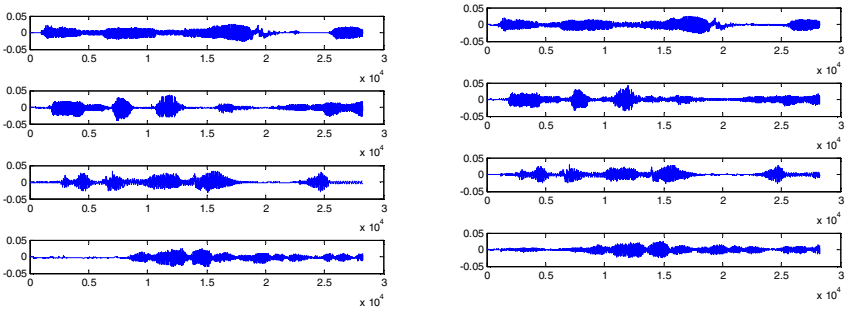
distribution. Zero-mean Gaussian noise is chosen with standard deviation 0.005. Signal-to-Noise Ratio (SNR) is used to evaluate the estimation quality. It is defined as $20\log(\|\mathbf{s}\|/\|\mathbf{s}-\hat{\mathbf{s}}\|)$, where $\mathbf{s}$ and $\hat{\mathbf{s}}$ denote the actual source and its estimation, respectively. Fig. 1 is the simulation results. Sparse signals are reconstructed with SNR=35.0037(average performance of 100 times experiments).

4.2 Experiment 2

In this experiment, the algorithm is applied to underdetermined blind source separation. A FourVoices data set is used for the experiment and there are two sensors. Mixing matrix A is

$$A = \begin{bmatrix} 0.9985 & -0.1142 & 0.2152 & 0.4027 \\ 0.0555 & -0.9935 & -0.9766 & 0.9153 \end{bmatrix}$$

Sparse representation for voices in time-frequency domain is better than that in time domain. So mixtures are first transformed into time-frequency domain using short FFT. Then the algorithm is implemented in time-frequency domain. Finally, reconstructed signals in time-frequency domain are transformed back into signals in time domain. Signal length for the experiment is 28244. The length of short FFT is 2048. Signal-to-Noise Ratio (SNR) is used to evaluate the estimation quality. Fig. 2 is the experiment results with SNR=13.396(average performance of 100 times experiments). Signals are reconstructed successfully and can be heard clearly using earphone.



(a) (b)

Fig. 2. Simulation for sparse reconstruction using voice data
(a) original waveforms (b) reconstructed waveforms

5 Conclusions

This paper studies the problem of underdetermined blind source separation based on the relevance vector machine. A hierarchical prior is adopted for sparse prior and sources are estimated based on MAP. A fast reconstruction algorithm is proposed. It

is experimented that the proposed algorithm reconstructs sparse signals using synthetic data with high SNR and four voices are estimated successfully when the algorithm is applied to UBSS.

References

1. Tipping, M.E.: The Relevance Vector Machine. In: Solla, S.A., Leen, T.K., Müller, K.-R. (eds.) Advances in Neural Information Processing Systems 12, pp. 652–658. MIT Press (2000)
2. Tipping, M.E.: Sparse Bayesian learning and the relevance vector machine. Journal of Machine learning Research 1, 211–244 (2001)
3. Figueiredo, M.A.T.: Adaptive sparseness for supervised learning. IEEE Trans. Pattern Anal. Mach. Intell. 25(9), 1150–1159 (2003)
4. Tipping, M.E., Faul, A.C.: Fast marginal likelihood maximization for sparse Bayesian models. In: Bishop, C.M., Frey, B.J. (eds.) Proc. of the Ninth International Workshop on Artificial Intelligence and Statistics (2003)
5. Ji, S., Xue, Y., Carin, L.: Bayesian compressive sensing. IEEE Trans. Signal Process. 56(6), 2346–2356 (2008)

Research on the Construction and Method System of Enterprise Information Ecosystem

Chen Song and Shuhang Guo

School of Information, Central University of Finance and Economics, 100081, Beijing, China
flyingarchers@hotmail.com

Abstract. Enterprise information ecosystem is a theory which describes and instructs the architecture of enterprise information system based on ecology theories. The paper defines the level of enterprise information ecosystem. It puts forward an idea that constructing the system of theories of enterprise information ecosystem bases on population theory and community theory. The definition of the idea is presented. The reason of why the idea is bring out is provided.

Keywords: Enterprise information system, Information ecosystem, Population and community theory.

1 Introduction

Enterprise information ecosystem is a brand-new concept in recent years. It is likely an efficient method to organize and construct enterprise information system. Since the terminology of information ecology proposed, scholars all over the world have researched this hotspot issue. To explore more effective and efficient instructions of enterprise information system strategy and IT program, we are attempting to bring up a theoretical system of enterprise information ecosystem based on theories of populations and communities, as well as relevant directions of research.

2 Enterprise Information Ecosystem

In 1989, German scholar Rafael Capurro [1] published his report at the "Information & Quality" seminar in Copenhagen and first officially suggested the terminology of information ecology. Information ecology is a science in order to research the relations between people and information with the help of ecological knowledge and method.

Based on the idea of information ecology, the concept of enterprise information ecosystem came into being in recent years. Enterprise information ecosystem was a system constructed by people, information, technology and practice. It runs commercial activities and information communication in enterprises.

Scholars all over the world have researched this hotspot issue. They systematically expounded the theory of information ecology, enterprise information system and ecosystem ecology on many sides. However, researching achievements about

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 671–677.
springerlink.com

enterprise information ecosystem are still rare. Current researches of enterprise information ecosystem can be divided into several parts, including basic elements, the modes of structure, the ecological characteristics, ecological chain, operating mechanism and methods of development and evolution.

Current researches of enterprise information ecosystem can be divided into several aspects, including the concepts, basic elements, organization mode, characteristics, information ecological chain, operating mechanism and evolution.

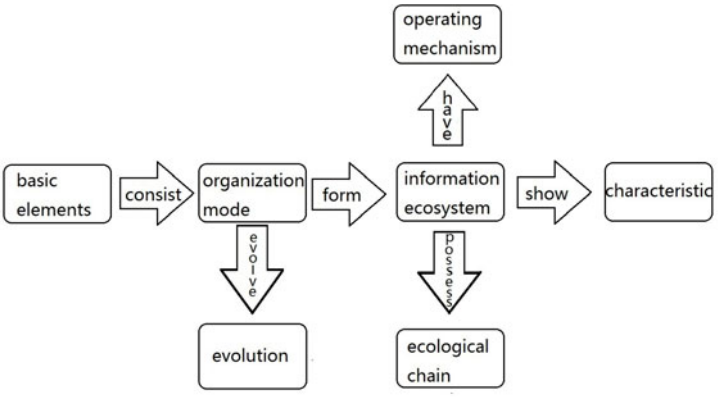


Fig. 1. The model of relations between the directions of researches of information ecosystem

There are some limitations among current researches of enterprise information ecosystem. There are no unified definitions on enterprise information ecosystem and lots of related notions. And the quotation from ecosystem is still lack of conformity and operability. Moreover, there are few effective researching methods, which has made the studies is difficult to do further researches and analysis. Besides, the theories in enterprise information ecosystem can rarely put into practice, and therefore, they can not give effective suggestions to the application of enterprise information ecosystem.

3 Enterprise Information Ecosystem

Researches of constructing enterprise information system are valuable in macroscopically level at present stage, yet still call for more details in concrete practice. Concerning about this, the concept that researching enterprise information ecosystem basing on theories of populations and communities is put forward.

3.1 Levels of Ecology

Ecology concerns about the relations that living organisms have between each other and between their natural environments. Ecology can be divided into four levels, individual level, population level, community level and ecosystem level. Individual

level investigates the physiology and behaviors of organisms, especially their adaptation to environmental factors. Population level focuses on all the organisms that belong to the same species and live in the same geographical area, which was influenced by itself and environment. Relatively, the community level is more complicated. It examines how interactions among species and their future prospect and diversity of species within communities. Ecosystems are of the most various kinds and sizes. They have their own features and characteristics, like individuality, homeostasis or energy system, etc.

Fig. 2. The level of ecology

3.2 Associations between Ecology and Enterprise Information Ecosystem

A population is all the organisms that both belong to the same species and live in the same geographical area in the same time. This definition stresses that a population contains one kind of specie only. The organisms in the same species share similar living styles and habits, and they are in the same position in food chain. In a certain subsystem in enterprise information ecosystem, several kinds of functional modules have similar properties. Most of them are in the same information ecosystem niche, and they work together to complete the functional tasks of subsystems. Therefore, the subsystem has the characteristics of population.

A community is all the populations that live in the same geographical area in the same time. A community is made up of several populations, which connect with each others. The populations in a community have internal environment and structure. In interior part of enterprise information ecosystem, different subsystems have their own functional characteristics and dynamic. Meanwhile, those subsystems stay in large enterprise information system, and they chain each other together through information. Therefore, they have the characteristics of community.

An ecosystem is a whole biological environment consisting of all the organisms living in a particular area, as well as all the nonliving, physical components of the environment with which the organisms interact. They form into integration through energy flow and material recycling. The ecosystem has introduced the inorganic environment of the objective world, and studied the position of the species in recycling and the exchange with the environment from the perspective of energy and material flow. In enterprise information ecosystem, the information system is not alone but interacts with the external world to complete information exchange through network or interfaces. An information ecosystem that does not exchange information with the external world is just like an ecosystem that loses the energy and material resources, thus the balance may be broken.

Population ecology mainly studies the ecological characteristics of population, its number as well as its connection with the environment. In the enterprise information ecosystem, for each single subsystem, the characteristics and total number of components, and the influence of the subsystem on the components are main points of researches.

Community ecology mainly studies composition characteristics of all the populations which are in the same area, the connections among each other as well as with the environment, the population structure and formation, and the internal mechanisms. In the enterprise information ecosystem, how to construct the logic structure among different subsystems, how the subsystems connect with each other and the whole system, the evolution and development process in the development stage and maintenance phase are research focus.

Ecosystem Ecology divides the components in the system into environment, the product, the consumer and the decomposer from the perspective of energy and material flow. Through the researches on the functions and characteristics of these four aspects, it analyzes the energy recycling and material recycling chain in the system. In the studies of enterprise information ecosystem, there are also similar theories that consider different functional modules as the roles in ecosystems. The components accomplish the useful information recycling through the information flow.

Basic on the discussions mentioned above, it can be drawn that when studying enterprise information ecosystem, the functional modules can be considered as individuals, the subsystems can be considered as populations, the whole information system can be considered as community, and the external environment that connects with the enterprise information ecosystem as well as the whole system can be considered as the ecosystem.

3.3 Define the Levels of Enterprise Information Ecosystem

Definition 1. Information Individual:

Information Individual = {programs, procedures}.

Program refers to those components that are used to realize the specific functions. The programs connect each other together through procedures, and then an integrated complete module is formed. The function of each module can be embodied in the system, while the data flow within the information individual is microcosmic.

Definition 2. Information Population:

Information Population = {functional modules, interfaces, rational}.

Functional module refers to the procedure set that is responsible for the accomplishment of some functions, and each functional module exchange the data through the interface. There are not absolute procedure rules and hierarchy among functional modules. The selection and topology of the modules are completed through some rational. The system flow as it is usually called actually is just an expression of the rational.

Definition 3. Information Community:

Information Community = {subsystems, connections, importance, structure}.

A subsystem deals with a certain category of data and realizes all functions in a particular field. The subsystem is an integrated and stand-alone functional system. The connections among subsystems achieve the topology. Subsystems have different importance. Dominant species in the information community can be distinguished through comparing the importance index. There exists the structure in the information community, and rational structure helps to run the whole system.

Definition 4. Information Ecosystem:

Information Ecosystem = {system, networks, constrains, environments}.

The system exchanges data and updates knowledge with the outside information environments through network. An information ecosystem also contains some constrains. Through filtering out unrelated information that has nothing to do with the system, they can delimit the scope for the information ecosystem.

3.4 Proposal of Focus on Information Population and Information Community

There are two reasons could explain the phenomenon that most ecologists lovely to focus more on populations. First, it is hard to establish accurate mathematical model to describe more than one population. Second, because species and environments are different and unique in details, so it is hard to use abstract analysis without fixing to particular species. [17] CJ Krebs [18] also points out that the best way to learn about a large and complicated system is to study small segments or details in great depth rather than to start by trying to develop a sense of the whole.

In enterprise information ecosystem, however, it is essential to study about communities. Communities include structures, which is really crucial to the system. Moreover, researching on information communities is easier than real communities. Mathematical model is still hard to be made, yet logical model is more important for an information community mathematical.

Theories of ecology about populations and communities have been applied widely in lots of researches. For instance, Timothy Bartram [19] research employee management depends on a mathematic model of population theory; PD Allison [20] uses the Logistic equation as a support in a research of SAS. Therefore, theories of populations and communities are sound enough to be quote and drew by researches of other subject.

In enterprise information ecosystems, details inside subsystems and the integrated whole system are the most important part. Subsystems are defined as information populations and systems without connecting to outside are defined as information communities.

Based on the discussions above, researches of enterprise information ecosystem should focus on populations and communities levels. Theories of communities provide tangible logic frame, process of evolve and organizational system. Theories of populations provide mathematic model, which could integrate reality with data, and make theories valuable for practice. The combination of theories of populations and communities could instruct the construction of the system of both microscopic and macroscopic theories. That means the researches will be more accurate directed and easier to operate.

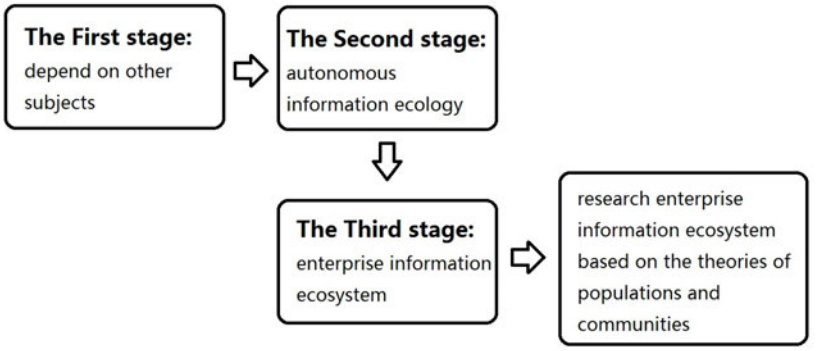


Fig. 3. The developing model of enterprise information ecosystem

4 Conclusion

The paper analyses existing researches of enterprise information ecosystem from different perspectives, including the basic elements, the modes of structure, the ecological characteristics, ecological chain, operating mechanism and methods of development and evolution. A concept of research enterprise information ecosystem using theories of population ecology and community ecology is put forward.

Acknowledgments. This work was supported by Project 70872119 of the National Natural Science Foundation of China, by Project 109016 of the Key Project of Chinese Ministry of Education of China, by Project 09YJC630240 of the Humanities and Social Sciences Fund of Chinese Ministry of Education.

References

1. Capurro, R.: Towards an information ecology. NORDINFO International seminar "Information and Quality", Royal School of Librarianship, Copenhagen, August 23-25 (1989)
2. Zhou, Q., Li, H., Zhu, J.: An Introduction to the Resear ch of Information Ecology and Basic Conception. Library and Information 6, 24–29 (2006)
3. Davenport, T.H.: Information Ecology—Mastering the Information and Knowledge Environment, pp. 140–142. Oxford University Press, US (1997)
4. Li, M.: The Analysis of Information Ecosystem. Journal of Intelligence (7)(1998)
5. Wang, X., Liu, D.: Research of the Elements and Evaluation Index Construction of Enterpr ise Information Ecological System. Library and Information Service 54(16), 22–25 (2010)
6. The ecology of interorganizational information sharing. Journal of International Technology and Information Management 13(2), 73–85 (2004)
7. Ma, J., Jing, J., Zhang, X.: Research on Information Organization Model of Information Ecosystem. Library and Information Service 10, 15–19 (2010)
8. Zhang, H., Yan, Y., Leng, X.: Logic Model and Operating Mechanism of Enterprise Information Ecosystem. Information Studies: Theory & Application 33(4), 6–9 (2010)

9. Nardi, B.A., O' Day, V.L.: Information ecologies: Using technology with heart, pp. 52–54. MIT Press, Cambridge (1999)
10. Xu, X., Bian, Y., Jiang, Y.: Analysis of Characteristics of Enterprise Information System based on Ecological Theories. Science and Technology Management Research 11, 375–377 (2009)
11. Sun, S., Yen, J.: Information Supply Chain: A Unified Framework for Information-Sharing. In: Kantor, P., Muresan, G., Roberts, F., Zeng, D.D., Wang, F.-Y., Chen, H., Merkle, R.C. (eds.) ISI 2005. LNCS, vol. 3495, pp. 422–428. Springer, Heidelberg (2005)
12. Xu, X., Li, X.: Building on Model of Enterprise Information System Growth Based on Ecological Theory. Journal of Intelligence 9, 17–20 (2009)
13. Gang, H., Zheng, T.: Information Ecological Chain: A Theoretical Frame. Information Studies: Theory & Application 1, 18–21, 32 (2007)
14. Zhang, X., Guo, J., Ma, J.: Research on Information Cooperation Model of Enterprise Information Ecosystem. Information Studies: Theory & Application 33(4), 10–13 (2010)
15. Mische, M.A., Bennis, W.: Reinventing Through Reengineering. Information Systems Management 13(3), 58–65 (1996)
16. Li, B., Jing, J., Wang, J.: Evolution Mechanism of Ecological Information Communities. Library and Information Service 10, 6–10 (2010)
17. Ulanowicz, R.E.: Information theory in ecology. Computers & Chemistry 25(4), 393–399 (2001)
18. Krebs, C.J.: Ecological methodology, 2nd edn. Benjamin Cummings, US (1998)
19. Bartram, T.: Employee management systems and organizational contexts: A population ecology approach. Management Research Review 34(6) (2010)
20. Allison, P.D.: Logistic regression using the SAS system: theory and application. SAS Publishing, US (2008)
21. He, Y.-G., Huang, L.H., Dai, W.H.: Research on Information Systems Evolution Based on Ecology. Science & Technology Review 1, 41–44 (2006)
22. Chen, Y., Chen, Z., Wang, J.: The Ultimate Goal of Enterprise Informatization: Make Up a Healthy Information System. Journal of Intelligence 6, 108–110 (2007)
23. Zhang, H., Yan, Y., Leng, X.: Research on Deconstruction and Ternary Natures of Enterprise Information Ecosystem. Library and Information Service 54(10), 11–14 (2010)
24. Lu, J., Guo, D.: Progress in Study of Information Ecology Theory. Journal of Intelligence 3, 82–84 (2007)
25. Zhou, X., Xia, Z., Dong, Y.: The Hot Spots Analysis and Forecast of Information Ecology. Journal of Intelligence 28(12), 178–181 (2009)
26. Niu, C., Lou, A., Sun, R.: Basic Ecology, 2nd edn. Higher Education Press, Beijing (2002)

Satisfation Statistical Analysis on Explanding the Channels of Employment in Hebei Province*

Yun-Chao Bai

College of Economics, Hebei University, Baoding 071002, China
yunchaobai@126.com

Abstract. The past two years, Employent situations have improved in hebei, but the situation is still grim. The first three quarters of 2011, sscience and engineering employment rate is less than 78%, literature and history less than 30%.Expanding the channels of employment satisfaction are directly related to employment rate. This paper is based on AHP and establish the emploment satisfaction model according to fuzzy comprehensive evaluation method. through a combination of qualitative and quantitative analysis, we give the objective evaluation of satisfaction on the expansion employment channels. Finally, according to reviewing results, we give a reasonable policy and recommendations.

Keywords: AHP, Fuzzy comprehensive evaluation, job satisfaction of college students.

1 Introduction

Facing severe employment situation colleges and uneversities in hebei province take a variety of means to improve employment rate, Mainly the following aspect: A_1:Vigorously promoting curriculum regorm professional settings of school are contact with the market demans; A_2 :Expanding the cooperation of school-enterprise and Inter school, So that the combination of the theoty and practice; A_3 :to helping students better grasp employment opportunites shoud pay attention to creer planning and strenghthening career guidance; A_4 :In order to reduce the employment pressure, focus on helping the graduates of economic and employent difficulties; A_5 :to establish students starting bases build the business long-term mechanism to promote employment.The article of satisfaction evaluation of college students are not much, in addition,the evaluation of one thing often involve multiple factors of multiple indicators, Evaluation is comprehensive in the interaction of multiple factors. When things have characteristics that he is also getting behind, the classic statstical methods seem powerles so it is necessary to introduce new statistical methods so that achives satisfac-

* The work is supported by Social Science Development Research of People Livelihood Special of Hebei Province(No:201101258)and Social Science Foundation of Baoding (No : 201102010).

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 679–684.
springerlink.com

tory, objective and realistic comprehensive evaluation. Fuzy mathematics is as new branch of mathematics that deals with the fuzzy things of concept for their study object, so that related disciplines are likely to use quantitative and mathimatical approaches to be described and handled which has shown strong vitality and penetration In this paper, AHP has be applied to determine the weight of these five areas then fuzzy comprehensive evaluation method gives the satisfaction of these five aspects of a comprehensive evaluation results, finally based on the evaluation results, this paper gives the reasonable policy recommendations.

2 Determining the Weight with AHP

This step is very imporment, we determine the weight according to importance different index and degree of reflect of student AHP method is also known as multi level decision making and weight analysis method. This approach bombines the qualitative with quantitative, based on the overall objective of the problem and systematic point, the problem is decomposed int a number of factors, according to the relations of domination, APH constitutes a hierarchical structure model, then apply pairwise comparison method to determine relative importance of the decision making programs and obtain a satisfctoy weight distribution. Each element evaluation fctors set has different position and role, this weight is called priority weight value. Determing the weights use calculation table of the relative importance of relevant rating, this method has high reliability and small error. Calculation table of the relative importance of the relevant rating see table 2-1and comparison table of relvtive importance see table 2-2

Table 1. Calculation table of the relative importance of the relevant rating

Score	Meaning	Explanation
1	Equally important,	Two factors are equally important to the same target
3	Slightly important,	Another factor is slightly more than a factor
5	More important,	One factor is more important than another
7	Much more important,	Another factor, compared with a factor, obviously the former is much more important
9	Absolutely essential,	Comparison of the two factors, the former has proven to be absolutely essential for all evaluation factors

Notation: 2, 4, 6, 8 were collected intermediate values from two adjacent odd number.

The relative weight of each factor is $\omega_i \approx \frac{1}{m}\sum_{j=1}^{m}(a_{ij}/\sum_{j=1}^{m}a_{ij})$ The weight distribution vector offive factors is $A=(a_1,a_2,a_3,a_4,a_5)$, meanwhile, $\sum_{i=1}^{m}a_i=1$ （Normalization principle). Thus we get the relative importance of factors comparison table (see Table 1-2). And the random consistency index by calculating the ratio is $C_R<0.1$, So that the five aspects of a reasonable weight distribution on colleges and universities to expand the employment channels is:

$$A=(a_1,a_2,a_3,a_4,a_5)=(0.391,0.290,0.086,0.051,0.182)$$

Table 2. Comparison table of the relative importance

	u_1	u_2	u_3	u_4	u_5	a_i
u_1	1	2	5	7	2	0.391
u_2	1/2	1	4	8	2	0.290
u_3	1/5	1/4	1	3	1/3	0.086
u_4	1/7	1/8	1/3	1	1/4	0.051
u_5	1/2	1/2	3	4	1	0.182

3 Case Analysis

we mainly select Hebei University students in the school to survey questionnaires. Mainly from the students on the five aspects of satisfaction start to survey. As more school students, we only select one part of choice to survey, a relatively dense flow of Hebei North University, South Campus cafeteria and hospital canteens to carry out the survey. Survey objects include different grades, different gender and different majors. This basically reflects the implementation of the channels of employment situation of the Hebei University,. The survey uses anonymous questionnaire survey method, we dispense questionnaires 1000,800 copies of the actual were recoveried, the recovery rate is 80%

3.1 Determining Domain U of Students' Satisfaction Model

Set $U=\{u_1,u_2,u_3,u_4,u_5\}$, Where u_1: Vigorously promoting curriculum regorm professional settings of school are contact with the market demans; u_2 Expanding the cooperation of school-enterprise and inter-school, So that the combination of the theoty and practice; u_3 to helping students better grasp employment opportunites shoud pay attention to creer planning and strenghthening career guidance; u_4 :In order to reduce the employment pressure, focus on helping the graduates of economic and employent difficulties; u_5 :to establish students starting bases build the business long-term mechanism to promote employment By choosing the five-level indicators. we can understand satisfaction of implementation of the expansion of employment from the overall.

3.2 Determing the Domain Model V of College Students' Satisfaction Level

Set $V=\{v_1,v_2,v_3,v_4,v_5\}=$ {Very satisfied, satisfied, more satisfied, generally satisfied, not satisfied} . Determination ofthe level of discourse must meet the requirements, that is, the number of levels are greater than 4 less than 9 , take a moderate value which is not too big or too small to affect the evaluation results. Here m is equal to 5.

3.3 Determining the Relationship Matrix of Satisfaction Model

Set

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} & r_{14} & r_{15} \\ r_{21} & r_{22} & r_{23} & r_{24} & r_{25} \\ r_{31} & r_{32} & r_{33} & r_{34} & r_{35} \\ r_{41} & r_{42} & r_{43} & r_{44} & r_{45} \\ r_{51} & r_{52} & r_{53} & r_{54} & r_{55} \end{bmatrix}$$

In this survey, students assessed results of the five indicators in Table 3-1:

Table 3. Evaluation results of residents' satisfaction

Factors set U	Evaluation set V				
	v_1	v_2	v_3	v_4	v_5
u_1	112	201	234	156	97
u_2	89	168	197	226	120
u_3	203	135	228	102	132
u_4	141	164	97	267	131
u_5	82	173	166	236	143

It can be seen through the table for vigorously promoting curriculum reform, the school of professional settings to establish contact with the market demand, there are 112 people that are very satisfied, satisfied that there are 201 people, 234 were considered to be more satisfied, there is general satisfaction that 156 people have 97 unsatisfactory, so we have come to fuzzy evaluation matrix is:

$$R = \begin{bmatrix} 0.14 & 0.25125 & 0.2925 & 0.195 & 0.12125 \\ 0.11125 & 0.21 & 0.24625 & 0.2825 & 0.15 \\ 0.25375 & 0.16875 & 0.285 & 0.1275 & 0.165 \\ 0.17625 & 0.205 & 0.12125 & 0.33375 & 0.16375 \\ 0.1025 & 0.21625 & 0.2075 & 0.295 & 0.17875 \end{bmatrix}$$

3.4 Comprehensive Evaluation of Students Satisfaction

This step is the most critical. In this step, we must choose operator. The choice of composite operator "$\circ$" have four methods under normal circumstances. They are $M(\wedge, \vee) \cdot M(\bullet, \vee)$, $M(\bullet, +)$ and $M(\wedge, +)$ .In this we use the $M(\bullet, +)$ and $M(\wedge, +)$.

1. We use $M(\bullet,+)$ and get results：

$$B = A \circ R = (0.391,0.290,0.086,0.051,0.182) \circ \begin{bmatrix} 0.14 & 0.25125 & 0.2925 & 0.195 & 0.12125 \\ 0.11125 & 0.21 & 0.24625 & 0.2825 & 0.15 \\ 0.25375 & 0.16875 & 0.285 & 0.1275 & 0.165 \\ 0.17625 & 0.205 & 0.12125 & 0.33375 & 0.16375 \\ 0.1025 & 0.21625 & 0.2075 & 0.295 & 0.17875 \end{bmatrix}$$

$$= (0.144,0.226,0.254,0.240,0.016)$$

2. we use $M(\wedge,+)$ and get results：

$$B = A \circ R = (0.391,0.290,0.086,0.051,0.182) \circ \begin{bmatrix} 0.14 & 0.25125 & 0.2925 & 0.195 & 0.12125 \\ 0.11125 & 0.21 & 0.24625 & 0.2825 & 0.15 \\ 0.25375 & 0.16875 & 0.285 & 0.1275 & 0.165 \\ 0.17625 & 0.205 & 0.12125 & 0.33375 & 0.16375 \\ 0.1025 & 0.21625 & 0.2075 & 0.295 & 0.17875 \end{bmatrix}$$

$$= (0.491,0.780,0.860,0.797,0.587)$$

By the results of above two calculations, based on the principle of maximum membership, satisfaction of employment channels for students in colleges and universities to expand is satisfied. Fuzzy comprehensive evaluation can be quantified fuzzy things, our result is still a vector, if we give a column vector C of assessment, then it can calculate a specific score. We make "very satisfied" scores 5 "satisfactory" score 4 "relatively satisfied" scores3, "generally satisfactory" score 2, "unsatisfactory" score 1•then $C = (5,4,3,2,1)^T$, Final score of evaluation of students satisfaction is:

$$Q = B \circ C = (0.144,0.226,0.254,0.240,0.016) \circ (5,4,3,2,1)^T = 2.982$$

Accordingly, a visual impression of the student satisfaction is between more satisfaction and general satisfaction and tends to general satisfaction. From the questionnaire itself and the actual situation, to expand school-enterprise, inter-school cooperation, to enable students to combine theory and practice and the establishment of university business base, to establish a long-term mechanism to promote employment start their first two scores is low than the other four, we suggested expanding the channels of employment in the implementation of colleges and universities to give these two aspects of the proper tilt. It will receive better results.

4 Conclusion

In this paper, we use AHP to construct the evaluation index system weight distribution, and can fully reflect the actual attitude to expand the channels of employment ,the use of fuzzy comprehensive evaluation method make each student be able to independently exercise their right satisfaction of the students can be more objective and true. This method can be people's minds and subjective judgments quantitative, it not only simplifies the systematic analysis and calculations, but also help policy makers to maintain consistency of thought processes and decision-making process. Survey only

reflects the attitude of some students, Therefore, the results of students' satisfaction only provide a reference for policy makers and universities

References

1. Tao, L., Zhen, Z.: The Application of Fuzzy Mathematics in the evaluation of teacher education and teaching ability. Journal of Mathematical Medicine 16(2), 183–184 (2003)
2. Liu, L.: Applied Fuzzy Math. Shaanxi Science and Technology Press (1996)
3. Zhao, H.-C.: AHP - A simple new method of decision-making. Science and Technology Press (1986)
4. Ran, M.-L., Ran, L.-F.: Employment of university students from the Perspective of the market mechanism talent agency Innovative Research. China Business 1 (2010)
5. Li, L., Yuan, X., Hu, J., Hao, L.: Employment of graduates of vocational hospital perspective and countermeasures

The Application of Improved DEA Model in Evaluation of China's Production Comprehensive Efficiency

Yuquan Cui[1], Lejun Shi[2], and Yuwei Fang[1]

[1] School of Mathematics, Shandong University, 250100
[2] Yantai Vocational College, 264025

Abstract. It's a very important problem that how to evaluate China's production efficiency, environmental efficiency and comprehensive efficiency. DEA model is a good method to solve this problem. But DEA model is mainly based on the linear model, therefore, the calculated results have differences with the actual results. In this paper, we improved DEA model. According to the function of output data and input data, we determine the form and number of input variables. If input variables are too many, we can use principal component analysis to reduce the dimensions, and then we use the improved DEA model to calculate. We use the improved DEA model to evaluate and analyze the efficiency of the production environment in China in recent years. Actual evaluation results excellently agree with the actual situation.

Keywords: DEA model, general efficiency, input variables, linear model, nonlinear model.

1 Introduction

Since the late 20th century, the Chinese economy has developed quickly and achieved sustained and rapid growth. The average of China's GDP annual growth reached 9.7% during 1979-2006, ranking the first over the same period in world economic growth. Long-term rapid economic growth has markedly enhanced China's comprehensive national strength and has improved Chinese people's living standard significantly. By 2007, China's GDP has been the second in the world and urban-rural income has increased about six times than that in 1978. However, the emissions associated with rapid economic growth have increased significantly. In 2005, industrial solid waste production is 13.4 million tons, which increased 12% than the previous year. Wastewater emissions increased 26% than that in 2000 and the annual average increase rate is 5.26%. Meanwhile, China's energy consumption is quite amazing during this period. In 2006, China's share of world GDP accounted for about 5.5%, but energy consumption accounted for about 15% of the world. Hw do we evaluate China's production efficiency, environmental efficiency and comprehensive efficiency more objectively over this period?

In this paper, we use the improved DEA model to make a more objective analysis and evaluation on China's production efficiency, environmental efficiency and overall efficiency of each year during this period. DEA model is proposed by U.S. operations

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 685–694.
springerlink.com

research expert A. Charnes, W.W. Cooper and E. Rhodes in 1978. This method is based on linear programming approach theory. By means of frontier production function, we construct production frontier according to certain criteria, and then we get the efficiency which is the gap between production frontier and departments or units to be assessed. This method can resolve the measurement of efficiency evaluation quite well. The advantages of it are as follows: handling the problem of multiple inputs and multiple outputs; calculating the projection of decision-making unit in production frontier; determining the relationship between the input and output elements according to symbols and size of dual price of inefficient decision units, solving optimization strength, arranging the order of optimization. Many scholars have used DEA models to do efficiency evaluation. Such as: Ruggiero(1996) considered endogenous factors and exogenous factors and used DEA methods to evaluate the efficiency of the public sector[1]; Gupta and Verhoeven(2001) estimated efficiency of public expenditure on education and public health spending in 37 countries in Africa from 1984 to 1995[2]; Antreas(2004) estimated the efficiency of all the branches of the tobacco companies according to DEA[3]; Afonso, Schuknecht and Tanzi(2005) built a comprehensive index system of seven sub-indicators constitutes on public-sector of 23 industrial countries, and then used non-parametric production frontier techniques to estimate input-output efficiency of public-sector[4]; Afonso and Fernandes(2006) used DEA technique to estimate efficiency of public expenditure of Portugal local governments[5]; Ning Zhang et al(2006) used DEA to discuss efficiency of health production[6]; Qiongzhi Liu(2007(3),2007(4)) studied the efficiency of social equity and the efficiency of economic growth that the public expenditure achieved[7][8]; Hao Xu et al(2010) studied the overall efficiency of interval DEA[9].Wu et al use DEA method to evaluate energy efficiency[10][11][12][13].

Most people use linear DEA model to evaluate efficiency, but practical problems are often complex. In practical problems the relationship between variables often manifested as nonlinear, non-linear model can better reflect the complex relationships of practical problems. Therefore, in this paper we first use sample data to analyze the relationship of output and each input factors and establish functions, then we use new variables to substitute all these functions and let new variables as input factors, at last we use DEA model to evaluate the efficiency. So the evaluation results are more objective and reasonable than that of linear model. Of course, if the dimension of new variables as input factors is high, we can use principal component analysis method to reduce the dimension, and then we use DEA model to evaluate efficiency. Actual evaluation results excellently agree with the actual situation.

2 Modeling

2.1 DEA Model

Let DMU_j (j=1,…,n) be decision unit, where input factors are $X_j = (x_{1j},\ldots,x_{mj})^T$ and output factors are $Y_j = (y_{1j},\ldots,y_{sj})^T$. The C^2R model that proposed by A. Charnes, W.Cooper, E.Rhodes is like follows:

$$\text{Max} \ \frac{\sum_{k=1}^{s} U_k y_{kj_0}}{\sum_{i=1}^{m} v_i x_{ij_0}} \qquad \text{S.T} \begin{cases} \sum_{k=1}^{s} U_k y_{kj} \Big/ \sum_{i=1}^{m} v_i x_{ij} \le 1 (j = 1, \ldots n) \\ U_k \ge 0 (k = 1, \ldots, s) \\ v_i \ge 0 (i = 1, \ldots, m) \end{cases} \tag{1}$$

Using Charnes-Cooper transformation: $t = 1/v^T x_0, \omega = tv, \mu = tu$, model(1) can be changed into:

$$\max \mu^T y_0 = \bar{V}_p \qquad s.t. \begin{cases} \omega^T x_j - \mu^T y_j \ge 0, j=1,\ldots,n \\ \omega^T x_0 = 1, \omega \ge 0, \mu \ge 0 \end{cases} \tag{2}$$

By the dual theory, the dual programming model of model (2) can be simplified as follows:

$$\min \theta \qquad s.t. \begin{cases} \Sigma_{j=1}^{n} \lambda_j x_j + s^- = \theta x_0 \\ \Sigma_{j=1}^{n} \lambda_j y_j - s^+ = y_0 \\ s^+ \ge 0, s^- \ge 0, \lambda_j \ge 0, j=1,\ldots,n \end{cases} \tag{3}$$

Property 1: Program (2) and (3) exist in solution, and the optimal value equal, that is $\theta = \bar{V}_p \le 1$.

Definition 1: If the solution ω^*, μ^* of program (2) satisfy $V_p = \mu^{*T} y_0 = 1$, DMU_{j_0} is called weak DEA efficient (C^2R).

Definition 2: If the solution of program (2) exist $\omega^* > 0, \mu^* > 0$, and $V_p = \mu^{*T} y_0 = 1$, DMU_{j_0} is called DEA efficient (C^2R).

Property 2: (1) The necessary and sufficient condition of that DMU_{j_0} are weak DEA efficient is that optimal value of program (4) is $\theta = 1$;
(2) The necessary and sufficient condition of that DMU_{j_0} are DEA efficient is that optimal value of program (3) is $\theta = 1$ and each optimal solution $\lambda^*, s^{*-}, s^{*+}, \theta^*$ of it satisfy $s^{*-} = 0, s^{*+} = 0$.

2.2 Improved DEA Model

From model (1) we know, the output y_k (k = 1,...,s) is a linear function of the input x_i (i = 1, ..., m), especially when k=1 $y_k = \Sigma_{i=1}^{m} \alpha_i x_i - \varepsilon_k$. In practical issues, the relationship between output y_k and input x_i is often nonlinear, that is $y_k = f(x_1, \ldots, x_m) - \varepsilon_k$. In this paper we consider the following method to establish

a nonlinear model. We analyze the function of each input x_i (i = 1,...,m) and the output y_k based on actual sample data, and then we get the function of each y_k and every x_i $(i = 1,..., m)$,that is

$y_k = f_{k1}(x_1) - \varepsilon_{k1},..., y_k = f_{km}(x_m) - \varepsilon_{km}$ (k=1,…,s), and then we use $(f_{11}(x_1),..., f_{sm}(x_m))$ as new input variables (the same $f_{ij}(x_j)$ can be combined), the output variable is still y_k. We still use the linear model to establish DEA model. This DEA model is more realistic than the original linear model. In this modeling process we change actually the original m-dimension input into the L-dimension input and we make a transformation on the input space.

Furthermore, we only consider the impact of $f_{ij}(x_j)$ on y_i about entry, and we do not consider the impact of cross terms and the second item on y_i. If these factors are taken into account, the dimension of input variable space is higher, the dimension of output is still the s-dimensional. Because the dimension of new model is higher, we consider reducing dimension of input space by means of principal component analysis (PCA).

The main idea of PCA is that we use linear transformation to change L-dimension input vector X_i =($x_{i1},..., x_{iL}$) to another L-dimension input vector $\overline{X}_i$ =($\overline{x}_{i1},..., \overline{x}_{iL}$), where the arrangement of each component of the new L-dimensional input vector $\overline{X}_i$ is based on the size of the characteristic values of linear transformation. Then, we take the first few components with large eigenvalues to approximate and reduce dimension.

3 Evaluation and Analysis of the Efficiency of the Production Environment in China in Recent Years

In this section we use the improved DEA model to evaluate and analyze the efficiency of the production environment in China in recent years. Analysis and evaluation will be carried out in two ways. On one hand we use a single-output, multiple-input improved DEA model, the GDP of China in recent years and waste emissions as output, capital, labor and energy consumption as input. We establish models to analyze and compare with the results of the original DEA model. On the other hand, we use multiple-input, multiple-input improved DEA model, the GDP of China in recent years and waste emissions as output, capital, labor and energy consumption as input. We establish models to analyze and compare with the results of the original DEA model.

3.1 Evaluation and Analysis of the Efficiency of the Production Environment of Single-Output, Multiple-Input Improved DEA Model

We establish improved DEA model, using the GDP of China in recent years as output, capital, labor and energy consumption as input. According to the data from China Statistical Yearbook, we use the actual GDP (1952 as the base year and the date from

1980 to 2007) as the y-axis, use the working population (data from 1980 to 2007) as the x-axis plot, we know that the actual GDP is similar to the exponential function of the working population. Thus, using the exponential function to approximate, we can get the following equation:

$\ln GDP = -1.5339 + 0.6529J$ and $R^2 = 0.9303, F = 347.1779$ (Unit of GDP is one hundred billion, unit of the working population is one hundred million).

Similarly, according to the data from China Statistical Yearbook, we know that the actual GDP is approximately a linear function of energy consumption and is a log-linear function of capital stock, so we get following equation:

GDP=-10.9216+2.2527N and $R^2 = 0.9634, F = 683.9435$ (N is energy consumption and unit is one million tons);

GDP=-29.1022+14.1551lnz and $R^2 = 0.9473, F = 466.9921$ (z is capital stock and unit is ten billion).

Further, we analyze solid wastes in China (GT is solid wastes and unit is one hundred million tons), wastewater discharge in China (S is wastewater and unit is one million tons), emission in China (Q is emission and unit is one billion tons). After analyzing the relationships between them, we can get :

lnGT=-1.0657+0.4681J, $R^2 = 0.6634, F = 35.4753$;GT=-1.0973+0.6675N, $R^2 = 0.9396, F = 276.573$;

GT=3.9802+0.0962z, $R^2 = 0.9417, F = 290.7073$;lnS=2.3978+0.19J, $R^2 = 0.6269$, $F = 30.2468$;

S=22.0195+1.3034N, $R^2 = 0.955, F = 382.0686$;S=6.9876+9.3004lnz, $R^2 = 0.9042, F = 169.8263$;

lnQ=1.3784+0.2234J, $R^2 = 0.5968$, $F = 26.6402$,P=0.0001(when P<0.05, regression equation is established);Q=9.4281+0.6451N, $R^2 = 0.8376$ $F = 92.8639$;Q=14.3375+0.0929z, $R^2 = 0.8391, F = 93.889$.

Therefore, we use y1 = GDP as output of DEA model and use x1 = e^J , x2 = N, x3 = lnz as input of DEA model, then we establish the CCR model that based on input and calculate the value of production efficiency in China each year, available Table 1:

Year	80	81	82	83	84	85	86	87	88	89
F0	1	0.991	0.9763	0.9886	1	1	0.9956	1	1	0.9786
θ_0	1	0.991	0.9763	0.9886	1	1	0.9956	1	1	0.9786
90	91	92	93	94	95	96	97	98	99	
0.5262	0.6094	0.6554	0.7005	0.7442	0.7754	0.7888	0.8136	0.8652	0.9077	
0.5262	0.6094	0.6554	0.7005	0.7442	0.7754	0.7888	0.8136	0.8652	0.9077	
2000	01	02	03	04	05	06	07			
0.9429	0.9746	1	0.9704	0.9347	0.941	0.9602	1			
0.9429	0.9746	1	0.9704	0.9347	0.941	0.9602	1			

We use y1 = GDP as output of DEA model and use x1 = e^J , x2 = N, x3 = lnz as input of DEA model, then we establish the CCR model as model (5) and calculate the value of production efficiency in China each year, available Table 2:

Year	80	81	82	83	84	85	86	87	88	89
F01	-1	-1.009	-1.0242	-1.0115	-1	-1	-1.0044	-1	-1	-1.0218
θ_{01}	1	1.0091	1.0243	1.0115	1	1	1.0045	1	1	1.0219

90	91	92	93	94	95	96	97	98	99
-1.6765	-1.6403	-1.5251	-1.4268	-1.343	-1.289	-1.2672	-1.228	-1.1548	-1.101
1.6772	1.641	1.5258	1.4275	1.347	1.2896	1.2678	1.229	1.1558	1.1017

2000	01	02	03	04	05	06	07
-1.06	-1.0258	-1	-1.0304	-1.0698	-1.0627	-1.0414	-1
1.0605	1.0261	1	1.0305	1.0699	1.0627	1.0415	1

Then, we use y2 = GT as output of DEA model and use x1 = e^J , x2 = N, x3 = z as input of DEA model and establish the CCR model which Based on the output of non-Archimedean infinitesimal. As more waste emissions, the more serious environmental pollution, so model is as follows:

$$\max \varepsilon(s^-, s^+)(1,\ldots,1)^{'} - \theta \qquad s.t. \begin{cases} \sum_{j=1}^{n} \lambda_j x_j + s^- = x_0 \\ \sum_{j=1}^{n} \lambda_j y_j - s^+ = \theta y_0 \\ s^+ \geq 0, s^- \geq 0, \lambda_j \geq 0, j = 1,\ldots,n \end{cases} \tag{5}$$

We calculate the value of production efficiency of solid wastes in China each year, available in Table 3:

Year	88	89	90	91	92	93	94	95	96	97
F1	-1	-1.03	-1	-1	-1.03	-1.18	-1.31	-1.396	-1.51	-1
θ_g	1	1.03	1	1	1.03	1.18	1.31	1.396	1.51	1

98	99	00	01	02	03	04	05	06	07
-1.12	-1.105	-1.12	-1	-1	-1.13	-1.14	-1.12	-1.08	-1
1.12	1.105	1.12	1	1	1.13	1.14	1.12	1.08	1

Note: In this table, more effective, the more serious environmental pollution; more invalid, the more light pollution.

Similarly, we use y3=Q as output of DEA model and use x1 = e^J , x2 = N, x3 = z as input of DEA model, then we establish the CCR model which based on output as model (5) and calculate the value of emission production efficiency in China each year, available Table 4:

Year	88	89	90	91	92	93	94	95	96	97
F2	-1	-1	-1	-1	-1.009	-1.003	-1.038	-1.067	-1.52	-1.120
θ_Q	1	1	1	1	1.009	1.003	1.038	1.067	1.52	1.121

98	99	00	01	02	03	04	05	06	07
-1.223	-1.039	-1	-1	-1	-1.039	-1.144	-1.035	-1	-1
1.223	1.039	1	1	1	1.039	1.144	1.035	1	1

Note: The same as Table 3, more effective, the more serious environmental pollution; more invalid, the more light pollution.

We use y4=S as output of DEA model and use x1=$_{e^J}$,x2=N ,x3=lnz as input of DEA model, then we establish the CCR model which based on output as model (5) and calculate the value of waste water production efficiency in China each year, available Table 5:

Year	88	89	90	91	92	93	94	95	96	97
F3	-1	-1.0405	-1	-1.074	-1.0123	-1.0826	-1.095	-1.114	-1.13	-1.107
θ_S	1	1.0486	1	1.074	1.0123	1.0826	1.095	1.114	1.13	1.107

98	99	00	01	02	03	04	05	06	07
-1.079	-1.044	-1.031	-1	-1	-1.013	-1.031	-1	-1.015	-1
1.079	1.044	1.031	1	1	1.013	1.031	1	1.015	1

Note: The same as Table 3, more effective, the more serious environmental pollution; more invalid, the more light pollution.

Above tables, F1, F2 and F3 are respectively the value of objective function, $\theta_G, \theta_Q, \theta_S$ are respectively the values of efficiency.

We use y2=GT as output of DEA model and use X1=J, X2=N, X3=z as input of DEA model, then we establish the original model and calculate the value of production efficiency of solid wastes in China each year, available Table 6:

Year	88	89	90	91	92	93	94	95	96	97
F11	-1	-1.01	-1	-1.06	-1.096	-1.21	-1.34	-1.36	-1.42	-1
θ_{1G}	1	1.02	1	1.06	1.096	1.21	1.34	1.36	1.49	1

98	99	00	01	02	03	04	05	06	07
-1.15	-1.13	-1.1	-1.01	-1	-1.16	-1.19	-1.11	-1.088	-1
1.19	1.18	1.14	1.04	1	1.16	1.19	1.14	1.09	1

We use y3=Q as output of DEA model and use X1=J, X2=N, X3=z as input of DEA model, then we establish the original model and calculate the value of emission production efficiency in China each year, available Table 7:

Year	88	89	90	91	92	93	94	95	96	97
F21	-1	-1	-1.004	-1	-1.02	-1.02	-1.05	-1.044	-1.428	-1.092
θ_{2Q}	1	1	1.04	1	1.02	1.02	1.05	1.046	1.448	1.092

98	99	00	01	02	03	04	05	06	07
-1.23	-1.048	-1	-1	-1	-1.027	-1.077	-1	-0.9996	-1
1.23	1.048	1	1	1	1.027	1.077	1	1	1

We use y4=S as output of DEA model and use X1=J, X2=N, X3=z as input of DEA model, then we establish the original model and calculate the value of waste water production efficiency in China each year, available Table 8:

Year	88	89	90	91	92	93	94	95	96	97
F31	-1	-1.0352	-1	-1.126	-1.099	-1.218	-1.222	-1.205	-1.191	-1.186
θ_{3S}	1	1.0484	1	1.126	1.099	1.218	1.222	1.222	1.226	1.186

98	99	00	01	02	03	04	05	06	07
-1.107	-1.053	-1.031	-1	-1	-1.036	-1.079	-1.008	-1.018	-1
1.107	1.053	1.031	1	1	1.036	1.079	1.02	1.018	1

Because the efficiency evaluation of waste discharge should be the reciprocal of the value of the efficiency θ of model (5), therefore, the reciprocal of the value of the efficiency θ in Table 3 to Table 8 is the value of the efficiency of waste emission. Values in Table 1 to Table 5 are calculated by the improved model, those in Table 6 to Table 8 are calculated by the original model. From the results, there are big differences: 1) a lot of efficiency values are different, as the efficiency value is 1.03 in 1992 in Table 3, but in Table 6, the efficiency value is 1.096 in 1992; 2) we use the improved model to get ineffective value, as the efficiency value 1.06 in 1991 and the efficiency value 1.04 in 2001 in Table 6, and we use the improved model to get effective value, as the efficiency value 1.00 in 1991 and the efficiency value 1.00 in 2001 in Table 3; 3) The improved model consider the function of the input variables and output variables, therefore, the improved model can better reflect the efficiency evaluation between outputs and inputs, the efficiency value of the improved model can be more realistic situation. In Table 5 using improved DEA model, the optimal efficiency of the wastewater discharge year is 1996, but with the original DEA model in Table 8 the optimal efficiency of wastewater discharge year is 1994. In fact, wastewater discharge in 1996 is less than in 1995 and 1997. It prove that the improved DEA model is more realistic.

3.2 Evaluation and Analysis of the Efficiency of the Production Environment of Multiple-Output, Multiple-Input Improved DEA Model

From the above analysis, we know that the relationship between the output variables and input variables are not always linear relationship, some of them are non-linear relationship. Therefore, multiple output, multiple input model, we use y1=GDP, y2=GT, y3=Q, y4=S as output of DEA model and use x1=e^J, x2=N, x3=lnz, X4=z, X5=Ne^J, x6=e^J lnz, x7=ze^J, X8=Nlnz, X9=Nz, X10=zlnz, X11=N^2, X12=$(e^J)^2$, X13=$(\ln z)^2$, X14=z^2 as input of DEA model, then we establish the CCR model which Based on the output of non-Archimedean infinitesimal. We calculate the value of production environment efficiency, available Table 9:

Year	88	89	90	91	92	93	94	95	96	97
Target value	-4	-4.0291	-4.1478	-4.1571	-4.0289	-4.1118	-4.2147	-4.2882	-4.7751	-4
θ_1	1	0.9915	1.0108	0.9811	0.9312	0.9045	0.8999	0.9241	0.9628	1
θ_2	1	1.0096	1.0358	1.064	1.0758	1.1617	1.269	1.3451	1.458	1
θ_3	1	1.0487	1.0541	1.1267	1.0498	1.1035	1.09	1.0718	1.0474	1
θ_4	1	0.9793	1.047	0.9852	0.9722	0.9421	0.9559	0.9472	1.3096	1
Year	98	99	00	01	02	03	04	05	06	07
Target value	-4.2636	-4.0946	-4.0267	-4	-4.0407	-4.1457	-4.1241	-4	-4.0266	-4
θ_1	0.909	0.9175	0.9545	1	0.8599	0.8268	0.9485	1	1.0097	1
θ_2	1.0573	1.0477	1.0511	1	1.1103	1.2963	1.0803	1	1.0405	1
θ_3	1.0983	1.0917	1.0469	1	1.0051	1.0105	1.0304	1	1.0101	1
θ_4	1.199	1.0378	0.9742	1	1.0654	1.0122	1.0649	1	0.9663	1

From the calculation results, the most efficient year include 1988, 1997, 2001, 2005 and 2007, we use θ_1 whose value reflects the output efficiency of China's GDP to evaluate output efficiency of y1 = GDP, the. We use θ_2, θ_3, θ_4 to evaluate the output efficiency of pollutants. So we use the reciprocal of them to evaluate the output efficiency of pollutants. From the results in Table 7, we know that the year with the best effectiveness of China's solid waste emissions is 1996, followed by 1995, then is 2003, 1994; the year with the best effectiveness of China's emissions is 1991, followed by 1993, then 1s 1998, 1999; the year with the best effectiveness of China's wastewater discharge is 1996, followed by 1998, then is 2002, 2004. From integrated pollution output efficiency, we know that the best year is 1996, followed by 1995, and then is 1998.

This is consistent quite well with the actual situation, GDP in 1996 has a larger growth compared to that in 1995; solid emissions in 1996 has a small growth compared to that in 1995, but has a larger growth compared to that in 1996; gas emissions in 1996 is smaller than that in 1995 and 1997. Therefore, from the actual data, the year with the best comprehensive efficiency is 1996.

4 Conclusion

Because of the linear nature of DEA model, there are many deficiencies in its application. In this paper we try to improve the original DEA model. From the corresponding results, the effect is very obvious. The improved model considers the function of the input variables and output variables, which can reflect the efficiency evaluation between the output and input better. the efficiency value of the improved model can be more realistic. The improved model is closer to the actual situation and can reflect the actual situation better. Of course, there are also some deficiencies in the improved model, we still need further work to improve it.

References

1. Ruggiero, J.: On the measurement of technical efficiency in the public sector. European Journal of Operational Research 90(3), 553–565 (1996)
2. Gupta, S., Verhoeven, M.: The efficiency of government expenditure-experiences from Africa. Journal of Policy Modeling 23(4), 433–467 (2001)
3. Antreas, D.A.: Using frontier efficiency models as a tool to re-engineer networks of public sector branches: an application to the Hellenic Tobacco Organization. European Journal of Operational Research 154(2), 533–547 (2004)
4. Afonso, A., Schuknecht, L., Tanzi, V.: Public sector efficiency: an international comparison. Public Choice 123(3-4) (2005)
5. Afonso, A., Fernandes, S.: Measuring local government spending efficiency: evidence for the Lisbon region. Regionnal Studies 40(1) (2006)
6. Zhang, N., Hu, A.G., Zheng, J.H.: Application of DEA method evaluation efficiency of the Regional Health. Economic Research Journal (7), 92–105 (2006) (In Chinese)
7. Liu, Q.Z.: Encourage independent innovation: the scale of public expenditure and Optimal. The Journal of Quantitative & Technical Economics (3), 81–90 + 101 (2007) (in Chinese)

8. Liu, Q.Z.: Public expenditure end: the Chinese government to implement public services to the poor hands of it? Management World (4) (2007) (in Chinese)
9. Xu, H., Sun, Y.H., Hua, Z.S.: An Interval DEA Method Based on Overall Efficiency. Chinese Journal of Management Science 18(2), 102–107 (2010) (in Chinese)
10. Wu, Q., Wu, C.Y.: Research on Evaluation Model of Energy Efficiency Based on DEA. Journal of Management Sciences 22(1), 103–112 (2009) (in Chinese)
11. Dong, L.: Energy Intensity Trend and Its Influertial Factors in China. Industrial Economics Research (1), 8–18 (2008) (in Chinese)
12. Wu, C.Y., Wu, Q.: Evaluation Model of Energy Efficiency Based on Super—Efficiency—DEA. Chinese Journal of Management 6(11) (2009) (in Chinese)
13. Liu, M.Z., Zhou, M.H., Yang, J.: Evaluation Model and Empirical Study of Urban Logistic Efficiency Based on DEA. Statistics and Decision (6), 50–52 (2009) (in Chinese)

Stable Fuzzy Adaptive Controller Design for Nonlinear Singularly Perturbed Systems*

Chenggong Sun and Li Li

College of Science, Tianjin University of Science & Technology, Tianjin 300457, China
scgtj@sina.com

Abstract. The controllers of the MIMO Nonlinear Singularly Perturbed Systems (NSPS) are usually developed based on the fast-slow decomposition and two-stage approach. It is necessary that the mathematical models of the plants are exactly known, which sometimes can't be achieved in the practical industry. We propose a new approach of the adaptive fuzzy control to solve the tracking problem of the NSPS with unknown models. Firstly, a direct adaptive fuzzy sub-controller with adjustable parameters is designed for the slow subsystem. The objective is to guarantee the slow states follow the expected trajectory. Then a fuzzy controller is designed as a fast sub-controller to make the fast subsystem stable. The final controller is the composition of the above mentioned ones. The simulation is given out to illustrate the effectiveness of the proposed method.

Keywords: Adaptive fuzzy control, TS model, nonlinear singularly perturbed systems, MIMO.

1 Introduction

The design of the controller for a class of physical plants with the slow and fast modes is relatively difficult because of the ill-conditioned and complexity computing problems. In the early decades, a lower-order model is chosen to approximate the high-order plant and then a finite order controller is proposed. Recently the singular perturbation theory and two-stage approach have been studied extensively in the numerous literatures [1]-[3]. The SP systems are firstly decomposed into the fast and slow subsystems, and then sub-controllers integrating optimal [2] or sliding control [3] themes are given. The final controller is the composition of the sub-controllers.

However, all of these approaches need the model of the SP systems exactly known in prior, which usually is not practical in the industrial field. Reza Langari constructs a framework incorporating Kalman Filter based optimal identification and hierarchically structured fuzzy controller for the nonlinear SP systems in 1996 [4]. He is the first person tried to introduce the fuzzy controller to the field of the controller design for the nonlinear SP systems. However, there is no a simulation example to illustrate the effectiveness of his idea. Lin[5] develops a two-time scale fuzzy logic controller for the

* This work was supported by the research fund of Tianjin University of Science & Technology (20080207).

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 695–701.
springerlink.com

flexible-link robot in 2002. The fast-subsystem controller can damp out the vibration of the flexible structure by an optimal control method and the slow-subsystem fuzzy controller can realize the trajectory tracking. Sun and his coworkers [6] have proposed the Fuzzy Singularly Perturbed Model (FSPM) via extending the ordinary TS fuzzy model. They have further developed H_∞ and H_2 [7] controllers. The stabilization analysis are given in [8].

Most of the above-mentioned methods need the dynamics of the fast subsystems known in prior and are only for the regulation problem. Therefore we propose an adaptive fuzzy controller based on the two-stage approach for a class of SISO SP systems[9]. In this paper, we extended the approach proposed in [9] to the control of the MIMO SP systems. An adaptive fuzzy sub-controller is designed to make the slow variables track the expected trajectory. A fuzzy sub-controller is developed guarantee the stability of the fast subsystem.

2 Problem Formulations

Suppose the MIMO SP system is as following:

$$\begin{aligned}
&[x_{11}^{(r_1)}(t)\cdots x_{1r_p}^{(r_p)}(t)]^T = \boldsymbol{F}_{11}(\boldsymbol{x}(t)) + \boldsymbol{F}_{12}(\boldsymbol{x}(t))\boldsymbol{z}(t) + \boldsymbol{G}_1(\boldsymbol{x}(t))\boldsymbol{u}(t) + \boldsymbol{d}(\boldsymbol{x},t)\\
&\varepsilon\dot{\boldsymbol{z}}(t) = \boldsymbol{F}_{21}(\boldsymbol{x}(t)) + \boldsymbol{F}_{22}(\boldsymbol{x}(t))\boldsymbol{z}(t) + \boldsymbol{G}_2(\boldsymbol{x}(t))\boldsymbol{u}(t)\\
&\boldsymbol{y}(t) = [y_1(t)\cdots y_p(t)]^T
\end{aligned} \tag{1}$$

where $\boldsymbol{x}(t) = [\boldsymbol{x}_1^T(t) \quad \cdots \quad \boldsymbol{x}_p^T(t)]^T$, $\boldsymbol{x}_i(t) = [x_{i1}(t) \quad \cdots \quad x_{ir_i}(t)]^T$ and $z(t) \in \boldsymbol{R}^{n_z}$ are the slow and fast variables of the system , respectively; n_z and r_i are the order of the fast and slow variable, respectively with $r_1 + r_2 + \cdots r_p = n$. $\boldsymbol{F}_{11}(\boldsymbol{x}(t)) = [f_1^1(\boldsymbol{x}_1(t))\cdots f_p^1(\boldsymbol{x}_p(t))]^T$, $\boldsymbol{F}_{12}(\boldsymbol{x}(t)) = [\boldsymbol{f}_1^2(\boldsymbol{x}_1(t))\cdots \boldsymbol{f}_p^2(\boldsymbol{x}_p(t))]^T$, $\boldsymbol{G}_1(\boldsymbol{x}(t)) = [\boldsymbol{G}_{11}^T(\boldsymbol{x}(t))\cdots \boldsymbol{G}_{1p}^T(\boldsymbol{x}(t))]^T$, $\boldsymbol{G}_{1i}(\boldsymbol{x}(t)) = [g_{i1}(\boldsymbol{x}(t))\cdots g_{ip}(\boldsymbol{x}(t))]^T$.

The control objective is to get the slow variables $\boldsymbol{x}(t)$ to track the specific states $\boldsymbol{x}_d(t)$ and all the signals are bounded. Define the tracking error as $e_i(t) = x_{i1}(t) - x_{di1}(t)$. The problem is to design a control law $\boldsymbol{u}(t)$, which ensures that $e_i(t) \to 0$, as $t \to \infty$ and the fast states $\boldsymbol{z}(t)$ are stable.

3 Controller Design

Let $\varepsilon = 0$ in (1), we have

$$\boldsymbol{F}_{21}(\bar{\boldsymbol{x}}(t)) + \boldsymbol{F}_{22}(\bar{\boldsymbol{x}}(t))\bar{\boldsymbol{z}}(t) + \boldsymbol{G}_2(\bar{\boldsymbol{x}}(t))\boldsymbol{u}_s(t) = \boldsymbol{0} \tag{2}$$

Assumption 1: $\boldsymbol{F}_{21}(\bar{\boldsymbol{x}}(t)) + \boldsymbol{F}_{22}(\bar{\boldsymbol{x}}(t))\bar{\boldsymbol{z}}(t) + \boldsymbol{G}_2(\bar{\boldsymbol{x}}(t))\boldsymbol{u}_s(t) = \boldsymbol{0}$ be solvable and the solution is

$$\bar{\boldsymbol{z}}(t) = \boldsymbol{h}_1(\bar{\boldsymbol{x}}(t)) + \boldsymbol{h}_2(\bar{\boldsymbol{x}}(t))\boldsymbol{u}_s(t) \tag{3}$$

where $\boldsymbol{h}_1(\bar{\boldsymbol{x}}(t)) = -\boldsymbol{F}_{22}^{-1}(\bar{\boldsymbol{x}}(t))\boldsymbol{F}_{21}(\bar{\boldsymbol{x}}(t))$, $\boldsymbol{h}_2(\bar{\boldsymbol{x}}(t)) = -\boldsymbol{F}_{22}^{-1}(\bar{\boldsymbol{x}}(t))\boldsymbol{G}_2(\bar{\boldsymbol{x}}(t))$.

Considering of (1) and (2), the slow subsystem is obtained.

$$\left[\bar{x}_{11}^{(r_1)}(t)\cdots\bar{x}_{1r_p}^{(r_p)}(t)\right]^T = \boldsymbol{F}(\bar{\boldsymbol{x}}(t)) + \boldsymbol{G}(\bar{\boldsymbol{x}}(t))\boldsymbol{u}_s(t) + \boldsymbol{d}(\bar{\boldsymbol{x}},t) \tag{4}$$

where $\boldsymbol{F}(\bar{\boldsymbol{x}}(t)) = \boldsymbol{F}_{11}(\bar{\boldsymbol{x}}(t)) - \boldsymbol{F}_{12}(\bar{\boldsymbol{x}}(t))[\boldsymbol{F}_{22}(\bar{\boldsymbol{x}}(t))]^{-1}\boldsymbol{F}_{21}(\bar{\boldsymbol{x}}(t))$

$\boldsymbol{G}(\bar{\boldsymbol{x}}(t)) = \boldsymbol{G}_1(\bar{\boldsymbol{x}}(t)) - \boldsymbol{F}_{12}(\bar{\boldsymbol{x}}(t))[\boldsymbol{F}_{22}(\bar{\boldsymbol{x}}(t))]^{-1}\boldsymbol{G}_2(\bar{\boldsymbol{x}}(t))$

To identify the fast subsystem, a fast time scale $\tau = t/\varepsilon$ and $z_f = z - z_s$ is designed. Suppose the slow variables are constant, we can get the fast subsystem:

$$\varepsilon\dot{\boldsymbol{z}}_f(t) = \boldsymbol{F}_{22}(\bar{\boldsymbol{x}}(t))\boldsymbol{z}_f(t) + \boldsymbol{G}_2(\bar{\boldsymbol{x}}(t))\boldsymbol{u}_f(t) \tag{5}$$

Assumption 2 and 3 are required before the design of the slow controller.

Assumption 2: $[\boldsymbol{F}_{22}(\bar{\boldsymbol{x}}(t)) \quad \boldsymbol{G}_2(\bar{\boldsymbol{x}}(t))]$ is stabiliable.

Assumption 3: Suppose $\boldsymbol{G}(\bar{\boldsymbol{x}}(t))$ is nonsingular; $\|\boldsymbol{G}(\bar{\boldsymbol{x}}(t))\|$ is bounded for all $\bar{\boldsymbol{x}}(t) \in S_x, t \geq 0$, where $\boldsymbol{S}_x \in \boldsymbol{R}^n$ is a compact set. That is, we suppose $\lambda_{\max}(\boldsymbol{G}(\bar{\boldsymbol{x}}(t))) < \infty$, $\lambda_{\min}(\boldsymbol{G}(\bar{\boldsymbol{x}}(t))) > 0$. $\lambda_{\max}(\boldsymbol{G}(\bar{\boldsymbol{x}}(t)))$ and $\lambda_{\min}(\boldsymbol{G}(\bar{\boldsymbol{x}}(t)))$ are the maximal and minimal eigenvalues of $\boldsymbol{G}(\bar{\boldsymbol{x}}(t))$, respectively. Suppose (4) is linearizible with the relative order $\boldsymbol{r} = [r_1 \cdots r_p]^T$ and the zero-dynamic is stable. Each entry of $\boldsymbol{G}(\bar{\boldsymbol{x}}(t))$ (except those on the main diagonal) is bounded by $\left|g_{ij}(\bar{\boldsymbol{x}}(t))\right| < \bar{g}_{ij}$, $i, j = 1,\cdots,p, i \neq j$. Those entries on the main diagonal satisfy $0 < \underline{g}_{ii} \leq g_{ii}(\boldsymbol{x}(t)) \leq \bar{g}_{ii} < \infty$ and $\sum_{j=1, j\neq i}^{p} \bar{g}_{ij} / \underline{g}_{ii} < 1$, $i = 1,\cdots,p$ and their derivatives satisfy $\left|\dot{g}_{ii}(\boldsymbol{x}(t))\right| \leq M_{ii}(\boldsymbol{x}(t))$, where $M_{ii}(X)$ is a known and bounded function.

Define the error is

$$e_{si}(t) = [k_0^i, \cdots, k_{r_i-2}^i, 1][e_{i1}(t), \cdots, e_{i1}^{(r_i-1)}(t)]^T \tag{6}$$

The track output vector is $\boldsymbol{y}_m(t) = [y_{m1}(t)\cdots y_{mp}(t)]^T$ and the coefficients of $1/\hat{L}(s) = 1/(s^{r_i-1} + k_{r_i-2}^i s^{r_i-2} + \cdots + k_0^i)$ should be picked so that the transfer functions are stable.

The adaptive fuzzy controller can be built by a TS fuzzy logic system and the l^{th} rule is of the form as

$$R^l: \text{IF } \bar{x}_{i1} \text{ is } A_{i1}^l, \text{ and } \ldots, \text{ and } \bar{x}_{ir_i} \text{ is } A_{ir_i}^l, \text{THEN } \hat{u}_{si}(t) = a_{i0}^l + \cdots a_{ir_i}^l \bar{x}_{ir_i}(t) \tag{7}$$

Using center-average defuzzifer, product inference, and singleton fuzzifier, the TS fuzzy model is

$$\hat{u}_{si}(t) = \xi_i(\bar{\boldsymbol{x}}_i(t))\theta_i \boldsymbol{X}_i(t) = \sum_{l=1}^{m} \xi_i^l(\bar{\boldsymbol{x}}_i(t))\boldsymbol{X}_i^T(t)\theta_i^l \tag{8}$$

where $\xi_i(\bar{\boldsymbol{x}}_i(t)) = [\xi_1^l(\bar{\boldsymbol{x}}_i(t))\cdots\xi_1^m(\bar{\boldsymbol{x}}_i(t))] \in \boldsymbol{R}^{1\times m}$, $\theta_i = [\theta_i^1 \cdots \theta_i^m]^T \in \boldsymbol{R}^{m\times(r_i+1)}$, $\theta_i^l = [a_{i0}^l \cdots a_{ir_i}^l]^T \in \boldsymbol{R}^{r_i+1}$, $\boldsymbol{X}_i(t) = [1, \bar{x}_{i1}(t)\ldots\bar{x}_{ir_i}(t)]^T \in \boldsymbol{R}^{r_i+1}$.

Suppose the i^{th} ideal control law is expressed as

$$u_i^*(t) = \xi_i(\bar{\boldsymbol{x}}_i(t))\theta_i^* \boldsymbol{X}_i(t) = \sum_{l=1}^{m} \xi_i^l(\bar{\boldsymbol{x}}_i(t))\boldsymbol{X}_i^T(t)\theta_i^{*l} + \varepsilon_i(\bar{\boldsymbol{x}}_i(t)) \tag{9}$$

where $\varepsilon_i(\bar{\boldsymbol{x}}_i(t))$ is the reconstruction error, θ_i^* is the optimal approximation parameters, which is defined as $\theta_i^* = \arg\min_{\theta_i\in\Omega_i}[\sup_{\bar{\boldsymbol{x}}_i(t)\in S_x, t\geq 0}\left|\xi_i(\bar{\boldsymbol{x}}_i(t))\theta_i^* X_i(t) - u_i^*(t)\right|]$, where Ω_i is a compact set with $\theta_i \in \Omega_i$.

Assumption 4: Suppose the unknown external disturbance $d_i(\bar{\boldsymbol{x}}(t), t) < D_i$, and the reconstruction error $\varepsilon_i(\bar{\boldsymbol{x}}_i(t)) \leq D_{\varepsilon i}(\bar{\boldsymbol{x}}_i(t))$, $\left|\xi_i(\bar{\boldsymbol{x}}_i(t))\boldsymbol{X}_i^T(t)\Phi_i(t)\right| \leq \bar{U}_i(\bar{\boldsymbol{x}}(t))$. Since $\xi_i(\bar{\boldsymbol{x}}_i(t)), \boldsymbol{X}_i^T(t)$ is known, such $\bar{U}_i(\bar{\boldsymbol{x}}(t))$ can be attained.

Then the adaptive fuzzy controller is

$$\boldsymbol{U}_s(t) = \hat{\boldsymbol{U}}_s(t) + \boldsymbol{U}_d(t) \tag{10}$$

$$\hat{\boldsymbol{U}}_s(t) = [\hat{u}_{s1}(t), \cdots, \hat{u}_{sp}(t)]^T, \boldsymbol{U}_d(t) = [u_{d1}(t)\cdots u_{dp}(t)]^T \tag{11}$$

$$u_{di}(t) = \operatorname{sgn}(e_{si}(t))(\sigma_i + \sum_{j=1, j\neq i}^{p} \bar{g}_{ij} U_{\max} / \underline{g}_{ii}) + \rho_i \tag{12}$$

$$\sigma_i = \sum_{j=1}^{p} \bar{g}_{ij} D_j(\bar{\boldsymbol{x}}_i(t)) / \underline{g}_{ii} + \sum_{j=1}^{p} \bar{g}_{ij} D_{\varepsilon j}(\bar{\boldsymbol{x}}_i(t)) / \underline{g}_{ii} + \sum_{j=1, j\neq i}^{p} \bar{g}_{ij}\bar{U}_j / \underline{g}_{ii} \tag{13}$$

$$\rho_i = 0.5 M_{ii}(\bar{\boldsymbol{x}}(t)) e_{si}(t) / g_{ii}^2(\bar{\boldsymbol{x}}(t)) \tag{14}$$

We required $\left|u_{di}(t)\right| \leq U_{\max}$, that is:

$$\left|u_{di}(t)\right| \leq \left|\sigma_i\right| + \left|\rho_i\right| + U_{\max} \sum_{j=1, j\neq i}^{p} \bar{g}_{ij} / \underline{g}_{ii} \leq U_{\max} \tag{15}$$

That is we need

$$U_{\max} \geq \max_{i=1,\cdots,p}[(\left|\sigma_i\right| + \left|\rho_i\right| / (1 - \sum_{j=1, j\neq i}^{p} \bar{g}_{ij} / \underline{g}_{ii}))] \tag{16}$$

The adaptive law is

$$\dot{\theta}_i = \Gamma_{ui}\xi_i^T(\bar{x}_i(t))\boldsymbol{X}_i^T(t)e_{si}(t) \tag{17}$$

where $\Gamma_{ui} \in \boldsymbol{R}^{m\times m}$ is a positive define diagonal matrix, representing the adaption gain. Projection algorithm can be employed to guarantee $\theta_i \in \Omega_i$.

Theorem: Suppose the reference trajectories $\boldsymbol{y}_m(t)$ and its derivation is bounded and the plant satisfies Assumption 3 and 4, using control law (10)-(16) and adaptive law (17), we can get:1)The states variables and output vector and their derivation are bounded. 2)The controller is bounded, that is $\|\boldsymbol{U}_s(t)\| \in L_\infty$. 3) The error vector decrease at least asymptotically to zero, that is $\lim_{t\to\infty} e_i(t) = 0$, $i = 1,\cdots, p$.

Here the proof is omitted. We suppose the slow variables are constant, the fast subsystem (5) is a linear system, which can be stabilized by many approaches. Since the model of the fast subsystem is unknown, a fuzzy controller is designed. The final control law is

$$\boldsymbol{u}(t) = \boldsymbol{u}_s(t) + \boldsymbol{u}_f(t) \tag{18}$$

5 Simulation

The adaptive fuzzy controller is applied in this section to make a nonlinear system follow an expected trajectory. The results of the simulation are shown as Fig.1-Fig.2.

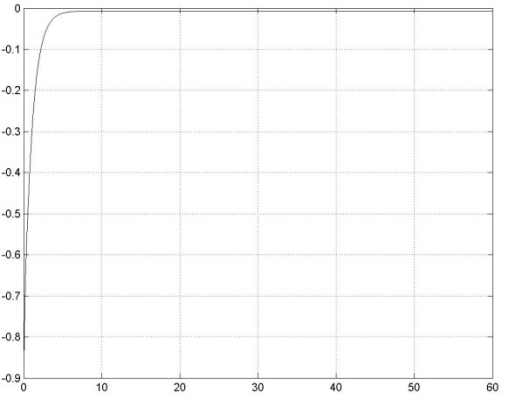


Fig. 1. The tracking error $e_{11}(t)$ between state x_{11} and its desired value $y_{m1}(t) = e^{-t}$

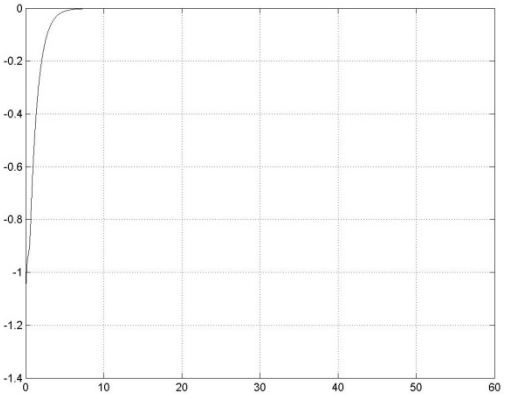


Fig. 2. The tracking error $e_{21}(t)$ between state x_{21} and its desired value $y_{m2}(t) = e^{-t}$

We can see that the dynamic tracking error is convergent to zero after 10 seconds and the fast state is bounded to a small range. The reason is that when the mathematic model of the SP system is unknown, the sub-controllers are the adaptive fuzzy and fuzzy controllers, respectively. Though there is approximation error of the SP system, the adaptive fuzzy controller can reduce it by adjusting the parameters.

6 Conclusion

A kind of composite controllers are presented to the MIMO nonlinear SP systems in this paper. Even if the mathematical model of the plant is unknown, it can still achieve better tracking performance since adaptive fuzzy controller is used. The simulation of a nonlinear MIMO SP system illustrates the performance.

References

1. Gajic, Z., Lim, M.-T.: Optimal control of singularly perturbed linear sytems and applications-hign accuracy techniques. Marcel Dekker, Inc., New York (2001)
2. Chow, J.H., Kokotovic, P.V.: A decomposition of Near-Optimum Regulators for Systems with Slow and Fast Modes. IEEE Transactions on Automatic Control 21(5), 701–705 (1976)
3. Zhang, X., Jin, H., Zhang, H., Li, G., Ji, M.: Robust Sliding Mode Control for a Class of Uncertain Nonlinear Singularly Perturbed Systems. SICe 2wz, August 5-7, pp. 2616–2621 (2002)
4. Langari, R., Li, W.: Analysis and Efficient Implementation of Fuzzy Logic Control Algorithms. In: 1996 Biennial Conference of the North American, June 19-22, pp. 1–4 (1996)
5. Lin, J., Lewis, F.L.: Fuzzy controller for flexible-link robot arm by reduced-order techniques. IEE Proc.-Control Theory &Application 147(3), 177–187 (2002)
6. Liu, H.-P., Sun, F.-C., et al.: Controller design and stability analysis for fuzzy singularly perturbed systems. Acta Automatic Sinica 29(4), 494–500 (2003)

7. Liu, H.-P., Sun, F.-C., et al.: H $_2$ state feedback Control for fuzzy singularly perturbed systems. In: Proceedings of the 42nd IEEE Conference on Decision and Control, Maul, Hawali USA, pp. 5239–5243 (2003)
8. Liu, H.-P., Sun, F.-C., et al.: Simultaneous Stabilization for singularly perturbed systems via linear matrix inequalities. Acta Automatic Sinica 30(1), 1–7 (2004)
9. Li, L., Sun, F.-C.: An adaptive fuzzy controller for nonlinear singularly perturbed systems. In: Proceedings of the IMACS Multi Conference on Computational Engineering in Systems Applications, pp. 1388–1394 (2006)

Resolution Performance of Modulated Acoustic Signal Inspired by Bat

Juqin Tan, Jian Yang, and Hejuan Chen

School of Mechanical Engineering, Nanjing University of Science and Technology,
200 Xiaolingwei, Nanjing, 210094, P.R. China
tanjvqin@126.com, {yang4446306,hjchenzj}@hotmail.com

Abstract. For smart ultrasonic detecting system of underwater weapons, there is Doppler effect cause by moving target for echo's complex envelope. The broad band ambiguity functions of symmetry triangle linear frequency modulation impulse signal, distance and velocity are derived by the definition of broad band ambiguity function. The symmetry characteristic of symmetry triangle linear frequency modulation impulse signal broad band ambiguity function is described. With broad band ambiguity function, the phenomena of the coupling echo from moving target for distance and velocity as well as wave distortion are explained. By analyzing the distance broad band ambiguity function, the conclusion is gained as that the distance theoretical resolution is inversely proportional to band width. Furthermore, the velocity resolution is correlative to velocity, time width, band width and carrier wave frequency, so that the theoretical foundation of parameters selection by the demand of engineering is put forward.

Keywords: acoustic detection, symmetry triangle linear frequency modulation signal, broad band ambiguity function, resolution.

1 Introduction

The underwater detection systems commonly use ultrasonic waves to detect target. In order to detect target accurately and fast for detectors, the emitting signals must have board band width to gain better distance and velocity resolution. The batty wide wave modulated frequency FM signal is this type of signals [1], and LFM signal [2] [3] is one kind of similar batty FM signals. The ambiguity function of single slope LFM signal is researched in detail by many literatures [4] [5]. For symmetry triangle LFM signal, the references [6] and [7] respectively analyzed the single and multiple periods ambiguity functions of symmetry triangle LFMCW signals on the condition of massive product of time width and band width. Because the time width of symmetry triangle linear frequency modulation impulse (STLFMI) signal is far less than the maximum time delays of target echo waves, the manner of target information pick-up is different from symmetry triangle LFMCW. Thus its' ambiguity function has another characters to symmetry triangle LFMCW signal. The reference [8]

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 703–710.
springerlink.com

detailedly introduced the ambiguity function characters of symmetry triangle frequency modulating impulse compress signals in narrowband models of radio fuse detecting target. Due to sound velocity c much less than electromagnetic wave velocity, the echo wave complex envelope Doppler effects of target relative movements can't be ignored for high velocity moving targets and the narrowband conditon of $v<c/(8TB)$ [9] (T is time width, B is band width, c and v are sound velocity in medium and target relative radial velocity) is unable to come into existance. Hence the narrowband ambiguity function disagrees with moving target echo wave information analysis for underwater dectection in this paper. The 2-dimensional broad band ambiguity fuction is established with target moving parameters and the resolution of detector to two adjacent targets by simulation, which supports the foundations for design of generator to produce optimal signal waves.

2 STLFMI Signal Broad Band Ambiguity Function

On the assumption that the target is single scatter dot model, the conclusion also can be applied to multiple scatter dots target models. The broad band ambiguity function [9] is that

$$\chi_s(\tau,\xi)=\sqrt{s}\int_{-\infty}^{+\infty}u_c^*(t+\tau)\cdot u_c(st)\cdot\exp(j2\pi\xi t)dt \tag{1}$$

In above formula, $s=1-\beta=(c-v)/(c+v)$ is Doppler compress factor, $\beta=2v/(c+v)$, $u_c(t)$ is the complex envelope of signal, f_0 is carrier wave frequency, subscript s shows the effect of broad band STLFMI signal ambiguity function by Doppler compress factor.

2.1 STLFMI Signal

Suppose that the up scanning frequency segment of STLFMI signal is $[-T,0]$ and down scanning frequency segment is $[0,T]$, then the complex envelope is

$$u(t)=\begin{cases}u_1(t)+u_2(t), & -\mathrm{T}<\mathrm{t}<\mathrm{T}\\ 0, & others\end{cases} \tag{2}$$

The complex envelopes $u_1(t)$ and $u_2(t)$ of up/down scaning frequency segments for symmetry triangle frequent modulation rectangular impulse signal is as follows

$$u_1(t)=\begin{cases}(1/\sqrt{2T})rect(t/2T)\exp(j\pi Mt^2) & -\mathrm{T}<\mathrm{t}<0\\ 0 & others\end{cases} \tag{3}$$

$$u_2(t)=\begin{cases}(1/\sqrt{2T})rect(t/2T)\exp(-j\pi Mt^2) & 0<\mathrm{t}<\mathrm{T}\\ 0 & others\end{cases} \tag{4}$$

2.2 Derivation of STLFMI Signal Broad Band Ambiguity Function

2.2.1 STLFMI Signal Up/Down Scanning Frequency Segment Self Broad Band Ambiguity Functions

Put formula (3) into formula (1), then the broad band ambiguity function of $u_1(t)$ can be gained as

$$\chi_{u_{11}s}(\tau,\xi)\approx\frac{\sqrt{s}}{4T\sqrt{|M\beta|}}e^{j\pi(\frac{\xi^2}{2M\beta}-M\tau^2)}\{[C(x_2)-C(x_1)]+\mathrm{sgn}(\xi)\cdot j[S(x_2)-S(x_1)]\} \tag{5}$$

C(x) and S(x) are Fresnel integrations, and $\mathrm{sgn}(\xi)=\begin{cases}1, & \xi>0\\ -1, & \xi<0\end{cases}$, $x_2=2\sqrt{|M\beta|}\,[b(\tau)-\xi/(2M\beta)]$, $x_1=2\sqrt{|M\beta|}\,[a(\tau)-\xi/(2M\beta)]$.

(1) When $s>1$

$$\begin{cases}a(\tau)=-T-\tau, b(\tau)=0 & -T<\tau<-T(1-1/s)\\ a(\tau)=-T/s, b(\tau)=0 & -T(1-1/s)<\tau<0\\ a(\tau)=-T/s, b(\tau)=-\tau & 0<\tau<T/s\end{cases} \tag{6}$$

(2) When $s<1$

$$\begin{cases}a(\tau)=-T-\tau, b(\tau)=0 & -T<\tau<0\\ a(\tau)=-T-\tau, b(\tau)=-\tau & 0<\tau<-T(1-1/s)\\ a(\tau)=-T/s, b(\tau)=-\tau & -T(1-1/s)<\tau<T/s\end{cases} \tag{7}$$

The broad band ambiguity function of $u_2(t)$ is alike as

$$\chi_{u_{22}s}(\tau,\xi)\approx\frac{\sqrt{s}}{4T\sqrt{|M\beta|}}e^{j\pi(M\tau^2-\frac{\xi^2}{2M\beta})}\{[C(x_4)-C(x_3)]-\mathrm{sgn}(\xi)\cdot j[S(x_4)-S(x_3)]\} \tag{8}$$

In above formula, $x4=2\sqrt{|M\beta|}\,[b(\tau)+\xi/(2M\beta)]$, $x3=2\sqrt{|M\beta|}\,[a(\tau)+\xi/(2M\beta)]$。

(1) When $s>1$

$$\begin{cases}a(\tau)=-\tau, b(\tau)=T/s & -T/s<\tau<0\\ a(\tau)=0, b(\tau)=T/s & 0<\tau<T(1-1/s)\\ a(\tau)=0, b(\tau)=T-\tau & T(1-1/s)<\tau<T\end{cases} \tag{9}$$

(2) When $s<1$

$$\begin{cases}a(\tau)=-\tau, b(\tau)=T/s & -T/s<\tau<T(1-1/s)\\ a(\tau)=-\tau, b(\tau)=T-\tau & T(1-1/s)<\tau<0\\ a(\tau)=0, b(\tau)=T-\tau & 0<\tau<T\end{cases} \tag{10}$$

2.2.2 STLFMI Signal Up/Down Scanning Frequency Segment Cross Broad Band Ambiguity Functions

The cross broad band ambiguity function $\chi_{u1u2s}(\tau,\xi)$ for $u_1(t)$ and $u_2(t)$ is that

$$\chi_{u_{1}u_{2}s}(\tau,\xi)=\sqrt{s}\int_{-\infty}^{+\infty}u_{1c}(st)u_{2c}^{*}(t+\tau)\exp(j2\pi\xi t)dt=\frac{1}{4T\sqrt{|M|}}e^{j\pi M\tau^2-\frac{j\pi\xi^2}{2Ms}}\{[C(x_6)-C(x_5)]+j[S(x_6)-S(x_5)]\} \tag{11}$$

In above formula, x6=$2\sqrt{|Ms|}$ [b(τ)+ ξ /(2Ms)], x5=$2\sqrt{|Ms|}$ [a(τ)+ ξ /(2Ms)].

(1) When s>1

$$\begin{cases} a(\tau)=-\tau, b(\tau)=0 & 0<\tau<T/s \\ a(\tau)=-T/s, b(\tau)=0 & T/s<\tau<T \\ a(\tau)=-T/s, b(\tau)=T-\tau & T<\tau<T(1+1/s) \end{cases} \tag{12}$$

(2) When s<1

$$\begin{cases} a(\tau)=-\tau, b(\tau)=0 & 0<\tau<T \\ a(\tau)=-\tau, b(\tau)=T-\tau & T<\tau<T/s \\ a(\tau)=-T/s, b(\tau)=T-\tau & T/s<\tau<T(1+1/s) \end{cases} \tag{13}$$

2.2.3 STLFMI Signal Broad Band Ambiguity Functions

On the foundation of ambiguity function combining characters [10]

$$\chi_u(\tau,\xi)=\chi_{u_{11}}(\tau,\xi)+\chi_{u_{22}}(\tau,\xi)+\chi_{u_1u_2}(\tau,\xi)+e^{-j2\pi\xi\tau}\chi^*_{u_1u_2}(-\tau,-\xi) \tag{14}$$

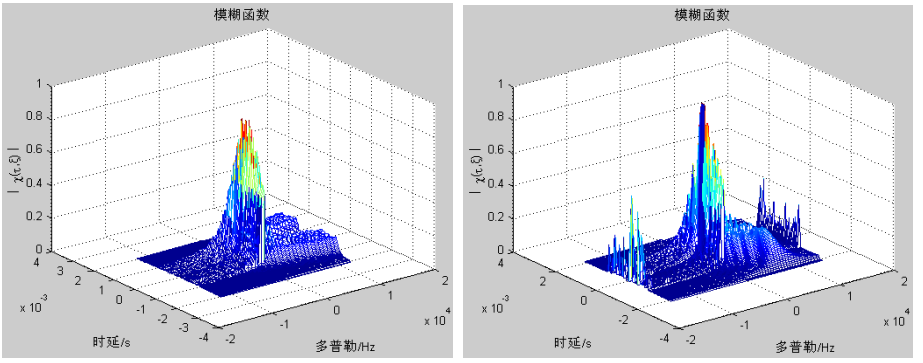
The STLFMI signal broad band ambiguity function can be obtained as

$$\chi_{us}(\tau,\xi)=\begin{cases} e^{-j2\pi\xi\tau}\chi^*_{u_1u_2s}(-\tau,-\xi) \\ \chi_l(\tau,\xi)+e^{-j2\pi\xi\tau}\chi^*_{u_1u_2s}(-\tau,-\xi) \\ \chi_{u_1s}(\tau,\xi)+\chi_{u_{22}s}(\tau,\xi)+e^{-j2\pi\xi\tau}\chi^*_{u_1u_2s}(-\tau,-\xi) \\ \chi_{u_1s}(\tau,\xi)+\chi_{u_{22}s}(\tau,\xi)+\chi_{u_1u_2s}(\tau,\xi) \\ \chi_m(\tau,\xi)+\chi_{u_1u_2s}(\tau,\xi) \\ \chi_{u_1u_2s}(\tau,\xi) \end{cases} \tag{15}$$

When s>1, χ_l isχ_{u11s} and χ_m is χ_{u22s}, $\tau\in$(-T(1+1/s),-T)$\cup$(-T, -T/s)$\cup$(-T/s,0)$\cup$(0, T/s)$\cup$(T/s,T)$\cup$(T,(1+1/s)T) ; when s<1, χ_l is χ_{u22s} and χ_m is χ_{u11s}, $\tau\in$(-T(1+1/s),-T/s)$\cup$(-T/s,-T)$\cup$(-T,0)$\cup$(0,T)$\cup$(T,T/s)$\cup$(T/s,(1+1/s)T).

2.3 Characters of STLFMI Signal Broad Band Ambiguity Function

The analysis of STLFMI signal broad band ambiguity function formulas shows that, $\chi_{us}(\tau,\xi)$ is the 2-dimemsional function of target moving parameter β, which is affected by moving target as well as signal time width T, band width B and carrier wave frequency f_0. These ingredients will bring the different integration limits for each integrates item. The values of T and B are not bigger enough in this detection environment, that makes the broad band width product of TB become small, the broad band ambiguity function turns to asymmetry and the volume of broad band ambiguity function is variable [9]. The STLFMI signal broad band ambiguity functions with different velocities are shown in figure 1. From the energy distribution, the max value is at the original point. The STLFMI signal output by receiver has distance coupling velocity and distinct multiple peak phenomena as target's velocity grows higher, which made the capability loss for target detection and resolution decreased.



(a)β＝0.02(v=15m/s) (b)β＝0.065(v=50m/s)

Fig. 1. Broad band ambiguity function of STLFMI signal in different velocities

3 Resolution Characters of Distance and Velocity for STLFMI Signal

3.1 Distance Resolution

When ξ=0, $\chi_{us}(\tau,0)$ is the distance broad band ambiguity function and its' formula is

$$\chi_{us}(\tau,0)=\begin{cases}\dfrac{1}{2T\sqrt{|M|}}e^{-0.5j\pi M\tau^2}[C(x_1)-jS(x_1)] & -2T<\tau<-T\\ \dfrac{1}{T}\{e^{j\pi M\tau T}\{\dfrac{\sin[\pi M\tau(T-|\tau|)]}{\pi M\tau(T-|\tau|)}(T-|\tau|)\} & \\ +\dfrac{1}{2\sqrt{|M|}}e^{-0.5j\pi M\tau^2}[C(x_2)-jS(x_2)]\} & -T<\tau<0\\ \dfrac{1}{T}\{e^{j\pi M\tau T}\{\dfrac{\sin[\pi M\tau(T-|\tau|)]}{\pi M\tau(T-|\tau|)}(T-|\tau|)\} & \\ +\dfrac{1}{2\sqrt{|M|}}e^{0.5j\pi M\tau^2}[C(x_3)+jS(x_3)]\} & 0<\tau<T\\ \dfrac{1}{2T\sqrt{|M|}}e^{0.5j\pi M\tau^2}[C(x_4)+jS(x_4)] & T<\tau<2T\end{cases} \tag{16}$$

In above formula, $x_1=2\sqrt{|M|}[T+(\tau/2)]$, $x_2=-\tau\sqrt{|M|}$, $x_3=\tau\sqrt{|M|}$, $x_4=2\sqrt{|M|}[T-(\tau/2)]$. The distance nominal resolution parameter (width is computed as -3dB) $\Delta\tau\approx0.44/B$, the minimum distance resolution corresponded to $\Delta\tau$ is as follows

$$\Delta R=c\Delta\tau/2=0.44c/2B \tag{17}$$

The formula (17) shows that, the distance inherent resolution of broad band STLFMI signal is inversely proportional to band width. The narrowband STLFMI signal distance resolution can be gained by the same manner and has the uniform result as formula (17). So the target movement has no effect to wave time resolution.

3.2 Velocity Resolution

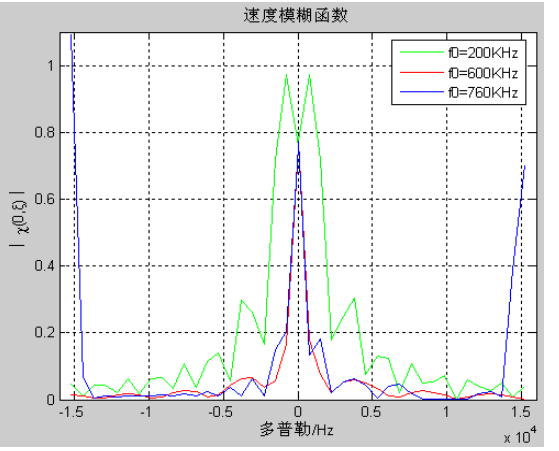
When τ=0, the velocity broad band ambiguity function can be expressed as

$$\chi_{us}(0,\xi)\approx\frac{\sqrt{s}}{4T}\cdot\frac{1}{\sqrt{|M\beta|}}\cdot\mathrm{sgn}\xi\{e^{j\pi\frac{\xi^2}{2M\beta}}\{[C(x_2)-C(x_1)]+\mathrm{sgn}\xi\cdot j[S(x_2)-S(x_1)]\}+e^{-j\pi\frac{\xi^2}{2M\beta}}\{[C(x_2)-C(x_1)]-\mathrm{sgn}\xi\cdot j[S(x_2)-S(x_1)]\}\} \tag{18}$$

In above formula, $x_2=2\sqrt{|M\beta|}\,[T+\xi/(2M\beta)](s<1)$ 或 $x_2=2\sqrt{|M\beta|}\,[-T/s+\xi/(2M\beta)](s>1)$, $x_1=2\sqrt{|M\beta|}\,[\xi/(2M\beta)]$.

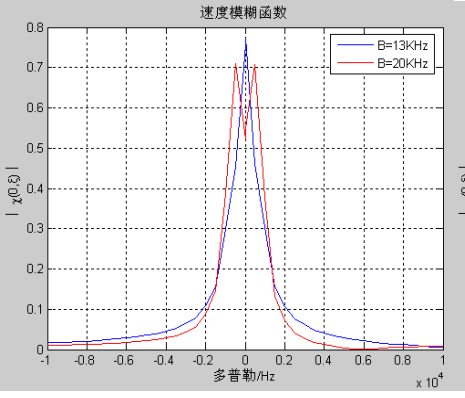
Usually the Fresnel integrations have no analytical solution, the relationship of velocity resolution with β, T, B and f_0 can not be made out from above formula. The figure 2 is the velocity broad band ambiguity function graphs with velocity resolution varying to β, T, B and f_0 as an example of water sonic signal by simulation method. The simulation results show that, when there is relative moving, the broad band STLFMI signal velocity resolution has following characters as

(1) When β, T *and* B are constants, the velocity resolution grows with f_0, but as f_0 increases to some value it will decrease, which is shown in figure 2(a).

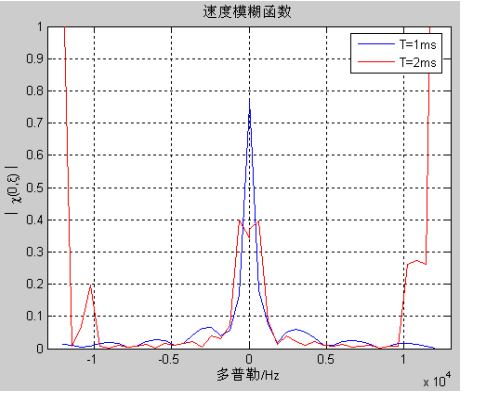


(a)TB=13,β=0.02(v=15m/s)

(2) When β、f_0[11] are constants, the velocity resolution decreases with B and T increase, which are shown in figure 2(b) and (c).

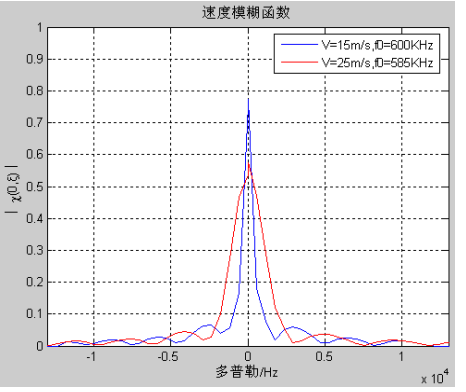


(b)β=0.02(v=15m/s),f_0=500kHz

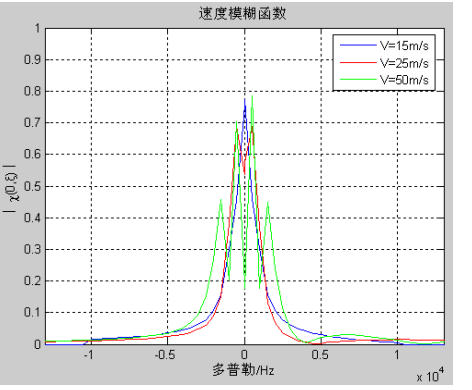


(c)β=0.02(v=15m/s),f_0=500kHz

(3) When T and B are constants, the carrier wave frequency f_0 of emitting signal for detecting high velocity target is lower to detecting low velocity target and the velocity resolution decreases with velocity increase, which is shown in figure 2(d).
(4) The figure 2(e) shown that, the Doppler frequency spectrum peak grows wide until multiple peaks appear with velocity increase and its' velocity resolution characters turn worse.



(d) TB=13



(e) TB=13, f_0=500kHz

Fig. 2. Velocity broad band ambiguity function of STLFMI signal in different conditions

4 Conclusion

According to the definition of broad band ambiguity function, the STLFMI signal, distance and velocity broad band ambiguity functions are derived. The STLFMI signal broad band ambiguity function has no symmetry because of different integrating limit for up/down scanning frequency segments. The moving target echo waves have phenomena of distance coupled velocity and wave distortion in the broad band ambiguity function point of view. The distance resolution is inversely proportional to band width by theoretical analysis of distance broad band ambiguity function. The velocity broad band ambiguity function is calculated by computer simulating method and results show that, this signal velocity resolution is relative to velocity, time width, band width and carrier frequency f_0, which decreases with velocity, time width and band width increase. The velocity resolution increases with f_0 increase, but when f_0 grows to some value, it will decrease again. These properties are valuable to use STLFMI signal and research new mode signal and could be the guide for the designing of "high resolution ratio emitting signal".

References

1. Feng, J.: Comparison Study on Some Batty Echolocation Signals and Morphological Features (Ph. D. Dissertation). Northeast Normal University, Changchun (2000) (in Chinese)
2. Oelze, M.L.: Bandwidth and resolution enhancement through pulse compression. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 54, 768–781 (2007)

3. Gerlach, K., Blunt, S.D.: Radar Pulse Compression Repair. IEEE Transactions on Aerospace and Electronic Systems 43, 1188–1195 (2007)
4. Zejak, A.J., Simic, I.S., Zrnic, B.M.: Chirp radar ambiguity function shaping. In: International Conference on Trends in Communications, EUROCON 2001, vol. 2, pp. 325–328 (2001)
5. Levanon, N., Mozeson, E.: Radar signals. John Wiley & Sons, Inc., Hoboken (2004)
6. Zhang, R., Yang, J., Xiong, J., Xiang, J.: Analysis of Ambiguity Functionof Symmetrical Triangular Linear Frequency Modulation Continuous Wave Signal. Acta Electronica Sinica 32, 353–356 (2004) (in Chinese)
7. Wu, L., Peng, S., Xiao, Z., Shi, X.: Multi-period Ambiguity Function Analysis of Symmetrical Triangle Linear Frequency Modulation Continuous Wave Signals. Journal of Nanjing University of Science and Technology (Natural Science) 33, 74–78 (2009) (in Chinese)
8. Zhou, X., Zhao, H., Liu, J., Gao, Z.: A Research on Ambiguity Function of Symmetrical Triangular Frequency Modulated Pulse Compression Signal. Journal of Detection & Control 31, 22–27 (2009) (in Chinese)
9. Lin, Y.: Active Sonar Detecting Information Theory. Ocean Press, Beijing (1990) (in Chinese)
10. Lin, M., Ke, Y.: Radar Signal theory. National Defence Press, Beijing (1984) (in Chinese)
11. Duan, Y.: Underwater Ultrasonic Transducer Array Signal Design (M. S. Dissertation). Nanjing University of Science and Technology, Nanjing (2006) (in Chinese)

Research on Pressure Control Method of Airport Pipeline Fueling System Based on Self-adaptive Fuzzy Control

Fanqin Meng, Libo Xi, and Guanghui Geng

POL and material Department of Xuzhou airforce college, 64 Xigejie Rd. XuZhou, china
mengfanqin@sohu.com, {Xilibo91,ggh76842}@163.com

Abstract. Pressure is an important parameter in the airplane fueling process. It is difficult for a traditional control method to keep pressure of airport pipeline fueling outlet constant in the large fueling flow and long pipeline. An adaptive fuzzy control method of airport pipeline pressure control is presented in this paper. A controller is designed and it is tested by simulation methods. It is proved that the fueling outlet pressure stability can be improved.

Keywords: Airport pipeline fueling, Fuzzy adaptive control, pressure control.

Airport pipeline refueling system belongs to liquid transportation system. The dynamic feature of the transportation system with characteristics of tight, large flow and complicated pipeline is difficult to express in mathematic model. It is more necessary to control the refueling pressure as increasing refueling flow and pressure. But it is too more difficult to control airport refueling system pressure because of its more conspicuous non-linear features with increasing pressure and flow. Self-adaptive fuzzy control method is free from the non-linear and time delay of plant and is not dependent on decisive mathematical model, so it is now in widespread use [1]. In this paper, self-adaptive fuzzy control methods is used in pressure control of airport pipeline fueling system, and the control accuracy and pressure stability was improved.

1 Airport Pipeline Fueling Mode and Control System Constitution

In the process of airport pipeline fueling, the fueling flow is changing with the number of fueling airplane at the same time. Airport pipeline fueling system has three modes to meet the changing fueling flow, first is parallel pump mode, second is single variable speed pump mode, and third is the combination of parallel pump and variable speed pump. But Third mode is now in widespread use. In this paper, three adjustable speed pumps are parallel connected to meet the changing flow and keep constant fueling pressure. The adjust process is as follows. Before refueling, a certain pipeline pressure is kept constant, when a fueling outlet is opened, the pipe pressure will decrease, the pressure change is taken by pressure transducer and is read by a IPC through a AD converter. The control system starts 1# pump to ignite a fueling process. The rotational velocity of the pump will keep up with the increasing fueling flow. If the rotational velocity of the pump has turned for 5~10S at rated value and the pressure of fueling position is still less than set value, 1# pump will be switched to power frequency

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 711–716.
springerlink.com

(50Hz) and 2# pump will be started by frequency converter. When two pumps work on rated rotational velocity and the pressure can be not get to set value, the third pump(3# pump) will be stared. On the contrary, when a fueling outlet is closed and the pressure will increase, the rotational velocity of pump will decrease. When the rotational velocity is less 5% of the rated rotational velocity of pump and the pressure of fueling position is still more than set value, a pump will be stopped. The constitution of airport constant pressure fueling system is shown in fig. 1.

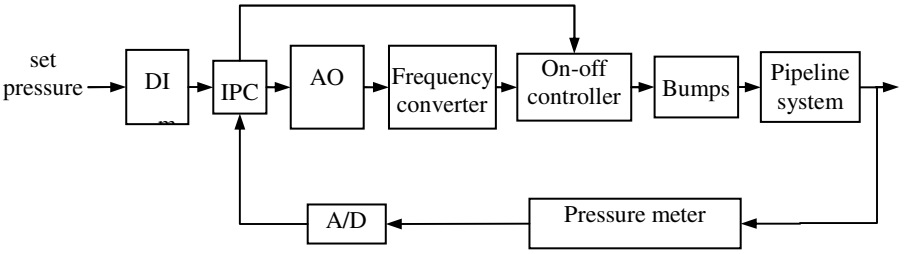


Fig. 1. Airport constant pressure fueling system

2 The Constitution of Fuzzy Adaptive Control System

Model reference adaptive system (MRAS) is a important adaptive control system, is used is many situation with its high adaptive calculation speed. In MRAS, a given performance indices set is replaced by a performance indices of a reference dynamic model, and the given ones become a reference performance indices. For getting the performance indices, an accessorial dynamic system of reference model which is driven by same excitation signals is used. A preset performance indices can be gotten by outputs and states of the reference model , the given performance indices and measured performance indices can be gotten by comparing output of adjusting system and reference model. The system parameters are adjusted by adaptive system with the difference of output of reference and adjusted system. Schematic of Fuzzy Adaptive control of airport fueling system is shown in fig. 2.

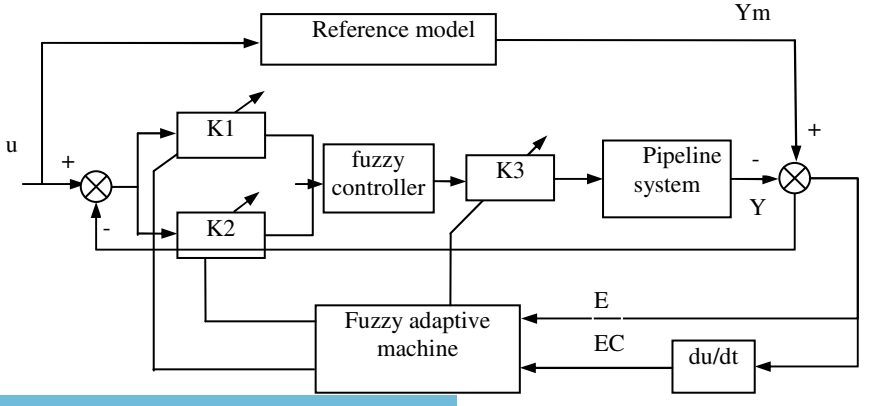


Fig. 2. Schematic of Fuzzy Adaptive control system

3 Design of Fuzzy Logic Controller

The fuzzy logic controller (fig. 2) is a two dimensions fuzzy controller, has two inputs: error of outlet pressure and a derivative of the error, and one output: rotational velocity of pump. The control system itself consists of three stages [2, 3]: fuzzification, fuzzy inference machine and defuzzification. The fuzzification stage converts real-number input values into fuzzy values while the fuzzy inference machine processes the input data and computes the controller outputs in cope with the rule base and data base. These outputs, which are fuzzy values, are converted into real-numbers by the defuzzification stage.

A possible choice of the membership functions for the three mentioned variables of the airport pipeline fueling system represented by a fuzzy set is shown in fig.3 .

The rule base used in the airport pipeline fueling system showed in Table1 with fuzzy terms derived by the designer's knowledge and experience.

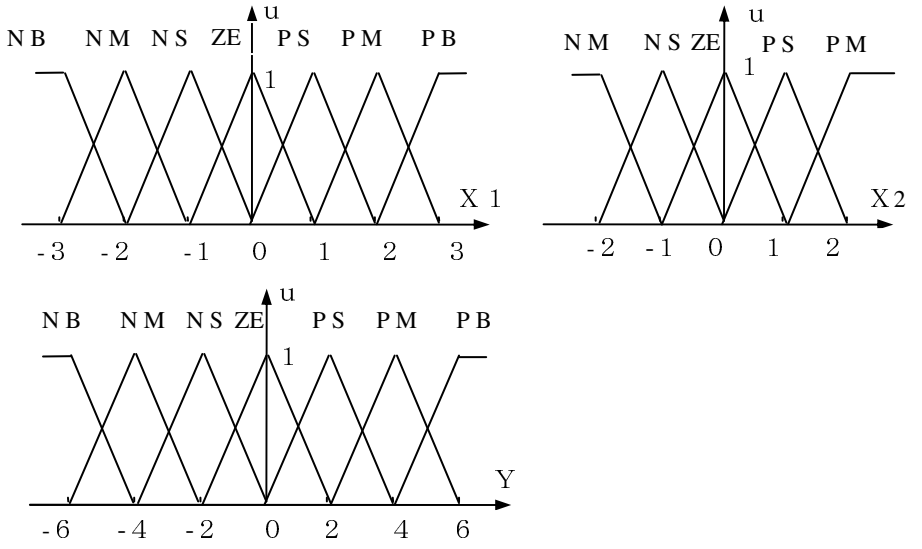


Fig. 3. Membership function for three variable of the fuzzy logic controller

The abbreviations used correspond to: NB is Negative Big, NM is Negative Medium, NS is Negative Small, ZE is Zero, PS is Positive Small, PM is Positive Medium, PB is Positive Big , x1 is error of outlet pressure x2 is derivative of the error and y is rotational velocity of pump.

Table 1. Rule base

	NB	NM	NS	ZE	PS	PM	PB
NB	PB	PB	PM	PM	PS	PS	ZE
NS	PB	PM	PS	PS	ZE	ZE	NS
ZE	PB	PM	ZE	ZE	NS	NM	NB
PS	PM	ZE	ZE	NS	NM	NB	NB
PB	ZE	ZE	NS	NM	NM	NB	NB

For example, the linguistic control rules of the fuzzy logic controller obtained from the table above used in such case are as follows:

R3:
If E(k) Is NB and EC(k) is NS then Y (k) is PB
R5:
If E(k) Is NB and EC(k) is PS then Y (k) is PS

The output of the fuzzy controller is a fuzzy set of control. As a process control usually requires a non-fuzzy value of control, a method of defuzzification called "centre of gravity method" (COG), is used here [4].

In the airplane fueling process, pressure overshoot is forbid, because the high pressure will damage the oil receiver system of airplane and pipeline system. But no overshoot and high response speed of a control system is conflicting; they are not able gotten at the same time. If a constant pressure and high response speed are need, we must use fuzzy adaptive machine to adjust the fuzzy logic parameter k1, k2, k3 . When the number of fueling airplane changes, the pressure of outlet of fueling system has a sudden change and the output of fuzzy logic controller should change greatly. So the k1 and k2 should be reduced and k3 increased. When the number of fueling airplane is not change, error and delta error is small; for improving the stability of system, decreasing the overshoot, k1 and k2 should be increased and k3 decreased.

So k1 , k2 and k3 can be calculated in off line state and a parameter adjusted table is given. Because of k1 and k2 change in the same direction, and k1 and k3 change in opposite direction, a quantization factors M is used to quantize E(error of outlet pressure) and EC(delta error). The value of M is defined by E and EC, an example of M is shown in table 2. The k1, k2 and k3 can be gotten by following equations.

$$K1 = K1 \times M \times 2 \tag{1}$$

$$K2 = K2 \times M \tag{2}$$

$$K3 = K3 / M \tag{3}$$

Table 2. Adjusting of parameter of fuzzy logic controller

	NB	NM	NS	ZE	PS	PM	PB
NB	1/4	1/3	1/2	1	1/2	1/3	1/4
NS	1/3	1	2	2	2	1	1/3
ZE	1	1	2	3	2	1	1
PS	1/3	1	2	2	2	1/2	1/3
PB	1/4	1/3	1/2	1	1/2	1/3	1/4

4 Simulation Analysis

The designed fuzzy adaptive controller is tested with simulation analysis methods by Matlab. The fueling pipeline model consists of three parallel variable speed pump and six fueling outlet. Fig.4 is control system step response result. Fig.5 is its response result with open and close of fueling outlet. In this process, it is assumed that the six

fueling outlet of the fueling system are opened and closed in sequence. By the simulation result, overshoot of the step response is zero and response speed is very quick; the pressure of fueling outlet will be stability in 5 s, the overshoot is less than the set upper limit of fueling pressure when more outlets are opened or closed.

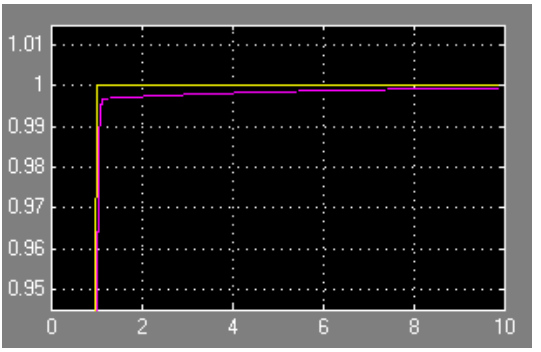


Fig. 4. Step response result

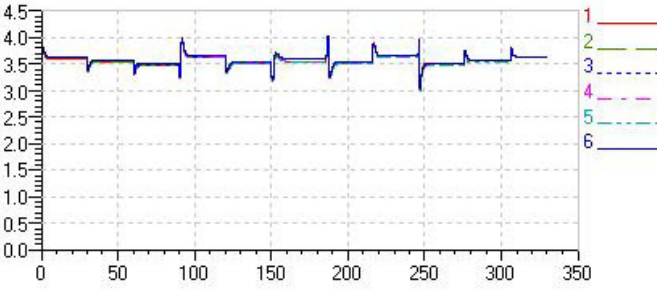


Fig. 5. The simulation result of outlet opening and closing

5 Conclusion

The idea of fuzzy adaptive control for controlling outlet pressure of airport pipeline fueling system was presented in this paper. Simulation results showed fuzzy adaptive control is very effective and can be used in airport pipeline fueling system. The new control system proposed in this paper achieves both constant pressure and less overshoot.

References

1. Li, L.-N., Liu, H.-Y., Luo, Z., et al.: Study on Improved Fuzzy-PID Composite Control Arithmetics and Its Applications. Journal of Northeastern University (Natural Science) (30), 274–278 (2009)

2. Mansour, S.E., Kember, G.C., Dubay, R., et al.: Online optimization of fuzzy PID control of a thermal process. ISA Transactions 44(2), 305–314 (2005)
3. Yang, S.: Designing and management of The oil pipeline, pp. 217–242. Petroleum Industry Press, Beijing (1995)
4. Li, C., Yu, X.: Engineering hydrodynamics, p. 191. China University of Mining and Technology Press, XuZhou (1995)

Research on the Selection between Absolute and Intensity-Based Limits of CO_2 Emissions

Chunmei Liu, Maosheng Duan, and Tao Pang

Institute of Nuclear and New Energy Technology, Tsinghua University
Beijing, China, 100084
liu6688h@gmail.com

Abstract. This paper introduced the background of limiting GHG emissions, presented a new issue: which is better between relative target and absolute target for achieving emission reduction target while taking account of national economic growth. First, establishing the model of relationship between absolute and Intensity-Based Limits under certainty, found that it is equivalent between choosing the absolute cap and intensity target if the economic growth is certain. Second, analysis according to the historical data, found that there is the similar trends of the curves of Carbon emissions and economic growth. The results of granger causality test shows that within the 98% confidence intervals, economic growth is granger cause of emissions when the duration of lag is 2, 3. That is economic growth would stimulate emissions. So we must reduce emissions through using new technology, not only limit the economic development. Third, the countries with the high absolute emissions may have relatively low intensities and vice versa. But the data is high of absolute and intensity target in China shows that economic development is relying on energy consumption and increasing emissions.

Keywords: Absolute cap, intensity based target, economic growth, GHG emissions.

1 Introduction

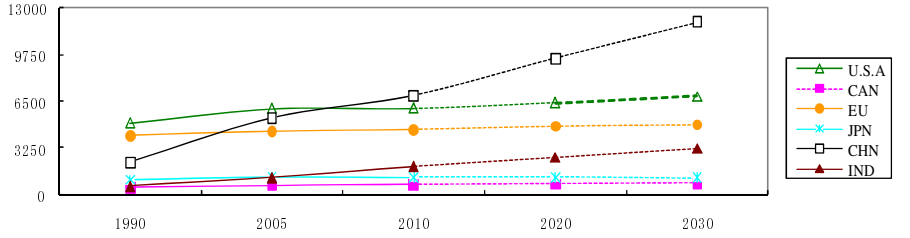
In 1988, IPCC (an Intergovernmental Panel on Climate Change) was created by the World Meteorological Organization and the United Nations Environment Programme (UNEP). This group issued a first assessment report in 1990 which reflected the views of 400 scientists. The report stated that global warming was real and urged that something be done about it. The forth assessment report considers both natural and anthropogenic drivers of climate change, including the chain from GHG emissions to atmospheric concentrations to radiative forcing to climate responses and effects. GHG emissions due to human activities have grown since pre-industrial times, with an increase of 70% between 1970and 2004. GHG emissions from energy use between 2000 and 2030 are the emissions scenarios special report projected an increase of 40 to 110% CO2e over the period (as shown in fig. 1).

In order to reduce the cost of abating greenhouse gas emissions, in December 1997, more than 150 countries negotiated the Kyoto Protocol (KP), a landmark agreement on

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 717–724.
springerlink.com

global climate change. KP recognized three flexible mechanisms: Emission Trading (ET), the Clean Development Mechanism (CDM), and Joint Implementation (JI). CDM and JI are project-based, and ET is cap and trade-based. Recently, ET becomes one of the focuses of environment. In the context of ongoing United Nations climate negotiations, most developed countries choose absolute target, and lots of developing countries choose intensity target for emission limits. In the KP treaty, 84 countries signed to commit industrialized nations to legally binding limits on their emissions of GHG. The limits are expressed in absolute emissions levels relative to a 1990 baseline. A new issue has been proposed: which is proper between a relative target and an absolute target, and how to select?

At the national level, intensity target is measured as emissions per unit of GDP. In order to achieve the emission reduction target while taking account of national economic growth and instability, some researchers take more attention to intensity-based limits on CO_2 emissions.



Data source: EIA, 2008

Fig. 1. 1990-2030 CO_2 emissions in some countries or regions (million tones of CO_2)

2 Literature Review

The proposal by the republic of Argentina in November 1999 at the 5th Conference of the Parties to the Kyoto Protocol first drew official attention to this subject (Argentina, 1999). Frankel proposed indexing GHG emission targets to GDP growth as a means of making Kyoto-type caps more acceptable to LDCs (Frankel, 1999). In February 2002, the Bush administration pledged a GHG intensity target——emissions per dollar of real GDP by 18% over the decade 2002-2012.

On the whole, there are two types of views: support and opposition. Strachan (2007) found the merit of intensity-based caps is the reduction of uncertainty in economic outcome. Jotzo and Pezzey (2007) showed that optimal intensity targets could reduce the cost of uncertainty and achieve significant increases in global abatement. Onno Kuik and Machiel Mulder (2004) assessed the pros and cons between absolute caps and relative caps at Netherlands, concluded that absolute caps lead to efficient emissions reduction, and mixed scheme would be a drawback which treated firms unequal. Gielen, Koutstaal and Vollebergh (2002) compared emission trading with absolute and relative caps using a partial equilibrium modal. Its main conclusion is that relative cap is less efficient than absolute cap because relative cap is

a combination of a price on emissions and a production subsidy. Timilsina, GR (2008) found that to achieve reduction target while maintaining expected economic growth, the global average CO_2 intensity would require a 68 percent drop from the 1990 level or a 60 percent drop from the 2004 level by 2030. Peterson, etc.(2011) assumed countries with emission targets may trade certificates, average reductions in GDP for countries with targets range between 0.1 percent and 0.7 percent in 2020 for the policy scenarios. While the GDP losses are larger for major non-Annex I countries with emission targets compared to Annex I countries. Van Vuuren, DP (2011) argued that the higher flexibility might be important in finding acceptable agreements on international climate policy. The difference of absolute caps and intensity-based caps is whether emission is association with economic growth, which one is better for controlling GHG emissions.

3 Model of Relationship between the Two Limits under Certainty

A nation can choose absolute or intensity-based limits, if it commits to constrain or limit its emissions at a target level in the future. The aggregate variables of Intensity targets include two separate sources, they are:

Output, measured by GDP, and denoted Y_t $\$/yr$, where the subscript t means time (year or period), $\$/yr$ means constant 2000 US dollars per year. However, the measure of output can vary significantly. A country uses GDP; and an industry or firm usually uses physical output (e.g., tons of steel, cement or aluminum).

GHG Emissions, measured in tons of CO_2 or CO_2 equivalent if CH_4 and N_2O etc. are factored in, denoted Q_t mt/yr, where mt/yr means million tones of CO_2 per year.

Now, we presume that the value of the subscript t range from 0 to N ($n \in [0, N]$), "I" denote emission intensity (intensity-based limits), "A" denote "absolute". Then:

$$I_t = \frac{Q_t}{Y_t} \tag{1}$$

Suppose that at the initial time when $t = 0$ one nation choose absolute limit, and decide the limit is Q'_T at the given time when t=T.

Expectations Devotes E. Based on the economy growth and emissions, Q_T^{BAU} is the emissions at t=T of BAU (business-as-usual) scenario. Then the emissions reduction is that:

$$r_T^A = Q_T^{BAU} - Q'_T \geq 0 \tag{2}$$

Suppose that the expectation of GDP at t=T is $E[G_T]$. Then the intensity-based target is that:

$$I_T' = \frac{Q_T'}{E[Y_T]} \tag{3}$$

When the expectation of GDP is equal to the true GDP or GDP is certain. The emissions reduction of the intensity based target is below:

$$r_T^I = Q_T^{BAU} - I_T' \times E[Y_T] \tag{4}$$

According to eq. (3), the emissions reduction of the intensity based target is equal to the emissions reduction of the absolute based target, that is

$$r_T^I = Q_T^{BAU} - I_T' \times E[G_T] = Q_T^{BAU} - Q_T' = r_T^A \tag{5}$$

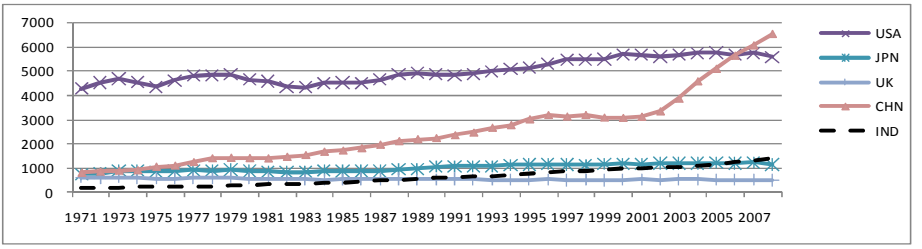
Eq. (5) indicates that the reduction between intensity based and absolute based target is the same in one nation, if the costs of the emission reduction are equal. On this occasion, for policymakers, there is little difference between the two target types of emission reduction limits.

4 Analysis of the Two Types Choosing in Typical Countries

4.1 Trends of Historical Data from CO2 Emissions and Economic Growth

Recently, the approach received a boost when Canada and China expressed their intention to implement intensity targets in the coming years. The KP set binding targets for 37 industrialized countries and European community for reducing GHG emissions is its major feature, and the average amount is 5 percent in absolute emissions levels relative to 1990 levels over the period 2008-2012.

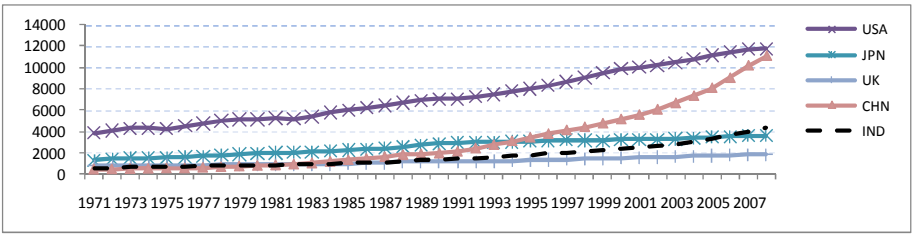
Launched in 2005, the European Union (EU) emission trading system works on the cap and trade principle which means there is a limit on the total amount or the absolute target. Under the KP, EU, USA and Japan all chose absolute target, and committed the reduction are 8, 7 and 6 percent of 2012 against 1990 levels. While some countries including India and China chose intensity-based target.



Resource: IEA, CO_2 emissions from fuel combustion, 2010

Fig. 2. CO_2 emissions (sectoral approach) (million tones of CO_2)

At the national level, as noted earlier, Intensity target is typically measured as emissions per unit of GDP. Fig.1 shows the carbon emissions from combustion in typical countries. Emissions in developed countries are stable and the changes are very slow from 1970s to 2008. While in developing countries, the emissions grow rapidly in the same period. The fastest-growing of emissions is in China, and the whole emissions surpassed the U.S. as the biggest emitter of GHG. Fig. 2 shows GDP using PPP in the typical countries. As is shown in the picture, we can see clearly the trends of economic growth in USA, JPN and UK is flat, while it grows at a very fast speed in CHN and IND in the period of 1971-2008. The curves between Carbon emissions and economic growth have the similar trend. Through the analysis of the model, we found that there is a positive correlation between Carbon emissions and economic growth.



Resource: IEA, 2010

Fig. 3. The trends of GDP using purchasing power parities (billion, 2000 US dollars)

4.2 Granger Causality Test

Through the intuitive historical trend figure above, the economic growth in developing countries is uncertain. So we need to analysis the relationship between economic growth (Y) and emissions (Q). In order to analyze the causality of emissions and economic growth, we will use granger causality test which was put forward by Granger (1969) [10] and Sims (1972). The basic hypothesis of granger causality test is that if x is the Granger-cause of y, the lagged variable of x will explain the lagged variable of y significantly. So built model as followed:

Unconditional limit model:

$$Y_t = \alpha + \sum_{i=1}^{m} \alpha_i \Delta Y_{t-i} + \sum_{j=1}^{k} \phi_j \Delta X_{t-j} + \varepsilon_t \tag{6}$$

Conditional limit model:

$$Y_t = \alpha + \sum_{i=1}^{m} \alpha_i \Delta Y_{t-i} + \varepsilon_t \tag{7}$$

ε_t is white noise time-series. α , ϕ is regression coefficient. t is the size of sample. m , k is lagged size of Y_t , X_t . The residual sum of squares of Eq. (6) is ESS_1 and The residual sum of squares of Eq. (7) is ESS_0 .

Null hypothesis is $H_0 : \phi_j = 0$, alternative hypothesis is $H_1 : \phi_j \neq 0$ (j=1,2,……,k)。 Built F-statistic as followed:

$$F = \frac{(ESS_0 - ESS_1)/m}{ESS_1/(n-k-m-1)} \rightarrow F(m, n-k-m-1)$$

If F-statistic>the critical value of F-distribution, null hypothesis will be rejected. The result (as shown in table 1) is output of granger causality test.

Table 1. The results of Granger causality test between Q and Y

Pairwise Granger Causality Tests				
Null Hypothesis	Lags	Obs.	F-Statistic	Prob.
	1	37		
Q does not Granger Cause Y			3.10873	0.0869
Y does not Granger Cause Q			5.31944	0.0273
	2	36		
Q does not Granger Cause Y			3.79397	0.0336
Y does not Granger Cause Q			6.23599	0.0053
	3	35		
Q does not Granger Cause Y			3.43176	0.0304
Y does not Granger Cause Q			4.44591	0.0112
	4	34		
Q does not Granger Cause Y			2.73022	0.0517
Y does not Granger Cause Q			2.81405	0.0468

The test results show that within the 98% confidence intervals, Y is Granger cause of Q when the duration of lag is 2, 3. Within the 95% confidence intervals, Y is Granger cause of Q when the duration of lag is 1~4, and Q is Granger cause of Y when the duration of lag is 2, 3, and Q is not Granger cause of Y when the duration of lag is 1, 4. So Y would stimulate Q. That is, in developing countries, there is a significant effect on Q from Y.

Since economic growth stimulates the emissions, then, developing countries in which economic growth is uncertain, when economic growth downturn, choosing intensity target is more advantageous to emissions reduction actually. And when economic growth slope, the generation of hot air can be prevented. That can promote energy-saving technical progress in reducing emissions. When economic growth rose, developing countries can reduce emissions coupled with economic growth. it is relatively easy to accept in developing countries.

4.3 Difference in Transforming between Absolute and Intensity-Based Target

Absolute targets emphasize reducing emissions growth, while intensity targets emphasize reducing emissions relative economic growth. Until there is emissions

reduction technology, economic development is integrally tied to emissions growth for foreseeable future. At country level, GHG intensity is measured as emissions per unit of GDP. This section focuses on difference in transforming of the top emitting countries in the absolute and intensity target. Converting the absolute to intensity target, and seek the difference of world rank. Table 2 shows GHG total emissions, intensity and the rank of the top emitting countries. It is quite distinct after comparing total emissions and intensity.

Table 2. Comparison between Absolute and intensity target of a hypothetical business-as-usual in top emitting countries (six GHG, 2000)

Country	Absolute emissions		GHG intensity		
	Emissions (Mt CO_2 eq.)	Rank	Emissions/GDP (tons CO_2 eq./$million. GDP)	Changes of Rank	
U.S.	6928	1	722	↓	4
CHN	4938	2	1023	↑	1
IND	1884	3	768	-	3
JPN	1317	4	396	↓	7
GER	1009	5	482	-	5
CAN	680	6	809	↑	2
U.K	654	7	416	↑	6

Note: GDP-PPP is a gross domestic product measured in purchasing power parity (in constant 2000 international dollars).

Source: Timothy H., Kevin, A. B. Analysis of greenhuose gas intensity targets[M]. World Resources Institute, 2006.

From the table, we can see that the countries with the high absolute emissions may have relatively low intensities and vice versa. For example, Japan ranks fourth in terms of absolute emissions, but seventh in terms of emissions intensity. On the other hand, China with absolute emissions of 4938 Mt CO_2 ranks second in terms of absolute emissions but first in intensity (1023 tons/million $).

5 Concluding Comments

First, it is equivalent between choosing the absolute cap and intensity target if the economic growth is certain. Through the analysis shows that, for developed countries, they can choose anyone of the two kind of target, because the economic development is certain, and the technology is advanced. It is not really for developing countries. Because of the uncertainty of economic, the intensity target can reduce the incidence of "hot air". With intensity target, when they suffers from an economic downturn, and there is accompany drop in emissions. When economic growth rose, developing countries can reduce emissions coupled with economic growth. It is relatively easy to accept in developing countries.

Second, Emissions in developed countries are stable and the changes are very slow, while in developing countries, the emissions grow rapidly in the period of 1971 to 2008. The curves of Carbon emissions and economic growth have the similar trends. The results of granger causality test shows that within the 98% confidence intervals, economic growth is Granger cause of emissions when the duration of lag is 2, 3. that is economic growth would stimulate emissions. So we must reduce emissions through using new technology, not limit the economic development.

Third, the countries with the high absolute emissions may have relatively low intensities and vice versa. But the data is high of absolute and intensity target in China shows that economic development is relying on energy consumption, and increasing emissions.

Acknowledgments. This work is supported by the Key Project of Chinese National Programs for Fundamental Research and Development (973 program) (No. 2010CB955504).

References

1. Argentina. Revision of the first National Communication to the UNFCCC. ARG/COM/2B (1999)
2. Frankel, J.: Greenhouse gas emissions. Policy brief 52. Brookings Institution, Washington, DC (1999)
3. Strachan, N.: Setting greenhouse gas emission targets under baseline uncertainty: the Bush climate change initiative. Mitigation and Adaptation Strategies for Global Change 12(4), 455–470 (2007)
4. Jotzo, F., Pezzey, J.: Optimal intensity targets for greenhouse gas emissions trading under uncertainty. Environmental and Resource Economics 38(2), 259–284 (2007)
5. Kuik, O., Mulder, M.: Emissions trading and competitiveness: pros and cons of relative and absolute schemes. Energy Policy 32(6), 737–745 (2004)
6. Granger, C.W.J.: Investing Casual Relations by Econometric Models and Cross-spectral Method. Econometric 37, 424–438 (1969)
7. Gielen, A., Koutstaal, P.R.: Comparing emission trading with absolute and relative targets. Presented at the 2nd CATEP Workshop on the Design and Integration of National Tradable Permit Schemes for Environmental Protection, London, vol. 3, pp. 25–26 (2002)
8. Timothy, H., Kevin, A.B.: Analysis of greenhuose gas intensity targets. World Resources Institute (2006)
9. Peterson, E.B., Schleich, J., Duscha, V.: Environmental and economic effects of the Copenhagen pledges and more ambitious emission reduction targets. Energy Policy 39(6), 3697–3708 (2011)
10. Van Vuuren, D.P., Riahi, K.: The relationship between short-term emissions and long-term concentration targets. Climatic Change 104(3), 793–801 (2011)
11. Robert, M., Ottmar, E.: Revisiting the case for intensity targets: better incentives and less uncertainty for developing countries. Energy Policy 38(9), 5048–5058 (2010)
12. Annex B Kyoto Protocol to the United Nations Framework Convention on Climate Change. United nations (1998)
13. Pizer, W.: The case for intensity targets. Resources for the future. Climate policy 5(4), 455–462 (2005)

Ontology-Driven Resource Aggregation Model on Geological Survey Information Grid*

Song Miaomiao[1], Xie Zhong[1], Li Chaoling[2], Wu Liang[1], and Lv Xia[2]

[1] Faculty of Information Engineering, China University of Geosciences, Wuhan, China
[2] Development Research Center, China Geological Survey, Beijing, China
songmiaomiao_2006@126.com

Abstract. The platform of China Geological Survey Information Grid (CGSI Grid) which is constructed in the distributed heterogeneous network environment is confronted with troubles of discovering, exchanging and describing spatial Data, because it is distributed in different grid nodes and of different formats and types. In order to solve these problems, Ontology-driven Resource Aggregation Model utilizes geological domain ontology description language (OWL) to describe geospatial information and GIS services and extend the semantic information of data, establishes peer-to-pear resource aggregation model, and builds multi-node resource discovery mechanism in distributed grid environments across the whole country, in order to promote the sharing of geospatial data and GIS capabilities, make sure that geological resources are published, discovered, retrieved, invoked automatically and intelligently, and actualize collaborative computing of heterogeneous data.

Keywords: CGSI Grid, Resource Aggregation, Geological Domain Ontology Modeling, Geographic Web Services.

1 The Architecture of CGIS

CGSI Grid (www.gsigrid.cgs.gov.cn) is an open, integrative, reconfigurable, reusable, scalable Geological Survey one-stop information sharing and service platform. It regards digital results of geological survey as a core, stores and manages multi-source, heterogeneous, massive geological data centrally, realizes integration and sharing of information resource of geological survey under the environment of Grid Computing, finally provides geological information services of different particle size grades for government, enterprises, companies and individuals [1].

The CGSI GRID platform has four modules, including Data Layer, Construction Layer, Resource Layer and Discovery and Integration Layer. In Data Layer, there is massive spatial data of different coordinate reference system, and models. The Construction Layer provides physical carrier and link transmission channel for spatial resource, constructs a distributed heterogeneous network environment and then forms large distributed grid service node clusters. According to Open Geospatial

* This work is sponsored by the National High Technology Support Program of China (Grant No. 2011BAH06B04.)

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 725–728.
springerlink.com

Consortium (OGC) service specifications, and Web Service Resource Framework (WSRF), Resource Layer builds grid services of different particle sizes and special services for different applications, shields the heterogeneity of hardware and software, and finally provides functional entities which are cross-platform, loosely coupled, highly reusable grid services. Discovery and Integration Layer contains resource aggregation middleware and peer-to-peer resource manager [2], uses the ontology-driven resources aggregation standard to describe, organize, register and manage resources, and provides efficient solutions for resource matching, dynamically binding between functions and data, and automatically combining services.

2 Ontology-Driven Resource Aggregation Model

Based on the platform of CGSI Grid, Resource Aggregation Model is designed to achieve a unified description, organization, discovery and integration of distributed resources. Following the one-stop aggregation standard of geological survey services, based on geological metadata of resources and geological domain ontology thesaurus, it establishes a resource aggregator, and the mechanism of discovery and integration. Therefore, it includes two important parts: resource aggregator and resource integrator.

Resource aggregator is used to build a relatively stable resource view and provide transparent, integrated access for users in complex internet environment. According to characteristics of resource, resource aggregator divides resources into Data Warehouse and Function Warehouse. Data warehouse contains data of Spatial Reference System, maps, Feature/Raster and metadata, and provides interfaces for registering, modifying, deleting and searching. Function warehouse is a collection of services which are grouped together according to certain principles, and is divided into atomic services, basic element services, professional services and service chains. Additionally, it is driven by the function engine to build services and form service chains by workflow modeling. In this way, resource aggregator organizes and describes all data and function resource uniformly, controls and manages them. Following the geological information metadata standards and OGC services metadata standard, Resource Integrator establishes geological domain ontology vocabulary and metadata databases, and uses Resource Description Framework (RDF) and Ontology Language of Web (OWL) to describe and express them. In Resource Integrator, the kernel is geological ontology model, which will be discussed in details in the next part.

The model works on the principle like that: first, it receives user requests, and analyses the type of resources required by task processing and other constraints (like the range of data, time, scale, and so on); Secondly, according to the rule constraints of resources, it sends a subscription request to the CGSI Grid aggregation middleware, and receives a metadata set of resources which meet the requirement; Then, complying with the filtering strategies, it chooses a certain amount of resources, establishes resource indexing, and returns the resource identifier.

3 Geological Domain Ontology Modeling

All geological resources in the platform of CGSI Grid are typed by geological domain ontology. Ontology is an explicit specification of a conceptualization [3]. The

geological domain ontology describes the relationship between concepts in the geological field.

Geological domain ontology divides geological entities into several corresponding successive levels, according to attributes (or features) of geological entities, by the method of linear classification [4]. Each level is divided into several categories, and then it constructs parallel relations between categories of the same level and subjective relations between categories of different levels. Each category contains several data items, like keywords, source, the equal word, the substitutive word, the relative word, alternative words, the type of attribute values, the range of fields, and so on. Ultimately, the classification and coding system of the geological field is established for each category and provides standards for resource classification and coding. The modeling procedure is as follows:

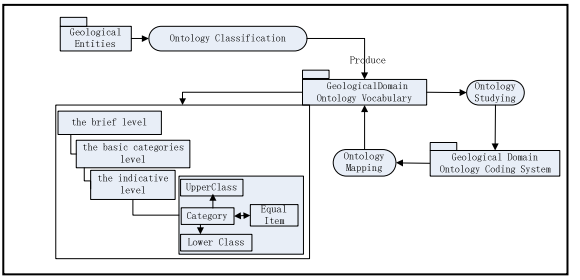


Fig. 1. The procedure of geological ontology modeling

In the geological domain ontology, we stores geological information in three levels: the brief level, the basic categories level and the indicative level [5]. The brief level is the first level and contains information of twenty-two brief categories. The basic categories level contains one hundred and four basic categories which are subclasses of brief categories. In the indicative level, there is category information in the third or upper levels. In the geological domain ontology, each category has upper and lower items in the brief level, the basic categories level and the indicative level. These all reflect the thought of "vertical pedigree and horizontal grid" [6].

In the geological domain model, there are more than 50,000 categories, each category stands for a type of geo-resource and every geo-resource has an identifier. The identifier of each geo-resource is well defined and contains ontology codes which stand for the brief category, the basic category and the indicative category. The brief category which is the entrance of professional geological classification is coded by two characters(e.g. KC) corresponding to the standard of GB / T 9649, the basic category which is the lower class of brief category is coded by one character from 1 to 9 and “A” to “Z” (e.g. 5), and the indicative category which is about attribute of the third or upper level is coded by two characters, like 01, 02, 89, A0-A9,B0-B9,…Z0-Z9,AA-AZ,BA-BZ,ZA-ZZ(e.g. 10070102). The ontology code of a category named “metallogenic prediction map” is “KC510070102” which is a whole ontology code.

Based on RDF (Resource Description Framework), we use web identifiers (like Unified Resource Identifier (URI), classes, individuals, instances, attributes and attribute values) to describe distribute geological resources. In CGSI Grid, OWL is

used to collect these distributive resources and make sure that they can be accessed directly via the network and invoked directly by URI. In the practice, we also use OWL and RDF to establish the mapping between ontology concepts and distributed resources, express semantic information of resources explicitly.

By the resource discovery mechanisms based on ontology description and ontology mapping, we first obtain distributed geo-information resource list information according to ontology classification, pick out the data we are interested in, and then view its RDF description. The description of resources by RDF contains a unique Unified Resource Locator (URL), an ontology code, data sources, and service address; through the service address, we can get the resource entities which contain spatial information and attribute information.

4 Conclusion

The platform of CGSI Grid is established in the heterogeneous distributed environment, using grid technology of geographic information, grid middleware, resource aggregative technology, and other spatial information processing technology. Geological resource aggregation model is built to eliminate "Spatial Information Island" gradually, according to the standard of Geological Domain Ontology Modeling. In CGSI Grid, Ontology-driven Resource Aggregation Model improves the sharing of data resources and serving degree of functional resources to ensure that multi-source heterogeneous distributed resources is organized, discovered and integrated automatically, intelligently. Based on it, CGSI Grid achieves storage and sharing of distributed mass geological resource and builds the environment of geospatial information service and resource connection and spatial collaborative computing. For more details, please visit the portal website (www.gsigrid.cgs.gov.cn).

References

1. Song, M., Xi, Z., Luo, X.: Research of Catalogue Service System based on China Geological Survey Information Grid Platform. In: The 18th International Conference on Geoinformatics, June 18-20 (2010)
2. Li, C.: The Technical Infrastructure of Geological Survey Information Grid. In: The 18th International Conference on Geoinformatics, June 18-20 (2010)
3. Gruber, T.: Toward principle for the design of ontologies used for knowledge sharing. Human Computer Studies 43(5/6), 907–928 (1994)
4. Wu, L., Xu, L., Chen, X.: Human-Earth-GIS Relations Discussion Based on Subject-Human and Geo-ontology. Geography and Geo-Information Science (01) (2006)
5. China Geological Survey Bureau, Geological domain ontology classification and code (unpublished)
6. China Geological Survey Bureau, Geological Information Metadata Standard, p. 2, No. DD2006-05 (December 2006)

The Design of Ecg Signal Generator Based on ARM9

Ai-ju Chen[1] and Yuan-juan Huang[2]

[1] School of Electronic and ElectricalEngineering, WUHAN TEXTILE UNIVERSITY, Wuhan, China 430073
[2] The Obstetrics and Gynecology Department of, QianJiangShi Hospital in Hubei province, Hubei, QianJiang 433100

Abstract. This topic design of ecg signal simulation generator, through digital signal reduction by ecg original analog signals to simulate human body, or send other analog signal to ecg products, achieve the purpose of test ecg products. Ecg simulation signal generator forms by the hardware and software of upper and lower level computer, the task of PC is sending ecg data down to machine work, the task of lower machine is responsible for receiving and storage ecg data. PC is based on Visual c + + integrated environment of software system. through Visual c + + complete program design. Lower place machine is based on embedded system by the ARM9, ARM9 core board, serial ports, Ethernet, DAC, LCD, input/output, and other plug-ins interface component.The whole system realizes several different format of ecg signal data of transmission through serial and net, storage and analog output and polarization voltage function, and can also produce sine wave and square-wave and normal ecg wave can be used waveform, such as amplitude, frequency of regulation.

Keywords: Ecg simulation signal generator, PC, Visual c + + programming, Ecg product testing, Embedded system.

1 Introduction

With the development of computer technology, technology innovation of medical equipment industry and constantly progress of ecg signal acquisition and processing technology. the requirements of ecg product's performance become more higher and higher, the quality of ecg product in relation to the user's healthy , so it is very important to make performance test for ecg products in the development process and before put into the market.

Because of the clinical testing for products in the development process, need to use actual human ecg data, at present the usual method is through taking the prototype to the hospital for data collection, need cooperate by clinicians and subjects, making test work not convenient .if we can develop an ecg simulation signal generator system, don't have to connect the actual human and also can test the actual clinical effect of ecg product , will make great convenience to product development.At present the function of ecg simulation signal generator is limit in current market, they can only produce some simple waveform, and because of without interface , it cannot directly read the original ecg signal from actual human body, thus cannot reduction out the actual human signal by collection,unable to meet the higher requirement of the ecg products test.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 729–733.
springerlink.com

Based on this subject, development of a ecg signal simulation, not only have the function of normal ecg signal generator, and can reduction out actual human signal in ecg data. it can also compared different manufacturers of ecg products' performance through standard of database of ecg data.

2 System Introduction

Ecg simulation signal generator including micro controller, Ethernet communication interface, serial communication interface, USB communication interface, power supply circuit, DA conversion circuit and signal processing circuit. This design using samsung's ARM chips as the whole system master S3C2410 chips.

The whole system hardware diagram as shown in figure 1 shows:

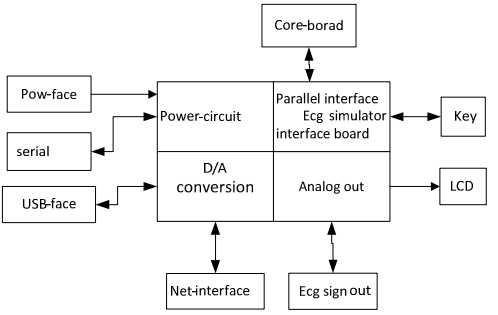


Fig. 1. Hardware diagram of the whole system

This system integrated circuit design, microcomputer control technology, the communication technology and the corresponding software technology etc., all research contents include: 1, S3C2410 lower level computer control program development; 2, PC user interface development, including read ecg data on PC, and display of serial ecg data files WangKou transmission; 3 the development the firmware of ecg simulation signal generator system.

Core board includes master ARM chips S3C2410 NAND FLASH, DRAM, and Ethernet devices such as CPLD chip, and formed 200MHz frequency of smallest system

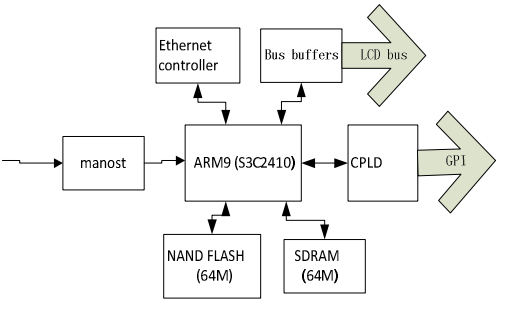


Fig. 2. Core board the principle diagram

Core board's main function is to control the whole system, receive, storage, management the ecg digital signals from the PC, response keyboard, output the digital ecg signal through SPI or data bus to DAC, displayed the data ID heart waves and output of state on LCD, output amplitude frequency sine wave designated square wave and triangle and etc. Figure 2 is the core board principle diagram.

3 System Software Design

This design software of mainly divided into the PC and lower place machine , PC based on Visual C + + to implementing software system integration environment, using Visual C + + language programming complete design. Lower level computer software design applications ARM development environment ADSv1.2, simulators ARM v2.2 Mutil - ICE, based on embedded operating system mu C/OS - Ⅱ, realize the data storage and receive data from PC and in response to the keyboard's button action .and display the operation interface on narrow LCD screen . Figure 3 for its total program flow chart.

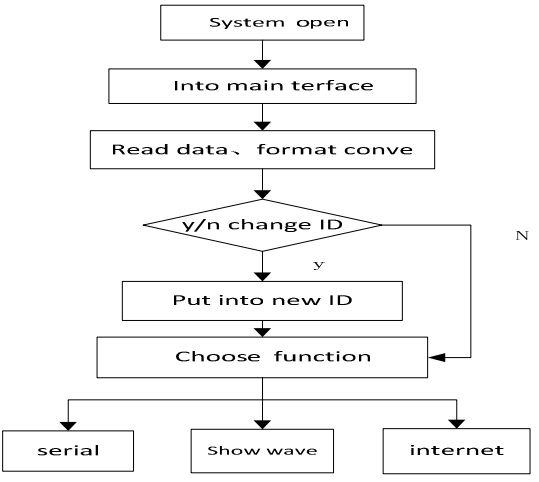


Fig. 3. Total program flow chart

4 Function Test

4.1 Waveform Display

Four formats of ecg data are normal display, for example the EDAN format data show,. EDAN format: eight guide league 10s of data, sampling rate for 1000Hz. EDAN format without filtering, ecg waveform ecg data for the original figure 4 the twelve guide league waveform figure.

Fig. 4. EDAN waveform

4.2 Filtering Effect Testing

MIT data format: the original sample rate for 250Hz, after conversion is 1000Hz.

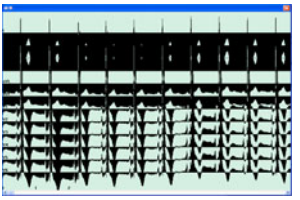


Fig. 5. MIT data filtering former waveform

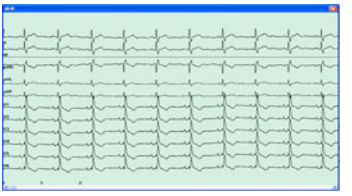


Fig. 6. MIT after data filtering waveform

Figure5, figure 6 MIT data filtering the contrast between the before and after waveform can be seen filtering, aliasing composition removed, to display clear ecg waveform.

The CSE data formats: sample rate for 1000Hz, after converting is 500Hz. The filtering effect is same as MIT data filtering , here no longer, for example.

4.3 Serial Communication Test

In order to watch data whether correct which sent from PC to down place machine, using a serial port debug assistant to examine send data whether is right , using a serial port debug assistant to simulate a machine up send request to under a machine . Batch download need continuously to send several files, and serial debugging assistant without function of continuous send frame , when batch download test by lower level computer.Can see effect of the PC send.as shown in figure 7 below.

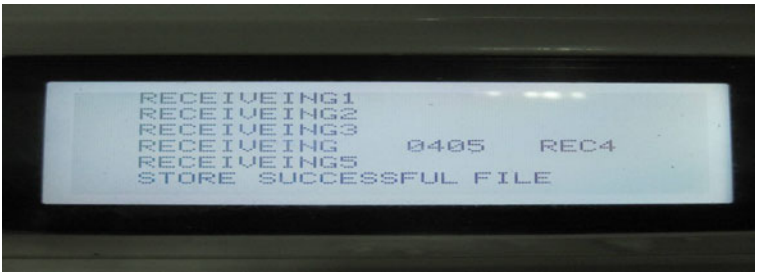


Fig. 7. Serial send finished

RECEIVEING1 means received the first frame, after received 3th frame, then received subcontractor data of the fourth frame.Last show RECEIVEING4until after received bag 405, Display STORE SUCCESSFUL FILE after collect all .if Lower place machine show receiving success, also suggests that PC data sending finished, communication is successful.

4.4 Ethernet Communications Test

Due to effect of realize each frame same the with serial communication, here no longer be detailed introduction. Figure 8 major display data frame continuity. Can see only four frame content.ethernet communications.

Fig. 8. Ethernet send finished

5 Closing

This topic research of ecg simulation generator basically achieve the expected effect, PC can clear display three different formats length of data of original ecg waveform. and can batch and continue sent to down machine by serial ports and ethernet, the under machine can reveive correctly and stored it. in output it will reduct to analog sign, the output waveform without distortion and analog signal amplitude accurate, and can choice poling voltage by any standard key, buttons can control the size of positive polarization voltage .This ecg signal generator not only with the function of normal ecg signal generator, but alos can reduction actual human signal which put in ecg data , can comparison different manufacturer of ecg product performance according to ecg data in the standard of ecg database , predictable it with prospect broad market.

References

1. He, L.S.: Design of 12 lead ECG signal generator based on DDS arithmetic. Chinese Journal of Scientific Instrument CNKI:SUN:YQXB.0 (2010-02-009)
2. Yu, X.F.: Principle and design of medical electronic equipment. South China University of Technology Press, Guang zhou (2003-05)
3. Jian, X.Z., Xu, L.: The embedded high-speed synchronous acquisition device based on ARM9 and FPGA. Microcomputer Information CNKI:SUN:WJSJ.0 (2011-02-012)
4. Hua, C.Y., Tong, S.B.: Analog electronics technology. Higher Education Press (2006-05)

Research on Medical Wireless Frequency Hopping Communication by nRF24L01

Ai-ju Chen[1] and Yuan-juan Huang[2]

[1] School of Electronic and Electrical Engineering, WUHAN TEXTILE UNIVERSITY, Wuhan, China 430073
[2] The Obstetrics and Gynecology Department of QianJiangShi Hospital in Hubei Province, Hubei, QianJiang 433100

Abstract. Medical electronics become more and more portable in modern time, the transmission of the data towards more wireless, it also need higher requirements on power and the instruments quality. Although the Bluetooth usually can be used in wireless transmission in the system of the data transmission, but it is too expensive, and power is too larger.

The platform is ARM in this system, and research on the wireless data transmission system of the short-range based on nRF24L01. Using the wireless chip nRF24L01 of the NORDIC company that can low the power consumption to 1mW and obtain better standby time performance. Using the technology of the DS/FH-SS(the combination of the direct sequence spread spectrum and the frequency-hopping spread spectrum) that can solve well the interference problems of the signal from the other device in the communication process of the 2.4G band ,to achieve the lower BER of the wireless data transmission and high reliability characteristic , on the basis of using the adaptive frequency-hopping technology.

The system is mainly applicable to complex environment, serious electromagnetic interference, the higher requirements of the data reliability, and the medical electronics field of the long-term stability work.

Keywords: Wireless data transmission; frequency-hopping spread spectrum, direct sequence spread spectrum, adaptive frequency-hopping.

1 Introduction

Wireless data transmission is refers to remote transfer various physical which output from industrial site equipment by wireless transmission module. Wireless data transmission equipment widely used in wireless data transmission fields, typical applications include remote control, remote sensing, and in the telemetry system of data acquisition, testing, alarm, process control and so on.

Taking into account the particularity of the medical electronic industry. The instrument need to collected biological signals from patient's body, then the collected signal will be transmit to host machine by the long lead wire and shielding wire for corresponding processing.however a large amount of data wires not only brought trouble to doctor,but also give inconvenience and pain to patients. Using wireless data

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 735–740.
springerlink.com

transmission, just need equipment connect with wireless module.It will save a lot of data wires and become move much easier. Relative cable transmission, wireless transmission with great convenient for patients and doctors. it is helpful to improve the patient's life quality. The wireless technology also can be applied to the remote medical care, Patients need a wireless acquisition module at home, transmit data to the computer through wireless then send to doctor might, in the process of collecting, patients can moving in a certain range, it is very convenient and save a lot of time. People is urgent demand medical instrument can be more friendly monitoring and treatment, wireless transmission with many benefits,makes people's demand become more and more urgent.

At present, there are much wireless data transmission system which based on single chip microcomputer control, However, most of the wireless transmission scheme is disquisition educes some communications, working in a fixed channel, Can't automatic changing channels to circumvent the interference, poor anti-jamming performance, not good as FH mode. Once meet jamming signal on the fixed channel, the rate of data packet loss will increase greatly, reliability is low, can not be used for medical electronics circumstance where requires very high reliability; There are some FH communication solutions, first in a fixed frequency communications, jump to next frequency point for communication until there is interference, although there are certain effect, but can't meet the FCC rules. For example: when frequency-hopped communication stay in each of communication channels no more than 400Ms, And a few wireless module as Bluetooth has become mature now, its bit-error rates above 10e-5, using adaptive FH communication, but it has two weakness: first, high power consumption; second, cost too much.

In view of the current situation, this paper examine a wireless frequency hopping communication system which used for wireless data transmission, whith the very high anti-electromagnetism interference , an extremely low bit-error rates. This system can be according to the information of real-time data acquisition channel, automatic avoid interference frequencies, with excellent anti-interference ability, it is fit to bad environments, need wireless communication, especially in the place demand higher anti-jamming of the wireless communication , can long stable and reliable work.

2 System Design

The wireless data transmission system is a specific application of the short-range wireless communication technology in the modern medical electronics. When collecting data, sensor will immediately transmit data to the next machine. When transmitting data in wireless connection and using ARM as platform, based on nRF24L01 we research the reliability of data transmission. The hardware is divided into two modules: transmitter module and receiver module, as shown in Figure 1.

In the wireless transmission system, there is power supply module, CPU, wireless transceiver modules. When choosing CPU, we use cost-effective, low power consumption of ST's Cortex-M3 core-based STM32F103VDT6. We use small, low power consumption, fewer external components, low-cost RF chip nRF24L01 of NORDIC companies as wireless module.

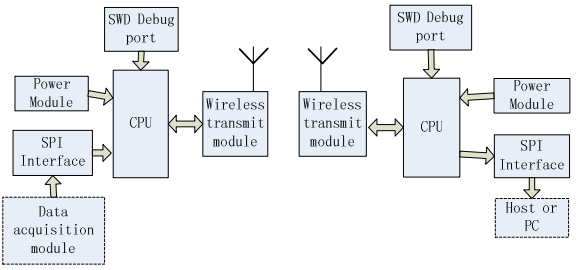


Fig. 1. The block diagram of system hardware

Considering the chip performance as well as the needs, we use 54 channels, 1,700 jumps / sec; in order to improve anti-jamming we use the adaptive frequency hopping scheme; in order to enhance immunity we combine the frequency hopping spread spectrum and direct sequence spread spectrum (FH / DSSS).

3 Software Design

The main part of the system is software design. Particularly the hopping part and the adaptive part are the core of this design. In Figure 2 it shows the block diagram of system software which includes the boot shake hands, time synchronization and calibration, spread spectrum and adaptive frequency hopping.

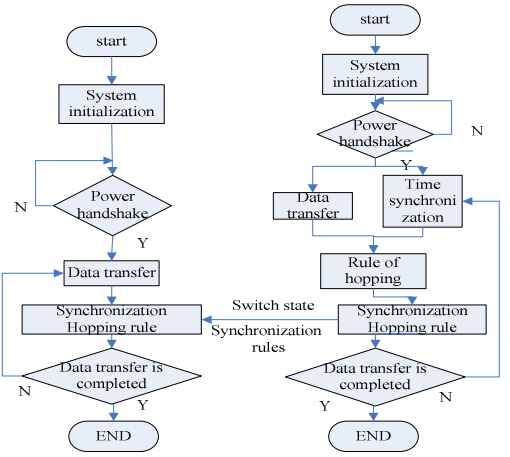


Fig. 2. Main flow of system

3.1 The Boot Shake Hands

At boot time sender and receiver contact each other to create a channel and synchronize the time to do the same frequency-hopping communication.

In this system, we use a four-way handshake, but the aim of the first two shake hands which equals the two shook hands in TCP / IP is to establish an initial path, followed by two handshake to transmit synchronization information. And then the two sides start the timer for synchronize.

When synchronizing time, we temporarily use the boot between the two sides on a single frequency to communicate. In the last three steps in the process of time synchronization, after the receiver Tx receives time synchronization signal of Rx, first it returns a ACK signal to Rx, and then starts Timer2. Because after three-step handshake the channel has been established. And the information which is received shows the channel is smooth. So it can be sent directly back the ACK; after Rx receives ACK signal from Tx, the signal which Tx received is synchronization signal. Rx immediately start Timer2 now. The handshake is completed.

3.2 Time Synchronization and Calibration

In the hopping process, due to various factors sender and receiver will be staggered, leading to communication interrupted. In order to ensure the normal communication, real-time calibration of frequency-hopping time for both parties is needed.

Transmitter TX and receiver RX are two relatively independent systems, even if they use the same CPU, but the crystal is different. In the current system, 8M crystal is used. And the frequency error is about 50ppm. After the oscilloscope measured, in 2～3S relative time offset of the two systems is about 15us, meaning after RX and TX experience 3400～5100 jumps, the error will reach about 15uS. In more than ten microseconds it will result to receive emission out of step.

In the system we use data synchronization method, meaning that the transmitter does not specifically add sync symbol and simply sends data. After receiver received data, the CPU enters the interrupt. It will determine whether it receives data symbol or not. Then it reads the data. So the time delay is greatly reduced. In order to achieve higher accuracy and reliability, considering the adaptive frequency hopping, synchronization at regular intervals is set to 3.0076 seconds.

3.3 Spread Spectrum

Medical electronics demands high communication quality, especially anti interference ability. The main purpose of spread spectrum is to improve the anti interference ability of the system, to decline bit-error rate and to increase the communication quality.

The frequency switch time of nRF24L01 is only 130 us, so that it can be used to frequency hopping. Besides, considering the automatic answering and retransmission of nRF24L01 will take large amounts of slots, after data are written into nRF24L01, a Phase-Locked Loop of 128us to stable the time will be needed. When transmission starts, a frame of 31 Bytes will be used, that means, it'll require 33+4*31=157us. After

data is sent, the receiving state is immediately set to wait for the reply of RX. 128us later, after data is received by RX, it is verified using CRC. If data is correct, a reply will be sent. After TX receives reply, interrupt pin IRQ will be pull down and it will inform CPU that the data is sent successfully. So sending a data frame containing 31bytes will spend 456us. After receiver received data and confirm that it is correct, it will spend 31Byte to input data into CPU. In fact, the time cannot be so precise. So 40 us will be left as margin. Overall, it needs 700us, about 1428 jump/s. So the number of hops will be greatly reduced. Therefore, in the solution of nRF24L01 response reply and response resender will not be used. Considering the feedback of data, an upward error correction protocol will be added in received.

In addition, based on FHSS the system adds DSSS. During sending data, each part of data will be sent on multiple channels to increase resistance to noise interference. In the system, every frame will be sent on five different channels. The receiver will pick up the number which occurs the most frequently. So it will not depend on whether the channel is good or not. If one of five channels is not interfered, the data will be transmitted successfully.

Besides, frequency hopping rules will changed adaptively during communication. That means frequency hopping rules will be planned every 3 to 6 seconds, during which the frequency which brings bad quality of communications encountered last time will be removed. So this time the system will communicate by same frame using the retransmission mechanism in 5 frequencies among the good frequencies, the error rate will be very low. Of course this is only avoid interference in channel, there are other factors influence the performance of the system. But this method can guarantee the reliability of communications with limited conditions.

3.4 Adaptive Frequency Hopping

Adaptive Frequency Hopping (AFH) is a active method to keep away from the interference source. Based on fixed frequency hopping rules, AFH adds adaptive frequency control. In this case, the rules of changing the carrier frequency are variable. Transmitter/Receiver can determine the channel quality by channel detection, then distinguish the channels by their communication qualities, so that we can select the channels with good communication quality, order the channels based on it and set out the new frequency hopping rule in the next time range. The frequency hopping will be base on the new rules after the rules are synchronous between the sender and the receiver.

Advantages: Excellent anti-interference ability. Avoiding automatically the frequencies that are interfered according to the real-time channel information. Even if a part of the selected frequencies are interfered, the reliability of the data transmission won't be affected.

Disadvantages: AFH leads to complicated protocols which are more difficult to realize, so the risk will be higher.

Adopting the AFH can make the system distinguish the interfered channels and avoid them to improve the anti-interference ability of the system. Comparing to the regular FH, the system is more stable and the bit-error rate will decline.

4 Closing

The medical wireless data transmission equipment's bit-error rates between 10e-5～10e-6, Can very good resistance to the interference of sporadic bluetooth ,Can be used in serious environment with electromagnetic interference, can be used as a medical electronic wireless data transmission. With characteristic of easy installation and maintenance, diffraction ability, network structure flexible, big range covers , suitable for some more, geographical environment and scattered complex applications, Achieve expected data transmission goal with high stability, high reliability and low cost, but in the test process there still has a frame of packet loss rate, so it need to be further perfected.

References

1. Qiao, T.-X.: Design of DS/FH signal transmission system based on FPGA. Electronic Design Engineering (2011-02)
2. Hao, W.: Wireless Environment Monitoring System Based on DSP and nRF24L01. Microcontrollers & Embedded Systems (2011-03)
3. Pan, Y., Guan, X., Zhao, R.: Design of smart wireless temperature measurement system based on NRF24L01. Electronic Measurement Technology (2010-02)
4. Tian, Z.-W.: Design of data communication system in wireless sensor networks based on Bluetooth. Transducer and Microsystem Technologies (2010-09)
5. Zhou, Y.: Implementation of RFID Wireless Communication System Based on FPGA. Modern Electronics Technique (2010-17)

Unknown Objects Grasp Planning Using Force/Torque Measurements

Shuang-quan Wen, Bing-xi Jia, Tie-jun Wu, and Shan Liu

Department of Control Science and Engineering, Zhejiang University,
Hangzhou 310027, China
sqwen@iipc.zju.edu.cn

Abstract. Geometrically unknown objects grasping is a challenging problem in robotic grasping and most researchers try to find suitable grasp through model reconstruction by vision. In this paper, a new three-fingered robotic hand grasping system is designed for three-dimensional geometrically unknown objects grasping. A six-axis force/torque sensor is mounted on each fingertip in our system to calculate the contact point and corresponding normal direction on the object once contact occurs. Thus it is capable of "exploration" the shape of unknown objects through continual contact to find suitable force-closure grasp configurations. We propose an intuitive algorithm to search possible good grasps and implement it in our system by means of a virtual instrument in LabVIEW™. The experimental results validate the feasibility of our grasp system.

Keywords: Grasp planning, Robotic hand, Unknown objects, Contact sensing.

1 Introduction

Grasping and exploration are two fundamental functionalities of human hands. "Grasping" indicates an action on an object consisting in preventing its motions relative to the hand [1], while "exploration" is to sense the real world by touching. Most of the planning algorithms assume that the object to be grasped is known a priori. For instance, the grasped object is a polyhedron in [2], while in [3], the object is approximated by a set of surface points and their corresponding normal direction. For novel or unknown objects grasping, most of literature resorts to computer vision. Stanfield [4] presented a general-purpose robotic grasping system for grasping 3D objects with unknown geometry. Wang et al. [5] proposed an unknown object grasp strategy by using 3D model reconstruction and a simulation environment. Although vision provides most of the world information for us, tactile sense is an important complement. This paper introduces a grasping system based on force/torque measurements. The grasp points are measured when fingertips touching the object, therefore geometrically unknown objects can be grasped by our system.

2 The System Setup

Our robotic hand grasping system consists of three robot manipulators as its fingers. One of the fingers is Motoman YP-HP3-A000 and the other two are built with

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 741–749.
springerlink.com

revolute modules from Shanghai xPartner robotics co., Ltd. A force/torque sensor (Gamma SI-65-5 from ATI Industrial Automation, Inc.) is mounted between the last joint and the fingertip of each finger. The center computer collects the force/torque applied on each fingertip using LabVIEW™ and receives joint angles for each finger through TCP message. The contact points and their corresponding normal direction are calculated to evaluate the grasp and generate path for each finger. Each finger is driven by a separated computer which receives planned path from the center computer.

3 Contact Points Estimation

We choose four revolute joints for each finger in this paper for simplification. The kinematics chain is represented in Fig. 1. For each finger, attach a frame S_i to the base of the finger and a frame $Tool_i$ to the fingertip. The homogeneous transformation matrix from base frame S_i to the global frame G is a fixed matrix, denotes as ${}^{G}_{S_i}\boldsymbol{T}$; while ${}^{S_i}_{Tool_i}\boldsymbol{T}$, the transformation matrix from $Tool_i$ to the base frame S_i, varies with all the revolute joints for the *i*th finger.

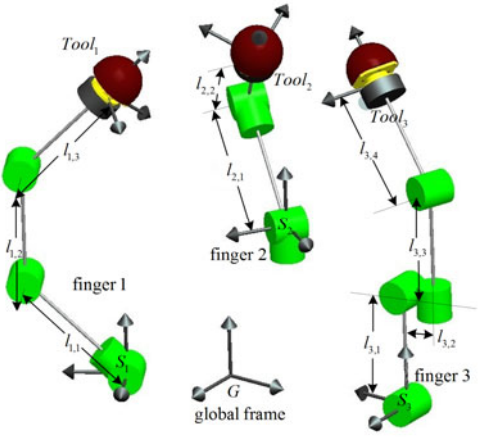


Fig. 1. Grasp coordinate frames

Each fingertip is mounted on an ATI force/torque sensor. Suppose contact occurs only on the semispherical part of the fingertip as is shown in Fig. 2a. Let $\boldsymbol{d}_i$ denote the center of the sphere w.r.t (with respect to) coordinate frame $Tool_i$, and R_i be the radius of the sphere. Then any contact point $\boldsymbol{c}_i$ at the surface must satisfy $(\boldsymbol{c}_i - \boldsymbol{d}_i)^T(\boldsymbol{c}_i - \boldsymbol{d}_i) = R_i^2$, and the unit inward normal $\boldsymbol{n}_i$ at this point is $(\boldsymbol{d}_i - \boldsymbol{c}_i)/R_i$. When soft finger contact model is adopted, the transducer reacts to applied force and torques at the contact point $\boldsymbol{c}_i$, i.e., $\boldsymbol{f}_i$ and $\boldsymbol{\tau}_i$. Note that $\boldsymbol{\tau}_i$ is a 3-dimension torque

along the contact $\boldsymbol{n}_i$ and $\boldsymbol{f}_i^T \boldsymbol{n}_i > 0$ since only push force emerges at the contact point. The *i*th ATI transducer readings, with total forces and torques w.r.t its coordinate frame $Tool_i$, would be $\boldsymbol{f}_{s,i} = \boldsymbol{f}_i + m_i {}^{Tool_i}_{G}\boldsymbol{R}\boldsymbol{g}$ and $\boldsymbol{\tau}_{s,i} = \boldsymbol{\tau}_i + \boldsymbol{r}_{g,i} \times m_i {}^{Tool_i}_{G}\boldsymbol{R}\boldsymbol{g} + \boldsymbol{c}_i \times \boldsymbol{f}_i$. Where $\boldsymbol{g}$ is the gravity acceleration at global coordinate frame G, m_i is the mass of the *i*th fingertip and $\boldsymbol{r}_{g,i}$ is the center of gravity w.r.t coordinate frame $Tool_i$, while ${}^{Tool_i}_{G}R$ is the matrix of relative orientation from the global frame G to frame $Tool_i$.

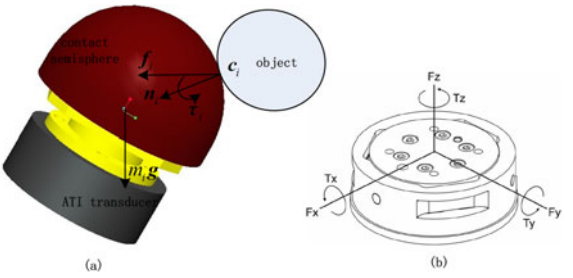


Fig. 2. (a) Force and torque at contat point $\boldsymbol{c}_i$. (b) ATI transducer measures 3-dimension forces and torques w.r.t. its coordinate frame $Tool_i$.

Let $\boldsymbol{f}_{m,i} = \boldsymbol{f}_{s,i} - m_i {}^{Tool_i}_{G}\boldsymbol{R}\boldsymbol{g} = [f_{x,i}\ f_{y,i}\ f_{z,i}]^{\mathrm{T}}$ and $\boldsymbol{\tau}_{m,i} = \boldsymbol{\tau}_{s,i} - \boldsymbol{r}_{g,i} \times m_i {}^{Tool_i}_{G}\boldsymbol{R}\boldsymbol{g} = [\tau_{x,i}\ \tau_{y,i}\ \tau_{z,i}]^{\mathrm{T}}$. All the fingertips have the same structure, i.e., $\boldsymbol{d}_i = [0\ 0\ z_0]^T$ and $R_i = R_0$ for $i = 1,2,3$. With the help of analysis in [6], the contact position $\boldsymbol{c}_i = [x_i\ y_i\ z_i]^{\mathrm{T}}$ can be calculated as the follows.

If $\boldsymbol{f}_{m,i}^{\mathrm{T}} \boldsymbol{\tau}_{m,i} = 0$, The contact position is

$$
\begin{aligned}
z_i &= \frac{z_0 f_{z,i}^2 + \tau_{y,i} f_{x,i} - \tau_{x,i} f_{y,i} \pm \sqrt{\varphi}}{f_{x,i}^2 + f_{y,i}^2 + f_{z,i}^2} \\
y_i &= \frac{\tau_{x,i} + f_{y,i} z_i}{f_{z,i}} \\
x_i &= \frac{-\tau_{y,i} + f_{x,i} z_i}{f_{z,i}}
\end{aligned}
\tag{1}
$$

where $\varphi = (\tau_{x,i} f_{y,i} - \tau_{y,i} f_{x,i} - z_0 f_{z,i}^2)^2 - (f_{x,i}^2 + f_{y,i}^2 + f_{z,i}^2)[\tau_{x,i} + \tau_{y,i} + f_{z,i}^2 (z_0^2 - R_0^2)]$. In this situation, we have $\boldsymbol{\tau}_i = \boldsymbol{0}_{3\times 1}$. Note that in the calculation, we suppose $f_{z,i} \neq 0$. If $f_{z,i} = 0$, we can select different equations and calculate alike. Two positions for the contact are found, one of them can be removed using $\boldsymbol{f}_i^T \boldsymbol{n}_i > 0$.

If $\boldsymbol{f}_{m,i}^T \boldsymbol{\tau}_{m,i} \neq 0$, let $\phi = (\boldsymbol{f}_{m,i}^T \boldsymbol{\tau}_{m,i})^2$,

$\psi = R_0^2 \boldsymbol{f}_{m,i}^T \boldsymbol{f}_{m,i} - \boldsymbol{\tau}_{m,i}^T \boldsymbol{\tau}_{m,i} + 2z_0(f_{x,i}\tau_{y,i} - f_{y,i}\tau_{x,i}) - z_0^2(f_{x,i}^2 + f_{y,i}^2)$ and

$k_i = -\dfrac{\mathrm{sgn}(\boldsymbol{f}_{m,i}^T \boldsymbol{\tau}_{m,i})}{\sqrt{2}R_0}\sqrt{-\psi + \sqrt{4R_0^2\phi + \psi^2}}$, then

$$\begin{aligned} \boldsymbol{c}_i &= \boldsymbol{H}_i^{-1}[\tau_{x,i}\ \tau_{y,i}\ \tau_{z,i} + k_i z_0]^T \\ \boldsymbol{H}_i &= \begin{bmatrix} k_i & f_{z,i} & -f_{y,i} \\ -f_{z,i} & k_i & f_{x,i} \\ f_{y,i} & -f_{x,i} & k_i \end{bmatrix} \end{aligned} \tag{2}$$

It is easy to prove that $\sqrt{4R_0^2\phi + \psi^2} \geq \psi$ since $\phi \geq 0$, thus k_i is a real number. From [6], for any nonzero contact force $\boldsymbol{f}_{m,i}$, if $\boldsymbol{f}_{m,i}^T \boldsymbol{\tau}_{m,i} \neq 0$, $\boldsymbol{H}_i$ if of full rank and $\boldsymbol{c}_i$ is conformed. In this situation, $\boldsymbol{\tau}_i = k_i[x_i\ y_i\ z_i - z_0]^T$.

4 Grasp Planning Scheme

4.1 Contact Force Constraint

Obviously, the contact force and torque applied on the object at each contact point $\boldsymbol{c}_i$ are $-\boldsymbol{f}_i$ and $-\boldsymbol{\tau}_i$. All of them are expressed in the coordinate frame $Tool_i$, thus we have to translate them into a unified frame to calculate the total force and torque. Let ${}^G\boldsymbol{n}_i$ and ${}^G\boldsymbol{r}_i$ be the unit inward normal of the object surface at the contact point and the contact position w.r.t the global frame G, respectively. We have ${}^G\boldsymbol{n}_i = -{}_{Tool_i}^{G}\boldsymbol{R}\boldsymbol{n}_i$ and $[{}^G\boldsymbol{r}_i\ \ 1]^{\mathrm{T}} = {}_{Tool_i}^{G}\boldsymbol{T}[\boldsymbol{c}_i\ \ 1]^{\mathrm{T}}$. To remedy the drawback of frame and unit variance for contact matrices, we adopted the contact matrix form in [7] for the *i*th contact as

$$\boldsymbol{A}_i = \begin{bmatrix} \boldsymbol{R}\,{}^G\boldsymbol{n}_i & \boldsymbol{R}\boldsymbol{o}_i & \boldsymbol{R}\boldsymbol{t}_i & \boldsymbol{0}_{3\times 1} \\ ({}^G\boldsymbol{r}_i - \boldsymbol{r}_0)\times {}^G\boldsymbol{n}_i & ({}^G\boldsymbol{r}_i - \boldsymbol{r}_0)\times \boldsymbol{o}_i & ({}^G\boldsymbol{r}_i - \boldsymbol{r}_0)\times \boldsymbol{t}_i & {}^G\boldsymbol{n}_i \end{bmatrix} \tag{3}$$

Where $\boldsymbol{r}_0 = ({}^G\boldsymbol{r}_1 + {}^G\boldsymbol{r}_2 + {}^G\boldsymbol{r}_3)/3$ is centroid of the contact positions, $\boldsymbol{o}_i$ and $\boldsymbol{t}_i$ are two unit tangent vectors satisfying ${}^G\boldsymbol{n}_i = \boldsymbol{o}_i \times \boldsymbol{t}_i$ and $R = (\|{}^G\boldsymbol{r}_1 - \boldsymbol{r}_0\| + \|{}^G\boldsymbol{r}_2 - \boldsymbol{r}_0\| + \|{}^G\boldsymbol{r}_3 - \boldsymbol{r}_0\|)/3$ is the average distance from the contact positions to the centroid.

The contact force applied on the object at ${}^G\boldsymbol{r}_i$ can be expressed in the frame $\{{}^G\boldsymbol{n}_i, \boldsymbol{o}_i, \boldsymbol{t}_i\}$ by ${}^o\boldsymbol{f}_i = [f_{in}\ f_{io}\ f_{it}\ f_{im}]^{\mathrm{T}}$. Where f_{in} is the normal force; f_{io} and f_{it} are tangential force; f_{im} is the spin moment around ${}^G\boldsymbol{n}_i$. We collect the force and

torque into a wrench as $\boldsymbol{w} = [\boldsymbol{f}^{\mathrm{T}}\ \boldsymbol{\tau}^{\mathrm{T}}]^{\mathrm{T}}$, the total wrench results from three fingertips is

$$\boldsymbol{w}_{res} = \sum_{i=1}^{3} \boldsymbol{A}_i\, {}^{0}\boldsymbol{f}_i \tag{4}$$

The contact force ${}^{o}\boldsymbol{f}_i$ must satisfy the contact constraints (soft finger contact with linear model [7, 8] is considered in this paper):

$$F_i = \{ {}^{o}\boldsymbol{f}_i \mid f_{in} \geq 0, \frac{\sqrt{f_{io}^2 + f_{it}^2}}{\mu_i} + \frac{|f_{im}|}{\mu_{si}} \leq f_{in} \} \tag{5}$$

Where μ_i and μ_{si} are the coefficients of tangential and torsion friction for the *i*th contact, respectively.

4.2 Force-Closure Grasp Test

Suppose all μ_i and μ_{si} are estimated beforehand. When all the three fingers contact on the object, then contact matrixes are determined. The ability to resist external wrenches depends on the contact matrixes, which is determined by the contact positions on the object. A grasp is force-closure [1] if and only if for any external wrench $\boldsymbol{w}_{ext} \in \mathbb{R}^6$, there exist contact force ${}^{o}\boldsymbol{f}_i \in F_i$ such that

$$\sum_{i=1}^{3} \boldsymbol{A}_i\, {}^{0}\boldsymbol{f}_i + \boldsymbol{w}_{ext} = \boldsymbol{0}_{6\times 1} \tag{6}$$

In [8], the friction cone for *i*th contact can be linearized by a polyhedral bicone with $m+2$ vertices $\boldsymbol{s}_{i,j}$. $\boldsymbol{s}_{i,j} = [1\ \ \mu_i \cos\frac{2j\pi}{m}\ \ \mu_i \sin\frac{2j\pi}{m}\ \ 0]^T$, $j = 1, 2, \cdots m$ and $\boldsymbol{s}_{i,j} = [1\ \ 0\ \ 0\ \ \pm\mu_{si}]^T$, $j = m+1, m+2$. Then the primitive contact wrenches [9] $\boldsymbol{w}_{i,j} = \boldsymbol{A}_i \boldsymbol{s}_{i,j}$. Let $\boldsymbol{P} = \frac{1}{3(m+2)} \sum_{i=1}^{3} \sum_{j=1}^{m+2} \boldsymbol{w}_{i,j}$, if $\boldsymbol{P} = \boldsymbol{0}_{6\times 0}$, the grasp is force-closure, otherwise, formulate the following linear programming (LP) problem

$$\begin{cases} \max \quad z = -\boldsymbol{P}^T \boldsymbol{x} \\ \text{s.t.} \quad (\boldsymbol{w}_{i,j} - \boldsymbol{P})^T \boldsymbol{x} \leq 1 \text{ for } i = 1, 2, 3.\ j = 1, 2, \cdots, m+2 \end{cases} \tag{7}$$

From [10], when the solution of the LP problem with $z_{\max} < 1$, the grasp achieves force-closure.

4.3 Grasp Points Selection

Finding grasp points which form a force-closure grasp is not a simple task because the object is geometrically unknown and grasp configurations are numerous, let alone finding the best grasp. We describe our intuitive method to construct possible good

grasps here. Firstly, we define a contact plane S which contains all three contact points ${}^{G}\boldsymbol{r}_i$ as is depicted in Fig. 3. Suppose the gravity direction is along the $-z$ axis of the global frame G. In our opinion, when three contact points form a largest equilateral triangle, with all normal direction ${}^{G}\boldsymbol{n}_i$ converge to a point just above the center of gravity of the grasped object, we achieve the "best" grasp. Obviously, such a rigorous grasp can be hardly obtained in practical application. We define the following simple grasp quality criterion to evaluate the grasp.

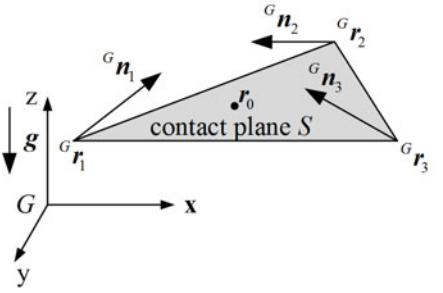


Fig. 3. Grasp quality criterion

$$J = \tfrac{1}{2}\left|\left({}^{G}\boldsymbol{r}_3 - {}^{G}\boldsymbol{r}_1\right)\times\left({}^{G}\boldsymbol{r}_2 - {}^{G}\boldsymbol{r}_1\right)\cdot\boldsymbol{e}_{3,3}\right| + \tfrac{R}{2}\sum_{i=1}^{3}\left(\boldsymbol{r}_0 - {}^{G}\boldsymbol{r}_i\right)^{\mathrm{T}}{}^{G}\boldsymbol{n}_i \tag{8}$$

where $\boldsymbol{e}_{3,3} = [0\ 0\ 1]^{\mathrm{T}}$ is the third column of 3-by-3 identity matrix, the operators $\times$ and $\cdot$ denote cross and dot product of two vectors respectively while |x| is the absolute value of x. The first term is the projection area of the triangle on the x-y plane. The second term measures the convergent level of normal direction to the centroid of the contact positions. Note that grasp quality J has the dimension of area and the bigger J will generate the better grasp intuitively. We use the following procedure to select an appropriate grasp.

Step 1) Select a plane parallel to the x-y plane at random. Let the center of each semispherical fingertip approach to the object spaced 120° apart on the plane.

Step 2) When all fingertips contact with the object, calculate the contact positions ${}^{G}\boldsymbol{r}_i$ and adjust the contact plane S to parallel to the x-y plane.

Step 3) Fix two fingers and move the third finger on the plane S left or right at a certain distance until obtaining maximum J from (8).

Step 4) Change the moving finger and return to Step 3. If no further improvement exists, solve the LP problem of (7).

Step 5) If $z_{\max} < 1$, the grasp points is found. Otherwise, go back to Step 1.

5 Experimental Results

To calculate the transformation matrix ${}^{G}_{S_i}T$ from base frame S_i to the global frame G, we first paste three marks (which emit infrared light) on the base of each finger and calibrate the position and orientation using NDI 3D Investigator (Northern Digital Inc.) as is shown in Fig. 4.

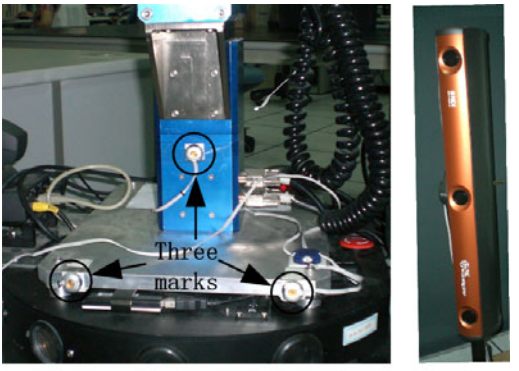


Fig. 4. (a) Three marks are pasted on the base of each finger. (b) NDI 3D Investigator.

Then, the force/torque applied on the fingertips is collected by means of virtual instruments within LabVIEW™ environment in the center computer. All the algorithms addressed above are also implemented by LabVIEW™. A box grasping task is implemented by our grasp system in the experiment. Suppose the tangential coefficient μ_i and torsion friction μ_{si} are 0.15 and 0.005m respectively for all the fingers and linearize the friction cone by a polyhedral bicone with 12 vertices. The box is a cube with approximate 0.3m on each side. After the "exploration", our grasping system finally found three contact points as list in Table 1. The initial and final scenes are snapshot as is shown in Fig. 5.

Table 1. The final grasp configuration

i	${}^{G}\boldsymbol{r}_i$	${}^{G}\boldsymbol{n}_i$
1	$[0.3767\ \ 0.2096\ \ 0.1528]^{\mathrm{T}}$	$[0.1262\ \ 0.9851\ \ 0.1170]^{\mathrm{T}}$
2	$[0.2268\ \ 0.3691\ \ 0.1550]^{\mathrm{T}}$	$[0.9955\ \ 0.0764\ -0.0563]^{\mathrm{T}}$
3	$[0.4062\ \ 0.5554\ \ 0.1741]^{\mathrm{T}}$	$[-0.0558\ -0.9984\ \ 0.0079]^{\mathrm{T}}$
	$J = 0.064\ \mathrm{m}^2$	$z_{max} = 0.8669$

Fig. 5. Snapshots of the (a) inital and (b) final scenes

6 Concluding Remarks

In this paper, a three-fingered robotic hand grasping system is established to grasp geometrically unknown objects. Compared to previous work based on computer vision, our system is simpler and model reconstruction is not necessary. Vision cannot sense the object weight and material on the other hand. Our grasping system calculates the contact point and corresponding normal direction on the object by touching the object, thus partial geometrical information of the grasped object is obtain by this "exploration" process. During the "exploration", we proposed an intuitive algorithm to find good grasps more effectively. Future work will extend grasp planning in this paper to smooth and robust grasping control. Meanwhile, visual feedback could be added to approximate the position and size of the object as people do.

References

1. Bicchi, A.: Hands for dexterous manipulation and robust grasping: A difficult road toward simplicity. IEEE Transactions on Robotics and Automation 16(6), 652–662 (2000)
2. Ding, D., Liu, Y.H., Zhang, J., Knoll, A.: Computation of fingertip positions for a form-closure grasp. In: Proceedings of the 2001 International Conference on Robotics and Automation, pp. 2217–2222. IEEE Press, New York (2001)
3. Roa, M.A., Suárez, R.: Finding locally optimum force-closure grasps. Robotics and Computer-Integrated Manufacturing 25(3), 536–544 (2009)
4. Stansfield, S.: Robotic grasping of unknown objects: A knowledge-based approach. The International Journal of Robotics Research 10(4), 314–326 (1991)
5. Wang, B., Jiang, L., Li, J., Cai, H.: Grasping unknown objects based on 3D model reconstruction. In: Proceedings of the 2005 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, pp. 461–466. IEEE Press, New York (2005)
6. Han, Z., Wang, T., Zhang, Y., Liu, J.: Selection of calculation model for contact sensing of robot hand using force/torque measurement. Journal of Beijing University of Aeronautics and Astronautics 30(05), 400–404 (2004) (Chinese)
7. Zheng, Y., Qian, W.H.: Improving grasp quality evaluation. Robotics and Autonomous Systems 57(6-7), 665–673 (2009)

8. Zheng, Y., Qian, W.: Linearizing the soft finger contact constraint with application to dynamic force distribution in multifingered grasping. Science in China Series E: Technological Sciences 48(2), 121–130 (2005)
9. Liu, Y.H.: Qualitative test and force optimization of 3-D frictional form-closure grasps using linear programming. IEEE Transactions on Robotics and Automation 15(1), 163–173 (1999)
10. Zheng, Y., Qian, W.H.: Simplification of the ray-shooting based algorithm for 3-D force-closure test. IEEE Transactions on Robotics 21(3), 470–473 (2005)

Vision-Aided Spiral-Surge Algorithm for Odor Source Localization in Indoor Natural Ventilated Environments

Yang Wang, Qing-Hao Meng, and Ming Zeng

School of Electrical Engineering and Automation, Tianjin University
Tianjin, China
{yangwang,qh_meng,zengming}@tju.edu.cn

Abstract. Mobile robot odor source localization (OSL) in indoor natural airflow environments is addressed. The fluctuant wind direction and local concentration maxima in indoor natural environments pose great challenges for a robot to localize an odor source. A novel OSL method using vision-aided spiral-surge algorithm is proposed, in which the top-down visual attention mechanism is applied to deal with visual clues. Real-robot experiments show that the proposed OSL method has higher success rate and shorter search time than the standard spiral-surge algorithm.

Keywords: Mobile robot, Odor source localization, Spiral-surge behavior, Visual attention mechanism.

1 Introduction

Olfaction is one of the important senses of animals. From 1990s, researchers began to use mobile robots equipped with gas and/or wind sensors to imitate animals' behaviors of finding odor source. Such study is called odor source localization (OSL) [1]. The potential applications of the OSL include locating toxic gas leakage sources, searching for survivors after disaster and humanitarian de-mining.

In indoor natural airflow environments, the detected wind direction is generally not reliable due to complicated boundary conditions, and the local concentration maxima caused by the odor accumulation often exist. Therefore, the mobile robot based OSL in such indoor environments is not a trivial task. To localize a previously unknown odor source, normally gas/odor and wind sensors are used. In recent years, vision sensing has also been considered by researchers. Martinez et al. [2] used vision together with olfaction to locate an ethanol source placed in front of a salient object, and a bottom-up visual attention mechanism was adopted. Ishida et al. [3] proposed a behavior-based architecture to combine vision and olfaction information. Kowadlo et al. [4] regarded the crack as the suspect leakage source and combined the vision and naïve physics model to locate the source. Jiang et al. [5] used the top-down visual attention mechanism (TDVAM) and shape analysis method for finding the suspect odor source in a complex indoor scene.

In this paper, we propose a vision-aided spiral-surge (SS) algorithm for the mobile robot based OSL in indoor natural airflow environments, where the TDVAM is used to correct the surge direction by finding the potential odor source.

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 751–755.
springerlink.com

2 Vision-Aided Spiral-Surge Algorithm

The proposed vision-aided SS (V-SS for short) algorithm mainly includes the TDVAM based visual search and the SS behavior. When the odor is detected, the robot moves according to the standard SS algorithm. If the difference between the upwind direction and the direction of the potential odor source found by the TDVAM algorithm is larger than a threshold (denoted using θ_{th}), the robot surges toward the potential odor source. Otherwise, the robot surges upwind.

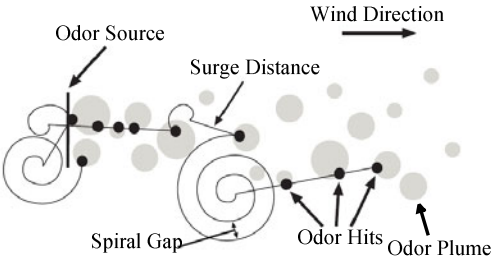


Fig. 1. Trajectory of the SS algorithm [1]

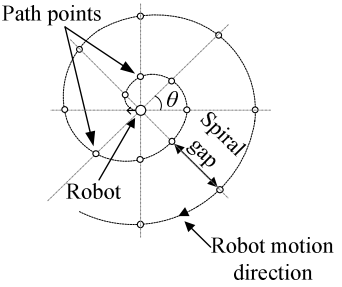


Fig. 2. Path points of the spiral motion [6]

Figure 1 shows the schematic diagram of the standard SS algorithm [1]. The SS algorithm [1] consists of the outward spiral motion and upwind surge motion. When the odor is detected, the robot surges upwind for a set distance. If another detection of the odor occurs during the surge, the robot resets the surge distance and continues to surge upwind. After the surge distance has been reached and no more new odor detection event happens, the robot starts the spiral motion to re-find odor plume. In our experiments, the robot was controlled by following a number of path points P_j (j=1, 2,…) on the spiral line to achieve the spiral motion as shown in Fig. 2. P_j is expressed as follow,

$$P_j = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \frac{j \cdot \theta \cdot d_{gap}}{360} \begin{bmatrix} \cos(\theta_r - j \cdot \theta) \\ \sin(\theta_r - j \cdot \theta) \end{bmatrix} . \quad (1)$$

Where (x_0, y_0) is the coordinate of the center. θ is the distribution angle of the path points, i.e. the path points are selected every θ degrees around the center. d_{gap} denotes the spiral gap. θ_r stands for the robot heading when the spiral motion begins.

The adopted TDVAM calculation model [5] was inspired by the "central-periphery difference" mechanism in biological visual system. Five spatial scales are created using Gaussian pyramid for the captured image. The priori knowledge of the odor source, including one brightness feature, two color features and four orientation features extracted from each scale, was used to improve the relevance of the salient regions to the potential odor source.

Suppose the central-periphery difference operation is denoted using the symbol Θ, and the symbols c and s stand for the center and periphery, respectively. Then the comparison maps of the brightness, color and orientation can be derived as follows,

$$I(c,s) = \left| I(c) \Theta I(s) \right| \quad , \tag{2}$$

$$C_{rg}(c,s) = \left| C_{rg}(c) \Theta C_{rg}(s) \right| \quad , \tag{3}$$

$$C_{by}(c,s) = \left| C_{by}(c) \Theta C_{by}(s) \right| \quad , \tag{4}$$

$$O_{\theta}(c,s) = \left| O_{\theta}(c) \Theta O_{\theta}(s) \right| \quad . \tag{5}$$

Where the symbols *I, C* and *O* represent the brightness, color and orientation features, respectively. The subscripts *r*, *g*, *b* and *y* stand for red, green, blue and yellow, respectively. The subscript $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ denotes the preferred orientation.

The salient regions were obtained by merging the comparison maps according to the mergence weights. Once the odor source is found in the image, its direction in the global coordinate system can be calculated.

3 Experiments and Results

The proposed V-SS algorithm and the standard SS algorithm were compared using real-robot experiments. The mobile robot platform and the odor source are shown in Fig. 3. A gas sensor (MiCS 5135, e2v Technologies (UK) Ltd.), an anemometer (Windsonic, Gill Instruments Ltd.), a laser rangefinder (LMS200, Sick AG), an electronic compass, and a PTZ camera (EVI-D100P, Sony) were mounted on the robot. The maximal speed of the robot was 0.4 m/s. The PTZ camera had a resolution of 320×240. An ultrasonic humidifier filled with liquid ethanol was used as the odor source.

The experiments were conducted in an 8m×4.5m area in our lab as shown by the red dash line in Fig. 4. During the experiments, the left door and five windows were open, other windows and the right door were closed, and the air conditioner was off. Twenty trials were implemented for each algorithm. For both algorithms, the spiral gap and the surge distance were set to 0.3 m and 0.9m, respectively. The threshold θ_{th} was set to 30°. A trial was considered as success once the robot was 0.5 m from the real odor source. If the robot couldn't successfully locate the odor source in 240 seconds, the trial was regarded as fail. Table 1 shows the experimental results. It

should be noted that only successful trials were considered in calculating the average tracing time and standard deviation. It can be easily seen that the proposed V-SS performs better than the SS. The smaller standard deviation of the V-SS implies that the algorithm is less impacted by the wind fluctuation than SS.

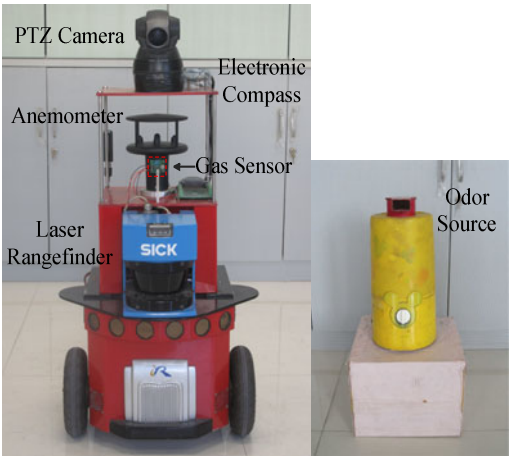


Fig. 3. Mobile robot platform and the odor source

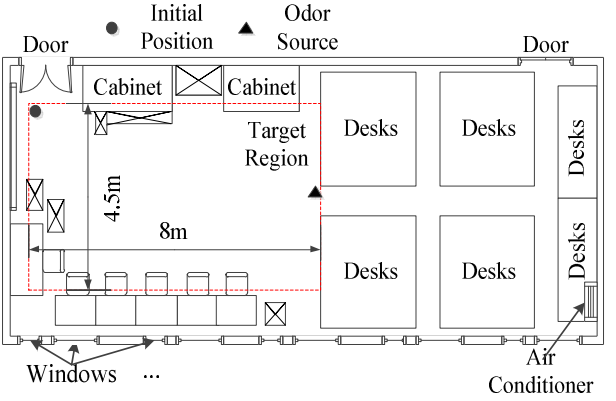


Fig. 4. The plan of the experiment site

Table 1. Comparison of the success rate, average tracing time and standard deviation

	Success rate (%)	Average tracing time (s)	Standard deviation(s)
V-SS	100	74.35	30.73
SS	60	146	52.54

4 Conclusion

Owing to the frequently fluctuant wind, eddies in corners and accumulated odor concentration in natural indoor airflow environments, the standard spiral-surge (SS) algorithm is hard to tracing plume to the source. The top-down visual attention mechanism could be a good choice for assisting the SS algorithm. Real-robot experiments show that the visual-aided SS algorithm can locate the odor source faster than the standard SS algorithm and is less impacted by the wind fluctuation.

Acknowledgements. This work is supported by China '863' Program (No. 2007AA04Z219), the NSFC projects (No.60875053, 60802051), Tianjin NSF project (09JCYBJC02100) and the Program for New Century Excellent Talent in University.

References

1. Hayes, A.T., Martinoli, A., Goodman, R.M.: Distributed odor source localization. IEEE Sensors Journal 2, 260–271 (2002)
2. Martinez, D., Perrinet, L.: Cooperation between vision and olfaction in a koala robot. In: Report on the 2002 Workshop on Neuromorphic Engineering, pp. 51–53 (2002)
3. Ishida, H., Tanaka, H., Taniguchi, H.: Mobile robot navigation using vision and olfaction to search for a gas/odor source. Autonomous Robot 20, 231–238 (2006)
4. Kowadlo, G., Rawlinson, D., Russell, R.A., Jarvis, R.: Bi-modal search using complementary sensing (olfaction/vision) for odour source localisation. In: Proc. of the IEEE Int. Conf. on Robotics and Automation, pp. 2041–2046. IEEE Press, New York (2006)
5. Jiang, P., Meng, Q.H., Zeng, M.: Mobile robot gas source localization via top-down visual attention mechanism and shape analysis. In: 8th World Congress on Intelligent Control and Automation (WCICA), pp. 1818–1823. IEEE Press, New York (2010)
6. Wang, Y., Meng, Q.H., Yang, W.X., Zeng, M.: Experimental comparison of spiral and zigzag algorithms for odor plume finding in an outdoor natural airflow environment. In: 5th International Conference on the Advanced Mechatronics (ICAM 2010), pp. 171–176 (2010)

Simulation and Analysis of the Stability of a PID Controller for Operation of Unmanned Aerial Vehicles*

Abolfazl Sheibani and Mohammad Ali Pourmina

Department of Mechateronics Engineering
Science and Research Branch-Islamic Azad University (SRBIAU)
Tehran, Iran
afshin_shei@yahoo.com, pourmina@srbiau.ac.ir

Abstract. In the study outlined we examined the state space model of an Unmanned Aerial Vehicles (UAV) along with all its transfer functions around both longitudinal and lateral axes. We then took advantage of the MATLAB software to design and simulate our proposed PID controller through either classic or poles placement methods across the longitudinal axis of the UAV when the motion dynamics are be known. Given the known defaults for designing a controller, the stability of the designed PID controller was precisely analyzed for each method separately using the results obtained from the MATLAB assisted analysis as well as the response step of both open loop and close loop of the system around the longitudinal axis of the UAV. The correctness of the designed PID controller eventually became evident through comparison of default values of the design with those determined after application of the controller.

Keywords: PID Controller, Transfer Function, Unmanned Aerial Vehicles (UAV), Poles Placement.

1 Introduction

An autopilot system is essential to steer and navigate an unmanned aerial vehicle. Its controller compartment is one of the most critical components of an autopilot system. Relying on its controller, an autopilot can steer and navigate the airplane on the predefined points via setting of certain parameters such as airspeed, altitude, roll, pitch and yaw angles.

For an autopilot system to accomplish these tasks a PID controller is required. The necessary orders to shift control levels of the airplane, including actuators of ailerons, elevators and radars, are controlled by a PID loop. It helps the drone to find the balanced state around its roll and pitch axes. Figure 1.a shows the actuators of UAV.

* Proportional–Integral–Derivative Controller (PID).

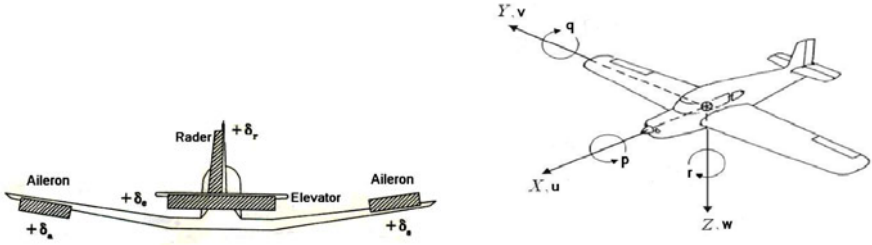
T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 757–765.
springerlink.com

All of the effective and used parameters for an UAV are summarized in table 1.

Table 1. The used parameters for an UAV

	Parameter	Description
1	ϕ	Roll angle, The angle between the aircraft's x axis and the horizontal plane.
2	θ	Pitch angle, The angle between the aircraft's y axis and the horizontal plane.
3	φ	Yaw angle, The angle between the aircraft's z axis and the horizontal plane.
4	p	Changes rate of the roll angle
5	q	Changes rate of the pitch angle
6	r	Changes rate of the yaw angle
7	u	Linear speed of the aircraft towards x axis
8	v	Linear speed of the aircraft towards y axis
9	w	Linear speed of the aircraft towards z axis
10	δ_E	Deviation elevator angle
11	δ_A	Deviation aileron angle
12	δ_R	Deviation rader angle
13	h	Altitude

Figure 1.b shows directions of some of these parameters over the UAV fuselage.



(a) (b)

Fig. 1. (a) The main control levels of an UAV. (b) Direction of the used parameters for an UAV.

The controller compartment of any autopilot system is typically composed of two separate PID controllers: a longitudinal and lateral controllers. Thus, two PID controllers are necessary for steering and navigating an UAV. The longitudinal PID controller is designed to control and set speed, pitch angle and altitude of an UAV. The PID loop is formed in proportion to the relevant errors of the indicated parameters and it sends the necessary orders using factors of the PID controller, including k_p (proportional gain), k_i (integral gain) and k_d (derivative gain), to actuators of the elevator and throttle and then corrects their errors. For example, in order to control the airspeed, the output of the PID controller loop sends the necessary orders to the throttle; hence, the propulsion of the airplane is set to the desired scale by increasing or decreasing revolution rate.

The lateral PID controller is designed to manage and set heading, roll angle and yaw angle. The necessary orders are sent to actuators of aileron and rader using PID controller lateral loop. Like the PID controller longitudinal loop, this loop uses its

controlling coefficients to correct the errors of desired parameters. Thus, the UAV is steered and navigated by the PID controller 's lateral and longitudinal loops across the indicated coordinates. We next employed an actual UAV model to simulate and analyze the stability of our PID controller. The various operations implemented on the proposed UAV are explained here.

2 Analysis of the Actual Model of an UAV

The state space model of an UAV should be specified in order to analyze the model of an UAV as eq. 1.

$$\dot{X} = AX + Bu \qquad Y = CX + Du \tag{1}$$

Specifying A, B, C and D factors of the dynamic model ascertains motion of the UAV. Thus, lateral and longitudinal equations of motion of an actual UAV model are considered separately as follows:

The longitudinal equations of motion for an UAV with known A, B, C and D factors are as: eq. 2, 3, 4, 5, 6 and 7, [3,4].

$$X = [u\ w\ q\ \theta\ h]^T \tag{2}$$

$$A = \begin{bmatrix} -0.6284 & -2.6762 & 0.0 & -9.81 & 0.0 \\ -0.2232 & -2.6764 & 16 & 0.0 & 0.0 \\ 0.0102 & -0.1384 & -0.1575 & 0.0 & 0.0 \\ 0.0 & -1.0 & 0.0 & 16 & 0.0 \end{bmatrix} \tag{3}$$

$$B = [-0.0610 \quad 0.1543 \quad 0.9686 \quad 0.0 \quad 0.0]^T \tag{4}$$

$$C = [0 \quad 0 \quad 0 \quad 1 \quad 0] \tag{5}$$

$$D = [0 \quad 0] \tag{6}$$

$$u = \delta_E \tag{7}$$

The lateral equations of motion for an UAV with known A, B, C and D factors are as :eq. 8, 9, 10, 11, 12 and 13, [3,4].

$$X = [v\ p\ r\ \phi\ \varphi]^T \tag{8}$$

$$A = \begin{bmatrix} -0.17 & 0.0 & -1.0 & -0.61 & 0.0 \\ -3.71 & -2.43 & 0.36 & 0.0 & 0.0 \\ 3.71 & -0.27 & -0.33 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \end{bmatrix} \tag{9}$$

$$B = [0\ 0.05;\ 2.62\ 1.02;\ -0.06\ -2.1;\ 0.0\ 0.0]^T \tag{10}$$

$$C = [0\,0\,0\,1\,0;\ 0\,0\,0\,0\,1]^T \tag{11}$$

$$D = [\,0\ 0\,] \tag{12}$$

$$u = [\delta_A \quad \delta_R]^T \tag{13}$$

However, it must be mentioned that measuring all of the equations of dynamic states of any UAV alongside the lateral and longitudinal directions will be easy through specifying airspeed, altitude and weight of aircraft as inputs of Aerosim[1] simulation software and executing its Trim_Aeresond.m file. Thus, execution of Trim-_Aerosonde.m code file in MATLAB software makes possible calculation of all necessary data for either lateral and longitudinal dynamics of UAV such as state matrix (A), controllability matrix (B) and observability matrix (C). Initially, transfer functions of the UAV in both longitudinal and lateral parts should be specified in order to simulate and analyze stability of the designed PID controller [7].

Introduction of Transfer Function of UAV

The longitudinal transfer function of the proposed UAV is indicated in eq. 14:

$$\frac{\theta}{\delta_E} = \frac{2.665e-0.15s\,\hat{}\,4+0.9686s\,\hat{}\,3+3.179s\,\hat{}\,2+1.029s}{s\,\hat{}\,5+3.462s\,\hat{}\,4+3.819s\,\hat{}\,3+2.099s\,\hat{}\,2+0.5708s} \tag{14}$$

Where, the final eq. 15 is obtained by removing "s" from either numerator and denominator and approximation of the eq.14:

$$\frac{\theta}{\delta_E} = \frac{0.9686s\,\hat{}\,2+3.179s\,\hat{}\,1+1.029}{s\,\hat{}\,4+3.462s\,\hat{}\,3+3.819s\,\hat{}\,2+2.099s\,\hat{}\,1+0.5708} \tag{15}$$

Given inputs of Aileron yaw angle (δ_A) and Radar yaw angle (δ_R), the lateral transfer functions of the proposed UAV are:

$$\text{Roll}: \left(\frac{\phi}{\delta_A}\right): \frac{2.62s\,\hat{}\,2+1.288s+9.641}{s\,\hat{}\,4+2.93s\,\hat{}\,3+5.078s\,\hat{}\,2+7.907s+0.06789} \tag{16}$$

$$\text{Yaw}: \left(\frac{\varphi}{\delta_A}\right): \frac{-0.06s\,\hat{}\,3-0.8634s\,\hat{}\,2-0.145s-5.794}{s\,\hat{}\,5+2.93s\,\hat{}\,4+5.078s\,\hat{}\,3+7.907s\,\hat{}\,2+0.06789s} \tag{17}$$

$$\text{Roll}: \left(\frac{\phi}{\delta_R}\right): \frac{1.02s\,\hat{}\,2-0.4315s-4.073}{s\,\hat{}\,4+2.93s\,\hat{}\,3+5.078s\,\hat{}\,2+7.097s+0.06789} \tag{18}$$

$$\text{Yaw}: \left(\frac{\varphi}{\delta_R}\right): \frac{-2.1s\,\hat{}\,3-5.55s\,\hat{}\,2-0.4135s+2.444}{s\,\hat{}\,5+2.93s\,\hat{}\,4+5.078s\,\hat{}\,3+7.907s\,\hat{}\,2+0.06789s} \tag{19}$$

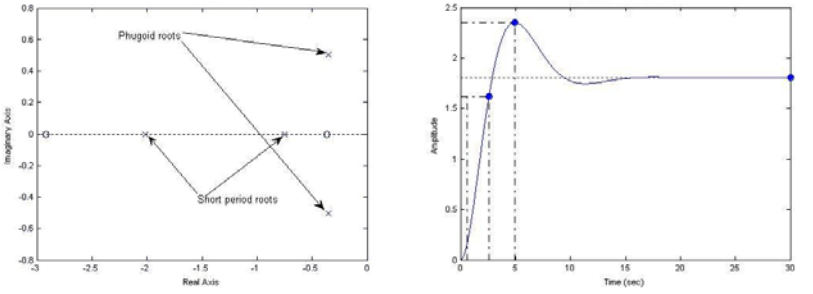
After specification of the transfer functions of the proposed UAV, we can begin to analyze and simulate the dynamic stability of UAV using MATLAB software with regard to PID controller in the longitudinal axis of the vehicle [2].

[1] Aeronautical Simulation.

3 Analysis of the Dynamic Stability in the Longitudinal Axis of UAV

Given eq.15, Figure 2.a indicates the special amounts and poles placement after being executed in MATLAB software.

Generally, the controller for UAV is designed aiming to bring about stability for the drone in the longitudinal axis. However, even if this is not the case (because of maneuverability of the UAV), the automatic control system of UAV is bound to maintain such state in order to specify longitudinal stability condition of the UAV over all situations and flight models. Step response of the open loop of the system to the longitudinal axis of the UAV is shown in Figure 2.b.



(a) (b)

Fig. 2. (a) Status of poles and zeros in the transfer function for δ_E as its input and θ as its output (b) Step response of open loop of the system around longitudinal axis of the UAV

Figure 2 shows that the system enjoys BIBO stability, but designer's expectations in terms of the longitudinal stability of UAV have not been met yet [2,3].

4 Designing the PID Controller for the Proposed UAV

Now, regarding the below defaults, we began to design a PID controller around the longitudinal axis of an UAV. Max Overshoot: > 3%. Rise Time: > 3 sec. Setting Time: 5< sec. The steady state error goes to zero.

Keeping correlation and interrelation of PID controller factors (k_p, k_i and k_d) in Table 2, we materialized the above designed criteria by setting PID controller factors.

Table 2. Changes of the main flight parameters by setting PID Controller factors

Gain type of PID	Rise Time	Overshoot	Setting time	S-S error
k_p	Decrease	Increase	Small change	Decrease
k_i	Decrease	Increase	Increase	Eliminate
k_d	Small change	Decrease	Decrease	Small change

Initially, we started and continued with k_p, k_I, k_d= 1. Response steps of both close and open loops (1) of the system are compared in Figure 3.

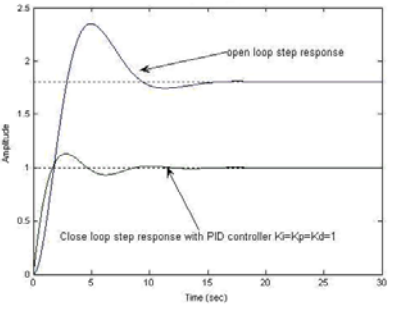
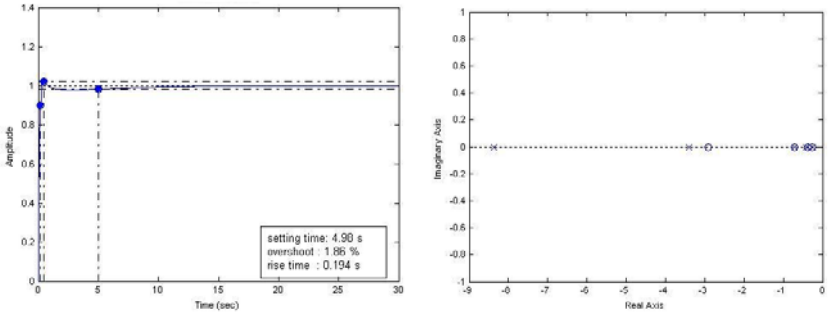


Fig. 3. Comparing response step of both close and open loops of the system using the PID controller

Eventually, as seen in the Figure 4.a, the outlined expectation of the design materialized with k_i =2, k_d =10 and k_p=10.

Figure 4.b shows system zeros and poles placements subsequent to application of the PID controller with unity negative feedback.



(a) (b)

Fig. 4. (a) Close loop Step Response with PID controller k_i =2, k_d =10 and k_p=10. (b) Close loop Pole-Zero Map(plant + PID controller + unity negative feedback).

For SETP, UNIRAMP and PARABOLIC as inputs, we had the steady state error as eq. 20:

$$\begin{aligned} STEP\_S\_S\_ERROR &= 0.0; \\ RAMP\_S\_S\_ERROR &= 0.5; \\ PARABOLIC\_S\_SERROR &= INF; \end{aligned} \tag{20}$$

Keeping in mind that the system belongs to degree 1 and, inspired by the results gained from indicated PID controller in figures 4, it can be concluded that all defaults for the UAV such as setting time, overshoot and rise time were met [3].

5 Designing and Analyzing the Dynamic Stability of the PID Controller via Poles Placement Method

Given the state space of the proposed UAV our system has both controllability and observability properties. The close loop poles of the system can be transferred in certain locations using the poles placement method. In addition, it would be possible to rely on the feedback law outlined in eq. 21. Thus, a state controller can be designed which is able to meet our demands in the close system.

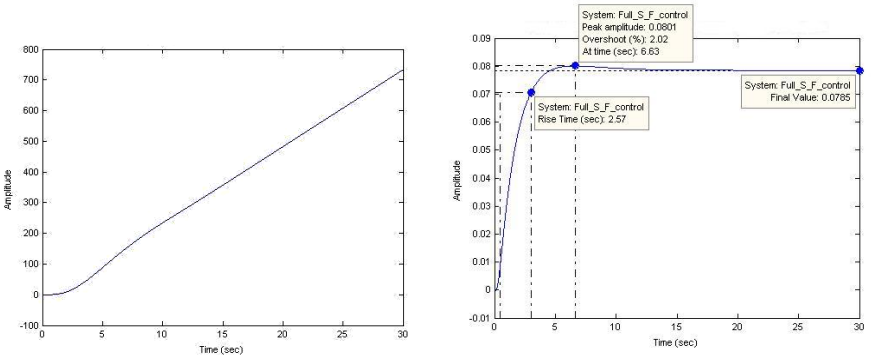
$$U = N.r - K.x \tag{21}$$

Figure 5.a indicates response step of the open loop of the system around the longitudinal axis through concentrating on matrix C.

By placing poles in p_1 =-8, $p_{2,3}$ =-0.5 ± 0.06i, p_4=-9, p_5= -10 and also by considering the steady state feedback as:

$$k = \begin{bmatrix} 7.5185 & -17.5734 & 28.6061 & 470.3481 & 12.7399 \end{bmatrix}$$

The response step of the close loop around the longitudinal axis and concentrating on Matrix C is shown in Figure 5.b.



(a) (b)

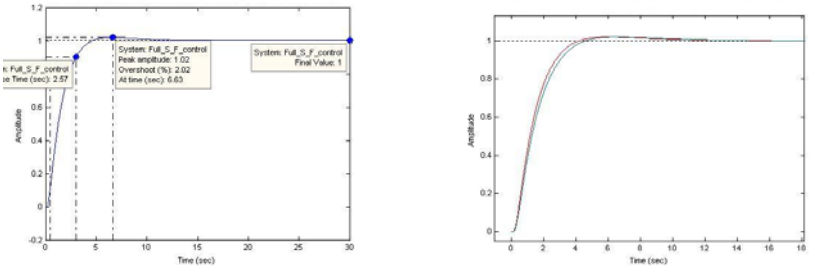
Fig. 5. (a) Step Response('C' focus – Altitude) of Longitudinal System. (b) Step response of the close loop around the longitudinal axis.

As it can be observed, there is a non-zero steady state error in the obtained response, which can be zeroed with a feed-forward gain called N (eq. 22).

$$N = -1/G_{cl}(0) = -1/\left(C\_Lon * \left(inv\left(A\_Lon - B\_Lon * K\right)\right) * B\_Lon\right) \tag{22}$$

Given state-feedback and feed-forward gain, the step response of the close loop of the system is N= 12.7399 (Figure 6.a).

Similarly, if we assume an error as large as 10% for system identification stage and specification of the matrix A from state space model of the proposed UAV, then the strength of the designed PID controller against poles placement method can be shown (Figure 6.b) [5,6].



(a) (b)

Fig. 6. (a) Close loop step response with state feedback and feed-forward gain N=12.7399. (b) Step response by the designed gain N,K and slightly different A matrix.

6 Conclusion

Considering the results of the MATLAB based design and simulation of the PID controller through both classic and poles placement methods, it was found that the designed PID controller meets all of our defaults for the important parameters of the UAV such as overshoot, setting time and rise time. It can also steer and control the proposed UAV model around the longitudinal axis in terms of stability. The results obtained from using the designed PID controller in the UAV and its predetermined defaults have been compared in table 3.

Table 3. Performance of the designed parameters in the longitudinal axis of UAV to the designed PID controller

Designed parameters on longitudinal axis of UAV	Default amount before application of the PID controller	Amount after application of PID controller	Performance
Setting time	< 5 sec.	4.98 s	Confirmed
Max overshoot	3%	1.86 %	Confirmed
Rise time	<3 sec.	0.194 s	Confirmed

References

1. Christiansen, R.S.: Design of an Autopilot for Small Unmanned Aerial Vehicles. a thesis Submitted to the faculty of Brigham Young University, pp. 9–25 (2004)
2. Hrng, L.J.: AM25 Unmanned Air Vehicle Flight Control. National University of Singapore, pp. 22–28 (2005)

3. Pisano, D.: Linear Control Design for Osprey UAV Control System Theory final project. University of Florida (2009)
4. Lewis, F.L., Stevens, B.L.: Aircraft Control and Simulation, 2nd edn., vol. 2. Wiley (2003)
5. Yiting, W.: Development and Implementation of a Control System for a Quad Rotor UAV. Master Thesis (2009)
6. The online resource for Control Systems, http://wikis.controltheorypro.com
7. Electronic publication: User Guide Aerosim, http://www.u_anamics.com

Image Retrieval Using Local Colour and Texture Features

E.R. Vimina[1] and K. Poulose Jacob[2]

[1] Department of Computer Science, Rajagiri College of Social Sciences, Kalamassery, Kochi, Kerala, India
vimina_er@yahoo.com
[2] Department of Computer Science, Cochin University of Science and Technology, Kochi, Kerala, India
kpj@cusat.ac.in

Abstract. This paper proposes a content based image retrieval system using the local colour and texture features of image sub blocks. The colour features are extracted from the histograms of the three channels of RGB colour space. Gray Level co- occurrence matrix (GLCM) is used for extracting the texture features. A combined colour and texture feature vector is computed for each sub block and Euclidean distance measure is used for computing the distance between the features of the query image and candidate images in the database. Preliminary experimental results shows that the proposed method provides better retrieving results than some of the existing methods based on colour and texture.

Keywords: Content Based Image Retrieval, GLCM, Colour histogram.

1 Introduction

As the volume of image database is growing at a phenomenal rate with the steady growth of computer power, declining cost of storage and increasing access to Internet, it has become a necessity to advance automated image learning and retrieval techniques to effectively manage the image information. In the traditional method of text-based image retrieval the image search is mostly based on textual description of the image found on the web pages containing the image and the file names of the image [6]. The problem here is that the accuracy of the search result highly depends on the textual description associated with the image. Also un-annotated image collection cannot be searched. An alternate method is to retrieve image information based on the content of the image. In content based image retrieval systems the visual contents of the image such as colour, texture, shape or any other information that can be automatically extracted from the image itself are extracted and is used as a criterion to retrieve content related images from the database. The retrieved images are then ranked according to the relevance between the query image and the candidate images in the database in proportion to a similarity measure calculated from their features [2][3].

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 767–772.
springerlink.com

2 Colour

Colour is one of the most effective, simplest and widely used low level visual features employed in CBIR. Among the proposed approaches based on colour, color histogram is extensively used [9][10]. The color histogram is obtained by counting the number of times each color occurs in the image array. Color histogram H for a given image is defined as a vector

$$H = \{H[0], H[1], H[2],\ldots,H[I],....H[N]\} \quad (1)$$

Where i represent the color in color histogram, H[i] represent the number of pixels of color i in the image and N is the number of bins used in color histogram. For comparing the histogram of different sizes, color histogram should be normalized. The normalized color histogram is given as

$$H'=H/p \quad (2)$$

Where, p is the total number of pixels in the image [7].

3 Texture

Texture can be considered as repeating patterns of local variation of pixel intensities. Unlike colour, texture occurs in a region than at a point. A number of techniques have been used for measuring the texture features such as Gabor filter [11], fractals [12], wavelets, co-occurrence matrix etc. Using these texture features like contrast, coarseness, directionality and regularity can be measured. The gray-level co-occurrence matrix (GLCM) is a statistical method of examining texture that considers the spatial relationship of pixels [8]. It is a matrix showing how often a pixel with the intensity (gray-level) value *i* occurs in a specific spatial relationship to a pixel with the value *j*. It is defined by P(i,j|d,Θ), which expresses the probability of the couple of pixels at Θ direction and d interval. Once the GLCM is created various features can be computed from it. The commonly used features are contrast, energy, correlation and homogeneity (Table 1).

Table 1. Texture features

Feature	Formula
Contrast	$\sum_{i,j} \lvert i-j \rvert^2 p(i,j)$
Correlation	$\sum_{i,j} \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i \sigma_j}$
Energy	$\sum_{i,j} p(i,j)^2$
Homogeneity	$\sum_{i,j} \frac{p(i,j)}{1+\lvert i-j \rvert}$

4 Proposed Method

In this paper a content based image retrieval system using the local color and texture features of the image sub-blocks is presented. The images are resized to 129x192 and each image is divided into 9 equal sub-blocks. For each sub-block colour features are extracted from histograms and texture features are computed from the gray-level-co-occurrence matrix.

4.1 Color Feature Extraction

Colour histogram is used for extracting the colour features. In this paper 5 bins are used for each channel in the RGB colour space. This results in a colour feature vector of size 15 for each sub block and that of size 135 for the entire image. Each element in the feature vector is normalized using equation (2), where p is the number of pixels in each sub block.

4.2 Texture Features

The gray level co-occurrence matrix is used for texture feature extraction. We have taken d=1 and Θ = 0, 45, 90 and 135. From the GLCMs, features such as contrast, correlation, energy and homogeneity are computed. 16 features are computed for each sub-block resulting in a total texture feature vector of size 144 for the entire image.

4.3 Retrieval Algorithm

A combined feature vector consisting of colour and texture is computed for each sub-block. Different steps involved in the retrieval of images are listed below.

1. Resize the image to 129x192.
2. Divide the resized image to 9 equal regions/ sub blocks
3. Compute the normalized histograms of each block using equation (2).
4. Compute the 16 texture features of each sub block from GLCM.
5. Construct the combined colour and texture feature vector for each sub block.
6. Calculate the similarity between the feature vectors of the query image and the images in the database using Euclidean distance measure. One-to-one comparison is done here. Ie., the feature vector of the sub-blocks of the query image are compared with the corresponding ones of the candidate image.
7. Arrange the retrieved results in ascending order of the distance.

5 Experimental Results

The Corel image database of 1000 images consisting of 10 categories is used for evaluating the performance of the proposed method. Each category contains 100 images. A retrieved image is considered to be correct if and only if it is in the same

category as the query. For each query, a preselected number of images are retrieved which are illustrated and listed in the ascending order of the distance between the query and the retrieved images.

The results of the proposed method of local colour histogram(LCH) + GLCM texture is compared with that of local colour histogram, GLCM texture and the method in [7] in terms of average precision. Precision (P) of N retrieved results is given by

$$P(I_k, N) = \frac{No.of\ relevent\ images\ retrieved}{Total\ No.of\ images\ retrieved(N)} \tag{3}$$

The average precision of the images belonging to the q^{th} category (Aq) is given by

$$\bar{P}_q = \sum_{k \in A_q} P(I_k) / |A_q|,\ q = 1, 2, \ldots 10. \tag{4}$$

The final average precision is

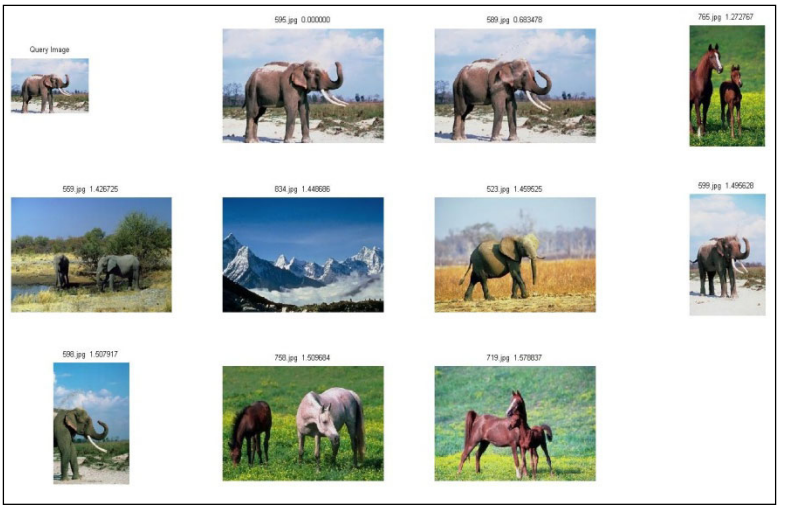
$$\bar{P} = \sum_{q=1}^{10} \bar{P}_q / 10 \tag{5}$$

Table 2. shows the average precision of the retrieved images for different methods.

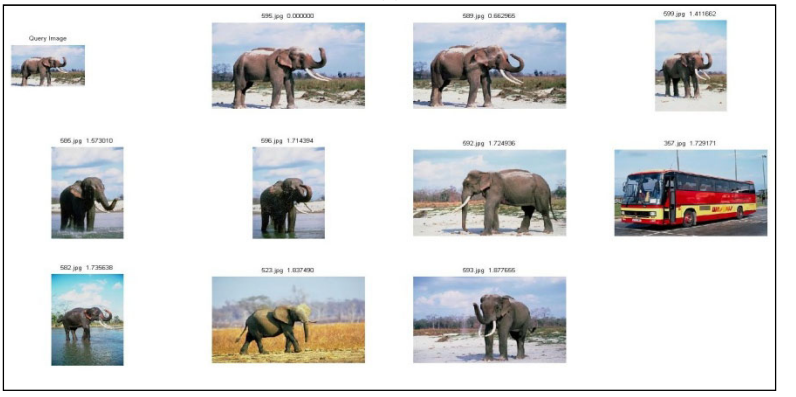
Table 2. Average Precision (N=10) of retrieved images using different methods

Category	LCH	GLCM texture	Colour histogram + Gabor transform [7]	LCH+ GLCM texture
Africa	64.3	48.2	79.50	66.9
Beaches	37.7	27.6	50.00	45.2
Buildings	38.9	28.9	46.80	31.8
Bus	55.1	59.5	51.70	60
Dinosaur	99.1	76.8	100	100
Elephant	64.1	49.7	61.70	72.9
flowers	67.6	90.6	79.50	90.7
Horse	87.6	75.9	90.80	89.5
Mountain	33	24.9	28.40	43
Food	61.6	37.2	62.80	52.7
Average	**60.99**	**54.16**	**64.76**	**65.286**

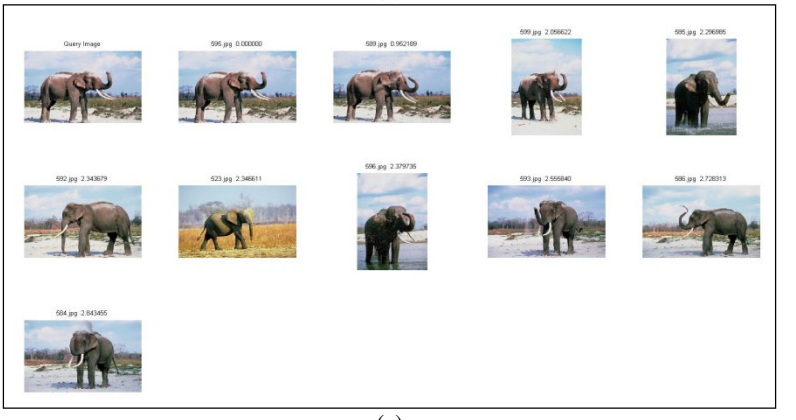
Figure1. depicts the top 10 retrieved images for a sample query image using GLCM texture(a), local colour histogram(b) and the proposed method(c). In each set, on top left corner is the query image and the retrieved images are listed according to their distance with the query image.



(a)



(b)



(c)

Fig. 1. Retrieved images: All images are ranked in the order of similarity with the query image from left to right and top to bottom (a) GLCM texture. (b) Local colour Histogram. (c) Proposed method.

6 Conclusion and Future Work

A content based image retrieval system using the local color and texture features of image sub blocks is presented here. The colour features are extracted from the histograms of the RGB image and texture features are computed from GLCM. Experimental results show that the proposed method provides better retrieving result than some of the existing methods based on color and texture. It is seen that combining the features provides better retrieval results than their individual performance. Also it is worth to note that colour is a prominent feature as colour alone could give close results to that of the combined ones. Further research can be done to improve the retrieval accuracy by using perceptually uniform colour space like HSV or by quantizing the RGB colour space.

References

1. Flickner, M., Sawhney, H., Niblack, J., Ashley, J., Huang, Q., Dom, B., et al.: Query by Image and Video Content: The QBIC System. IEEE Computer 28, 23–32 (1995)
2. Chen, Y., Wang, J.Z., Krovetz, R.: CLUE: Cluster-based retrieval of images by unsupervised learning. IEEE Transactions on Image Processing 14(8), 1187–1201 (2005)
3. Wang, J.Z., Li, J., Wiederhold, G.: SIMPLIcity: Semantics-sensitive Integrated Matching for Picture LIbraries. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(9), 947–963 (2001)
4. Smeulders, A.M.W., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of early years. IEEE Transactions On Pattern Analysis and Machine Intelligence 22(12), 1349–1380 (2000)
5. Carson, C., Thomas, M., Belongie, S., Hellerstein, J.M., Malik, J.: Blobworld: A System for Region-Based Image Indexing and Retrieval. In: Proceedings of the Third International Conference on Visual Information Systems, pp. 509–516 (1999)
6. Li, J., Wang, J.Z.: Real-time computerized annotation of pictures. In: Proceedings of the 14th Annual ACM International Conference on Multimedia, pp. 911–920 (2006)
7. Murala, S., Gonde, A.B., Maheshwari, R.P.: Color and Texture Features for Image Indexing and Retrieval. In: IEEE International Advance Computing Conference (IACC 2009), Patiala, India, pp. 1411–1416 (2009)
8. Haralick, R.M., Shanmugan, K., Dinstein, I.: Textural Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics SMC-3, 610–621 (1973)
9. Swain, M.J., Ballard, D.H.: Color indexing. International Journal of Computer Vision, 11–32 (1991)
10. Hafner, J., Sawhney, H.S., Wquitz, W.: Efficient color histogram indexing for quadratic form distance functions. IEEE Transactions Pattern Analysis and Machine Intelligence 17(7), 729–736 (1995)
11. Manjunath, B.S., Ma, Y.W.: Texture Features for Browsing and Retrieval of Image Data. IEEE Transactions on PAAMI 18(8), 837–842 (1996)
12. Kaplan, L.M.: Fast texture database retrieval using extended fractal features. Storage and Retrieval for Image and Video Databases, VI 3312, 162–173 (1998)

Implementation of DWT Domain-Video Watermarking Fast Algorithm in Blackfin DSP

Gong Zhaoqian, Gao Fei, and Sun Cheng

School of Information and Electronics, Beijing Institute of Technology, 100081, Beijing, China
gongzhaoqian007@163.com, {gaofei,chpsun}@bit.edu.cn

Abstract. Aiming at the normal DWT Domain-Video Watermarking algorithms are hard to satisfy the Real-time at present, this paper brought out a fast video watermarking algorithm based on Discrete Wavelet Transform (DWT).Using the ADSP-BF561 of Blackfin-DSP family, we initiate a video watermarking mini-system. The fast algorithm is tested on the system. Results of experiment show that the algorithm based on the Blackfin DSP's strong capability of parallel processing and DMA can satisfy the proper request of Real-time, and it also has strong robustness.

Keywords: DWT, Video Watermarking, ADSP-BF561.

1 Introduction

In recent years researches on digital watermarking technology have developed rapidly, especially in the still image watermarking technique have made a lot of results. In video watermarking technology, due to a wide range of video images, dynamic stronger, large volume of data, processing requirements of real-time higher, only the use of blind watermark detection can have practical value. Additionally the original digital video is subject to lossy compression, so it needs more challenge. These factors make the video watermarking technology lag behind the image watermarking technology. Many video watermarking algorithms are still in the stage of simulation.

In order to meet the real-time requirements of video watermarking, this paper presents a fast wavelet-based video watermarking algorithm. Using the ADSP-BF561 of Blackfin-DSP family, we initiate a video watermarking minimum processing system, and on this platform the performance of the fast algorithm is verified.

2 DWT Domain-Video Watermarking Fast Algorithm Description

Based watermarking technology and video encoding systems with different ways, the current video watermarking algorithm is mainly divided into non-compressed domain and compressed domain[1-2]. The algorithm in this paper is aimed at non-compressed domain, that directly extracted watermark information from the decoded video frames. The video watermarking algorithms on non-compressed domain can be divided into

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 773–778.
springerlink.com

spatial and transform domain. Spatial domain watermarking algorithms usually embed the watermark information into the luminance component or the color component, the advantage is low complexity, easy to compute, the disadvantage is poor visibility and robustness. Transform domain video watermarking algorithms usually embed and extract watermark information from a special transform domain. Commonly used algorithms are Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) [3]. Such algorithms can utilize the characteristic of human vision, video sequence inherent in time and space, frequency domain transform and the latest communications technology to continuously improve the non-visible and robustness of video watermarking algorithms. Due to the unique time-frequency localized features of DWT, the algorithm can achieve better results than the DCT transform in the traditional areas of the video image coding, so it became the focus of research.

Although the speed of wavelet-based video watermarking algorithm has greatly improved, for time-critical application it has considerable difficulty. For example, domestic PAL video format, its effective size is 720x576, frame rate is 25 frames per second, that requires the algorithm complete the DWT and extract the watermark information within 40ms. The traditional two-dimensional DWT firstly make row wavelet transform, and then make the column wavelet transform based on the changed row transform data [6-8]. The transformation process can be shown in Figure.1. Select a 8x8 macro block to indicate a frame video data, dashed line for row wavelet transform, solid line for column wavelet transform.

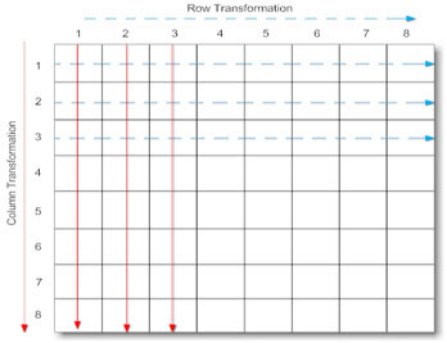


Fig. 1. Two-Dimension DWT

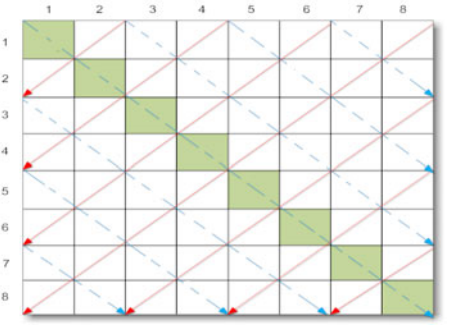


Fig. 2. Two-Dimension diagonal DWT

Based on the characteristics of traditional two-dimensional wavelet transform, combined with parallel processing and DMA capabilities of DSP, and with the practical application, this paper presents an improved diagonal wavelet transform, it can make the wavelet transform of two diagonal directions at the same time. One diagonal direction includes vertical and horizontal components. That makes the speed of video watermarking algorithms greatly improved. The transformation process can be shown in Figure.2. The dashed and solid lines represent the two directions of diagonal DWT. Due to the data in both directions is independent, the transform can process parallel to shorten the processing time.

2.1 Embedding Algorithm

① Read one frame from the video sequence, and extract the luminance component Y.
② Separate the luminance component Y into 8x8 macro blocks, select 20 macro blocks to embed the watermark information. Choice is controllable, you can choose the location according the key or provided locations. This paper selected 20 macro blocks from the middle of video frame.
③ Do the above-mentioned diagonal 5/3 wavelet transform to each 8x8 macro block.
④ From the wavelet coefficients of 8x8 macro block, select the middle 2x2 factor group [C44,C45,C54,C55], based on 20 bits watermark information to modify the relationship between groups, if the watermark is 1, then adjust the macro block coefficients to set |C44-C55|>|C45-C54|, if the watermark is 0, then adjust to set |C44-C55|<|C45-C54|.
⑤ Process two-dimensional diagonal discrete wavelet inverse transform, and re-combined into new watermarked video frame.
⑥ Repeat steps ① - ⑤ , until all the video sequences are completely embedded.

2.2 Extracting Algorithm

Extracting watermark is the inverse process of embedding watermark.

3 ADSP-BF561 Video Watermarking Detecting Minimum System

ADSP-BF561 (below referred to BF561) is an ADI and Intel's launch of the Micro Signal Architecture technology-based fixed-point DSP, integration of the traditional architecture of the advantages of DSP and RISC controller. The device uses a modified Harvard architecture and multi-stage pipeline structure. Its core has two independent and symmetric 600 MHz high-performance Blackfin processors.

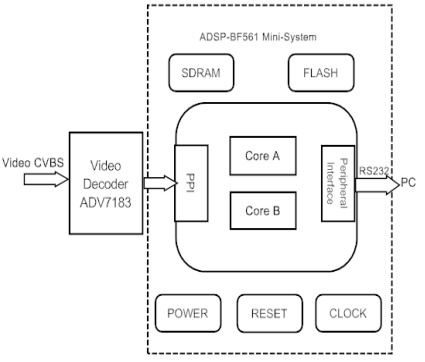


Fig. 3. ADSP-BF561 Mini-System

In order to verify the algorithm, this paper designed a video watermarking detection mini-system based on BF561.The basic structure is shown in Figure.3.

The entire system takes the BF561 as the core, combines with the front-end-video decoder module and the back-end-serial communication module. The BF561 is made up of the power circuit, reset circuit, the clock and the external SDRAM and FLASH. [4-5] The system workflow is simple, the video stream that has been embedded watermark is sent to the video decoder, and then is imported to the core by the PPI interface to process, the processed-result is sent to the host PC via the serial port for data storage. Using the characteristic of DSP's parallel processing, the algorithm can do the wavelet transform in both directions simultaneously, so the processing time is reduced greatly. Based on the powerful Blackfin DSP's DMA, the algorithm can easily do the one-dimensional and two- dimensional data movement, as each of diagonal-lines' data interval is constant. It is shown in Fig.2, identified by the shadow, Spacing between the adjacent rows is 8 widths. The DMA can easily set an appropriate step to complete the transformation of data.

The process of algorithm on the ADSP-BF561 minimum system is as follow: select a raw YUV video sequence, the resolution size is 352x288, select the watermark information for 20-bit random (1, 0) sequence, firstly loading the video into the DSP core, embedding the watermark according the fast embedding algorithm, secondly encoding the video into CVBS flow, then through a closed loop, the CVBS flow input the decoder module, lastly completing the watermark extraction by the fast extracting algorithm.

4 Results and Discussion

4.1 Test of Invisibility

Completing the watermark embedded on ADSP-BF561 minimum system, the none-watermarking and watermarking images from the video stream are shown in Figure.4 and Figure.5 .We can see from the figures that the video quality nearly has not been affected. The Figure.6 shows the PSNR value of 50 video frames. Normally if the PSNR value is greater than 30dB, it states that the changed image is very close the original image. In Figure.6 the PSNR value is more than 40dB averagely, proving that the algorithm is able to meet the application requirements.

Fig. 4. Original Image

Fig. 5. Watermarked Image

4.2 Test of Robustness

The digital video is subjected to lossy compression, so the important robustness of digital video watermarking technology is anti-video encoding and decoding compressed. The most commonly used video codec compression standard is MPEG-2. The watermarked video in the experiment did the MPEG-2 compression and decompression processing. Two of YUV video sequences were tested in the experiment, they were coastguard_cif.yuv and container_cif.yuv. The experiment's flow is as follow: embedding the watermark according the fast algorithm, doing the MPEG-2 compression and decompression processing, finally extracting watermark from 50 frames and 100 frames. The algorithm has taken some statistical methods to filter the experiment results. Because the watermark information is artificially controlled, we can choose the watermark that contains specific meaning or the random binary watermark. In this paper we selected (0, 1) random sequence to ensure the random 0 and 1 take the equal probability. Actually there are often some poor quality of video frames in a video stream, that leads the extracted watermark information have a large different probability between 0 and 1. Based on the statistical theory, if the probability difference is beyond a threshold, the algorithm will remove the watermark information to ensure the extraction accuracy. The final results are shown in Table.1.

Comparing with the none-codec processing methods, the accuracy rate nearly remains the same. That proves the algorithm can effectively resist the MPEG-2 compression.

Table 1. Comparison of Accuracy

Video Name	Frames	None MPEG-2 Accuracy	MPEG-2 Accuracy
coastguard_cif.yuv	50	90%	85%
container_cif.yuv	50	92%	86%
coastguard_cif.yuv	100	92.5%	87.5%
container_cif.yuv	100	93%	88%

4.3 Test of Real-Time

This paper mainly tested the extraction algorithm whether it can meet the requirements of fast processing. We designed the algorithm on the Integrated Development Environment-Blackfin Visual DSP++ and transplanted the algorithm on the BF561 to run, through the Integrated Development Environment's CYCLE counter to count the cycles. The experiment results shown that extracting a resolution of 352x288 video frame cost 43075 cycles, according to BF561's kernel 600NHZ clock basis, it spent 72us,fully meeting the real-time processing requirements. Also we made a comparison test between a diagonal direction wavelet transform and a traditional two-dimensional wavelet transform. We took a 8x8 macro block to do two kinds of wavelet transforms and computed the cycles.

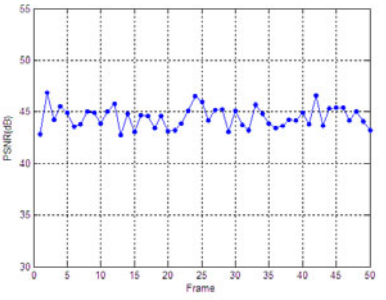


Fig. 6. Image PSNR

The diagonal direction wavelet transform cost 6783 cycles, while the traditional wavelet transform cost 11870 cycles, beyond about 57% than the diagonal wavelet transform. So it states that using the parallel processing and DMA of BF561 the algorithm meets the real-time requirements.

5 Conclusion

In order to solve the real-time requirement of video watermarking algorithm, this paper presents a fast algorithm in DWT domain. The algorithm does the wavelet transform in both diagonal directions simultaneously, improving the transform speed. Experiment results show that the algorithm can meet the real-time requirement and make a foundation for further application.

References

1. Hanjalic, A., Langelar, G.C., van Roosmalen, P.M.B., Biemond, J., Lagendijk, R.L.: Image and Video Databases: Restoration, Watermarking and Retrieval. In: Advances in Image Communications, vol. 8, Elsevier Science, New York (2000)
2. Podilchuk, C.I., Delp, E.J.: Digital Watermarking: Algorithms and Applications. IEEE Signal Processing Magazine 1, 33–46 (2001)
3. He, Q., Su, G.-C.: Watermarking algorithm based on the triple set of wavelet coefficients. Computer Applications and Software (11), 34–35 (2004)
4. Chen, F.: Blackfin family of DSP theory and system design. Publishing House of Electronic Industry, Beijing (2004)
5. Li, G., Gao, F.: Implementation of real-number parallel IFFT algorithm in Blackfin DSP. Application of Electronic Technique, Beijing (2009)
6. Daubechies, I.: Ten Lectures on Wavelets. Philadelphital City Press, Philadelphital (1992)
7. Mallat, S.: A wavelet tour of signal processing. China Machine Press, Beijing (2003)
8. Li, B.-C., Luo, J.-S.: Wavelet analysis and its application. Publishing House of Electronics Industry, Beijing (2003)

Verification of Behavioral Domain-Specific Languages with a Model Checker

Christian Ammann

University of Applied Sciences Osnabrück
Postal Box 1940, 49009 Osnabrück, Germany
c.ammann@hs-osnabrueck.de

Abstract. Model-Driven Development proposes to create models and automatically transform them into executable software. Textual models can be implemented with custom domain-specific languages (DSLs) and verified with a model checker. This paper presents a framework which can be used to develop behavioral DSLs and automatically transform them into a model checker input language. Therefore, a developer with just minimal knowledge about model checking can focus on the development of the DSL while he gets verified applications.

1 Introduction

Model-Driven Development (MDD) [11] is a software engineering approach which proposes to create models and transformation templates for parts of a system. The model transformation templates perform an automated translation into executable software which can increase a products quality and speed up the development process. This work focuses on textual models which can be implemented with domain-specific languages (in contrast to graphical models which can be implemented with the *Unified Modeling Language* and corresponding tools).

Modern frameworks for the Eclipse IDE like Xtext and Xpand [1] allow the creation of custom domain-specific languages, their integration in software projects and transformation into a target language. Xtext acts as a parser generator but also provides additional features like the automated generation of an editor with syntax highlighting and validation for the implemented DSL. Xpand is a template language which can be used to parse the abstract syntax tree of Xtext DSLs and to transform them into source code.

The automated generation of software increases the software quality but the models itself can still contain errors. Before the transformation process is invoked, a model checker can be used to verify whether a model with behavioral aspects (implemented with a custom DSL) meets all requirements. This work uses the model checker Spin [6] (and its input language Promela) for model verification.

The usage of a model checker is challenging: A developer with in-depth knowledge about formal verification has to transform a model in the corresponding model checker input language to reduce the models state space size (models with a large state space cannot be verified anymore because the model checker exhausts the available memory). Especially in projects which use several custom DSLs, formal

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 779–782.
springerlink.com

verification becomes difficult because an additional transformation template into the model checker input language has to be created for each DSL.

This paper presents a DSL verification framework which can be easily integrated into a given Xtext/Xpand project to allow the formal verification of behavioral DSLs with the model checker Spin. The advantage is: A developer with just minimal knowledge about model checking can focus on the development of his DSL and the corresponding models while he gets verified applications. The framework consists of a set of transformation templates and a set of production rules to implement expressions, variables, statements, etc. in a domain-specific language.

We implemented two DSLs in two case studies which use the verification framework to achieve an additional transformation to Promela. The first case study is an extension for the movement tracking system *AssyControl* which monitors workers at a workbench in the industrial assembly process and was developed by the *Soft2Tec GmbH* for the *Otto Kind AG*. The corresponding DSL is used to implement parts of AssyControl with finite state machines. The second case study is a DSL for the implementation of web applications with the *Google Web Toolkit* (GWT) [2].

This work is part of the *KoverJa Project* [3] and is supported by the German Federal Ministry of Education and Research (BMBF). KoverJa aims for the increase of software quality of (distributed) Java applications which includes methods as testing, model-driven development and formal verification.

2 Verification of DSLs

The development of DSLs with Xtext and Xpand consists of two steps: A developer writes with Xtext a grammar which contains of a set of production rules. Xtext automatically generates a parser and an editor for the DSL with syntax highlighting and validation. Afterwards, Xpand is used for the development of transformation templates which translate the previously described DSL into the target language.

The DSL verification framework consists of a set of Xtext production rules and Xpand transformation templates to Promela. When integrated in an Xtext/Xpand project, the production rules extend the corresponding DSL with features like statements, expressions, variable declarations and the specification of requirements. It is also possible to declare classes, to instantiate them and to call their methods. The models state space can be reduced with symmetric arrays and property preserving abstraction (both framework features are described in more detail in [5]). The framework transformation templates perform an automated translation of the previously described features into Promela which allows a formal verification of the DSL with the model checker Spin. The workflow of the framework is:

1. A developer writes manually the production rules which represent structural aspects of his DSL.
2. The production rules of the framework have to be integrated in the grammar and are used to express behavior (e.g. statements and expressions).
3. The transformation templates of the framework have to be injected in the DSL project to perform an automated transformation into Promela.

4. The developer writes manually transformation templates for the structural aspects of the DSL.

We demonstrate this approach with our first case study: AssyControl is a system which monitors the manual assembly process at workbench. Workers wear special gloves which contain an ultrasound transmitter. AssyControl receives the signal and calculates the position of the workers hand using the signal time delay. During the assembly process, a worker grasps components, moves and assembles them. AssyControl is programmed by a supervisor with a recipe which contains the components assembly order (e.g. grasping object 1 from box 1, moving it to the work place, grasping object 2 from box 2, moving it to workplace, assemble object 1 and 2). During the assembly process, AssyControl keeps track of the workers movements, informs him when an error occurred (e.g. grasping the wrong components in the wrong order) and proposes a correction if possible.

AssyControl is a reactive system and parts of it can be modeled with finite state machines. Therefore, we implemented a DSL for statecharts with the usage of the verification framework. The DSL provides features like simple states, composite states, transitions, entry actions, exit actions and signal events. The following listing shows a part of the DSL's grammar and the integration of the verification framework:

```
EntryAction:
   "entry" "{"
      (statements+=Statement)+
   "}";
```

It contains the implementation of entry actions with Xtext. An entry action has the format *entry{ statement$_1$, statement$_2$, ...}* and contains a list of action language statements. The production rule *Statement* is part of the verification framework. The complete statechart DSL with the verification framework and examples for Xtext}/Xpand 1.0 can be found at [4].

The DSL in the second case study can be used to generate web applications with the GWT. Static elements are implemented with HTML while dynamic elements are developed with Java. The client side of an application (which is executed in a browser) is automatically transformed into JavaScript. A GWT web application can also use *remote procedure calls* (RPCs) for the communication with a server. Pure HTTP is stateless and therefore GWT supports session-ids which are implemented with session objects. The GWT DSL uses the verification framework which allows a transformation into Promela and a verification of requirements (e.g. whether it is possible for a user to access certain elements of a web application without a prior login).

3 Related and Further Work

We have not found an approach in the literature which which tries to provide a generic framework to verify custom, behavioral DSLs. Nevertheless, there exist many

papers which deal with the formal verification of statecharts and model-driven development.

[8] and [9] describe an approach for the verification of UML statecharts. Jones et al. [7] describe a combination of MDD, model checking and a case study about mobile health systems. Those systems are small units for measuring human body data like temperature, blood pressure, etc. Further related work is focused on specific modeling languages (e.g. [10]).

Our second case study about GWT is not yet finished completely. Therefore, we plan as further work to finish and release it as a second example for the usage of the verification framework. Furthermore, we plan to extend the verification framework with a generic support for signal passing because signals are necessary in both case studies.

References

1. Eclipse Xtext - Language Development Framework, `http://www.eclipse.org/Xtext` (accessed January 31, 2011)
2. Google Web Toolkit, `http://code.google.com/webtoolkit` (accessed August 18, 2011)
3. KoverJa Projekt - Korrekte verteilte Java Applikationen, `http://www.edvsz.hsosnabrueck.de/kleuker/CSI/KoverJa` (accessed September 14, 2010) (in German)
4. Statechart DSL and Verification Framework, `http://home.edvsz.hsosnabrueck.de/chammann/sc` (accessed September 2, 2011)
5. Ammann, C., Kleuker, S., Pulvermüller, E.: From Business Modeling to Verified Applications (to be published), Informatik 2011. In: Protocol based Modelling of Business Interactions Workshop (2011)
6. Holzmann, G.: Spin model checker, the: primer and reference manual. Addison-Wesley Professional (2003)
7. Jones, V.: Model driven development of m-health systems (with a touch of formality). In: Fourth Annual IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2006, pp. 580–584 (March 2006)
8. Merz, S.: Model checking and code generation for UML state machines and collaborations. In: Schellhorn, G., Reif, W. (eds.) 5th Workshop on Tools for System Design and Verification (FM-TOOLS), pp. 59–64 (2002)
9. Paltor, I.P.: The Semantics of UML State Machines. Tech. rep. (1999)
10. Pulvermüller, E., Feja, S., Speck, A.: Developer-friendly Verification of Process-based Systems. Journal on Knowledge-Based Systems (KNOSYS) 23, 667–676 (2010)
11. Schmidt, D.C.: Guest editor's introduction: Model-Driven Engineering. Computer 39, 25–31 (2006)

Systematic Analysis of Tolerance Leadership Based on Core Capability of Corporation

Gao Hong

Business School, University of Shanghai For Science and Technology, Shanghai, 200093, China

Abstract. The concept and connotation of the tolerance leadership based on core capability of corporation are defined. The based on systematic elements of the tolerance leadership are analyzed and determined. By using the theory and method of systematic engineering, the systematic hierarchical structure model for the tolerance leadership based on core capability of corporation is formulated. After an analysis of the systematic hierarchical structure model proposed, suggestions for carrying out the development of the tolerance leadership based on core capability of corporation are given.

Keywords: core capability of corporation, tolerance leadership, knowledge, learning.

1 Introduction

With the promotion and popularization of information intelligent, Interdisciplinary cross of technology innovation, the corresponding focus of market competition has been transferred from cost to performance, quality of products and even to the development speed of new products; the further intensify of economic globalization, High mobility and openness, permeability and complementary, intensive and monopolistic coexist; Information, knowledge capital, replacing physical material resources, have become an important factor of economic development. The growth of value in enterprises mainly rely on the increase of knowledge, and due to the intensify of environmental pollution, a new scientific outlook has gone deeply into the public's minds, it has become a consensus and a new trend to protect natural resources, to reduce the environmental pollution.

Obviously, the business environment, which the traditional enterprises rely on, has changed significantly in the aspects of concepts, technology and market etc. Facing such earth-shaking changes, how to create and maintain the core capability and advantages[1] in such an unrest change environment, mainly depend on the leadership of a organization and unique high value core capability of enterprises. Enterprise's survival and development depends more on leadership of each lever than ever. The process of leadership is a kind of complicated behavior.

2 The Definition of Enterprise Core Capability

The basic of enterprise's survival and sustainable development depend on its core capability, especially the enormous pressure of competition that Chinese enterprises

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 783–791.
springerlink.com

are facing, in a deeper way, this is actually the competition of core capability between enterprises.Thus the premise of analyzing and establishing the competitive advantage is to define enterprise core capability scientifically and reasonable. Since C.K Prahalad and G. Hamel published the article “enterprise core capability” on 《Harford Business Review》 in the 1990s, core capability theory has become hotspot issue of enterprise development and enterprise strategic management research, and the hotspot in the practice of Industry and commerce administration.The definition of enterprise core capability varies from experts to experts, and two of them are influential. They are: (1) Prahalad and Hamel[2] think that: enterprise core capability is the accumulative knowledge in the organization, especially knowledge on how to coordinate different producing skills and organically combine various kinds of skills; （2） McKinsey consulting company’s point is: enterprise core capability is the combination of a series complementary skills and knowledge inside enterprises, it has the ability to make a business or more business on a world-class lever[3].Besides, overseas scholars produce a series of ideas[3] after studies on enterprise core capability, including core capability integrated view; skills network view, coordination view, combination view, knowledge carrier view, component architecture view, platform view and skills view.Chinese scholars have also done extensive research on core capability, the definition on enterprise core capability competitiveness on literature 4 was divided into three fields: narrow, medium and wide. The point of narrow thinks that core competitiveness is a combination of a series of complementary skills and knowledge inside enterprises, mainly compose by the ability of insight and prediction and the ability of execution at the site; the medium point view hold the idea that core competitiveness is key skills, invisible knowledge and intelligence capital that enterprises have; the wide point view thinks that core competitiveness is a composite system constituted by many kinds of elements.

Through learning and summarizing, this article believes that enterprise core system is a dynamic organic intelligentized system with the ability of integrating, coordinating and organizing self-learning, creating high value, which exist in people, organization, environment, assets and equipment, carrier and so on in the form of enterprise culture, values, rules, skills, knowledge, expertise, know-how, and resources etc. and through internal control of the organization and management mechanism to conduct collaborative integration, performing as excellent organization knowledge.It can not only reflect the real comparative advantage under the conditions of enterprises currently, but also can present the potential capability that enterprises can maintain long-term and strengthen the competitive advantage. The specific connotations are as follows: (1) make available and effective enterprises resources (tangible and intangible and strategic resources) as the foundation; (2) Build learning, innovation and incentive mechanism, through self-learning of organization, transfer enterprise resources into different kinds of skills, information and knowledge (basic ability) (3) Take core values and common vision as a scaling device, through internal control of organization and management system, integrate basic ability into special key knowledge and core elements; (4) Through organic combination, establishing an intelligence operation system of enterprises basic on information technology, under this system , key knowledge and core elements are turned into special core intelligence capital system of enterprise organization, and form performance output.

(5) Through strategic decision and management system, make organization check itself sustainable, calibrate internal relationships between enterprise basic abilities, core competence, operating environment and enterprise tasks, and make actions to improve continuous and loop dynamically.

3 The Definition of Tolerance Leadership Based on Core Capability of Corporation and the Option of Systematic Elements

In a relative long period, the research on leadership is much more than any themes that are relative to enterprises or other organizations.Such as Doc Hessian, the founder of American leadership development center, he defines leadership as: leadership is a process that produces influence on others.He pointed that: a leader persuade followers to struggle for some targets, while these targets reflect the leader and the followers' common values and motivation, wishes and needs, revenge and ideal.The famous American leadership experts: John Maxwell describe leadership in his book <leaders' 21 qualities> as: a position cannot make one exert leadership, instead a leader can make a position play a role; a leader find a target first, then he or she will find a group of followers, while the common person always find a leader first, then he or she starts to identity the leader's target, only the leaders who can trigger other people's dynamic can create kinetic energy[5].The model of leader competence that Hollenbeck[6] has created is: personal qualities, building ability of internal system, cognitive ability, collaboration and cooperation ability, risk defense capabilities, organization control ability. James Kouzes and Barry Posner[7] from amrican believe that: leadership is a kind of art which mobilizes people to struggle for their common visons.Leadship is everyone's business. Credit and sincerity is the cornerstone of leadership, they also have put forward five habitual behavior model of excellence leadership: play an exemplary role, shared vision, challenge the status quo, and make everyone do, inspiring. The group of studies on technology leadership from Chinese Academy of Sciences[8] has also created a five competitive forces model constructed by charisma, vision, influence, determination and control.

From the above discussion and description, we can see that, leadership is influence, is a process that influences other people. As long as you successfully influence others' behavior and achieve the organization's goals, then you are using leadership. These description have also hinted us that no one was born with leadership, most people have the possibilities and the potential to become an excellent leader. That means leadership can be educated. You can get it through subjective efforts and proper path in a specific environment and organization.According to the above exploration and combine the author's many years of practice in the position of executives, in this article we think the tolerance leadership based on core capability of enterprises is a dynamic and harmonious leader system, is a component of core capability of enterprises, is a subsystem inferior to the system of core capability of enterprises. To foster, enhance and maintain enterprise core competence and make the organization produces optimal performance output, which is the general goal. The process of leadership happens in a certain organization, during this process, the equal opportunities and fairness of distribution and the share of developmental fruits are

emphasized; Leadership exists in each lever of the organization in the form of people, leadership is everyone's business, and it relates closely to the leadership process and the pursuit of organization goals, the final achievement of goals is the result of joint actions from the leader and the followers, which is under influence of leadership.The specific descriptions are as follows:(1) First, tolerance leadership based on core capability of corporation is a kind of smart knowledge, most exist in each level of an organization in the form of invisible knowledge, manifests as a kind of non-positional influence, its result is the achievement of goals.The general ability that a individual has is basic intermediate ability, it can be also called the foundation of capability. Refine general ability to smart knowledge is how leadership is formed. General ability is the ability to do a specific job with the structured using of the basic quality.Basic quality and literacy are the fundamental origins, are the resourceful basic of the formation of leadership, they continuously complement the nutrition for the transformation and ascension that a leader need personally. General ability is transformed from individual basic qualities.In the process of leadership formation and the process of leadership, the equal opportunities that each member shared, fairness of distribution and share of developmental fruits are always emphasized and insisted, besides, a kind of harmonious leader style, atmosphere and environment are advocated. Learning, innovation and summary are a converter and a development way to the operating of the system of tolerance leadership, which is based on core capability of corporation. From the basic quality and literacy to general ability, from general ability to smart knowledge, from smart knowledge to leadership, they are all summarized through learning, the forms of learning mainly are: conventional study, action learning and strategic study.The assurance of system elements is the basic of system analysis, according to the above analysis, on the basic of reference to the relevant documents [2]~[10], the expertise on consulting business and the author's many years of practice in the position of executives, combining modern enterprise management theory, apply system scientific methods, this article choose the following to be the system elements of tolerance leadership.

B_1 ~tolerance leadership, which is based on core capability of corporation; B_2~influence charisma; B_3~ integrated control; B_4~ strategic insight; B_5~ resolute determination; B_6 ~executive willpower; B_7 ~core values; B_8 ~skills and capability; B_9 ~management ability; B_{10} ~Culture and knowledge; B_{11}~ professional knowledge ; B_{12} ~social image; B_{13} ~mental models; B_{14}~ innovative thinking; B_{15}~ learning awareness; B_{16} ~personality traits; B_{17}~ Character charm ; B_{18}~ Physique; B_{19}~ passion; B_{20}~ tolerance; B_{21} ~credibility.

4 Construction of Structure Model of Tolerance Leadership Based on Core Capability of Corporation

4.1 System Contracture's Adjacency Matrix

From the above analysis, we can see that tolerance leadership system is composed by 21 system elements.

$$B=\{ b_1,b_2,\ldots\ldots,b_n \}, \quad n=21 \tag{1}$$

Remember to R_{ij} (hereinafter referred to as R) with the dual relation between b_i and b_j.

If (1) B_i and b_j some binary relation—R, that is b_i R b_j; (2) B_i and b_j don't have some binary relation—R, that is $b_i \bar{R} b_j$. Set the adjacency matrix (A) to demonstrate system elements of basic binary relation between elements of system (directly relations), that is $A=\{a_{ij}\}_{nxn}$.

$$\text{Then } a_{ij}=\begin{cases} 1 & Rb_j \\ 0, & b_i\overline{Rb_j} \end{cases} \tag{2}$$

Besides, as for any constituent elements of system for b_i and b_j, if b_i R b_j and b_j R b_i, then we call this relation strong connection relation.

To do some preliminary analysis research on the interaction and influence between above system elements of tolerance leadership, and according to the formula (2), set up adjacency matrix of tolerance leadership system, $A=(a_{ij})_{21x21}$.

At adjacency matrix $(a_{ij})_{21x21}$, all line elements of $\{b_1\}$ are zero, whole column elements of $\{b_{18}, b_{19}, b_{20}, b_{21}\}$ are zero, so, $\{b_1\}$ is a output element (which is also called terminate set elements) of tolerance system, $\{b_{18}, b_{19}, b_{20}, b_{21}\}$ are input elements(which is also called starting set elements). Due to b_{16} R b_{17} and b_{17} R b_{16}, b_{16} and b_{17} are strong connections.

4.2 The Derivation of System Structure's Reachability Matrix

Set matrix $M=(m_{ij})_{21X21}$,m_{ij} as system reachability matrix, and the most road long and the transfer number r under the condition without loop, and 0«t«r ,then the reachability matrix[11] is defined as.

$$m_{ij}=\begin{cases} 1, e_i R^t e_j \\ 0, e_i \overline{R^t} e_j \end{cases} \tag{3}$$

Matrix A and M elements are "1" or "0",is 21 * 21 order 0-1 matrix, and accord with Boolean algebra operations rules. Through the operation of the adjacency matrix A, according to we can draw out reachability matrix M of tolerance leadership system based on the follow formulas (4) and (5).

$$M=(A+I)^r \tag{4}$$

$$(A+I) \ddagger (A+I)^2 \ddagger (A+I)^3 \ddagger \ldots\ldots \ddagger (A+I)^{r-1} \ddagger (A+I)^r = (A+I)^{r+1}=\ldots\ldots=(A+I)^{21} \tag{5}$$

4.3 The Establishment of Tolerance Leadership System Structure

System elements set of termination of this system integrating the essential factor is the highest elements, the highest element of this tolerance system is $\{b_1\}$ .According to Idea of grading the elements, remove b_{17} from b_{16} and b_{17} who has strong connections, thus we can get reachability matrix M' as you see. In M' , the"1"elements are arranged from less to more. Adjust M's rows and columns, and infuse a box. Each box represents a level.

So the tolerance leadership system structure can be divided into 5 levels, If the L_r represents the r (r = 1,2,3,4,5) level, then results are as follows: $L_1=\{b_1\}$, $L_2=\{b_2, b_3, b_4, b_5, b_6, b_7\}$, $L_3=\{b_8, b_9, b_{13}, b_{14}, b_{15}\}$, $L_4=\{b_{10}, b_{11}, b_{12}, b_{16}\}$, $L_5=\{b_{18}, b_{19}, b_{20}, b_{21}\}$. Take the principle of not considering the effect of passing a level , based on the above analysis structure and will delete the strong connection element b_{17}, we can get the explaining structure model of tolerance leadership, as figure 1 shows.

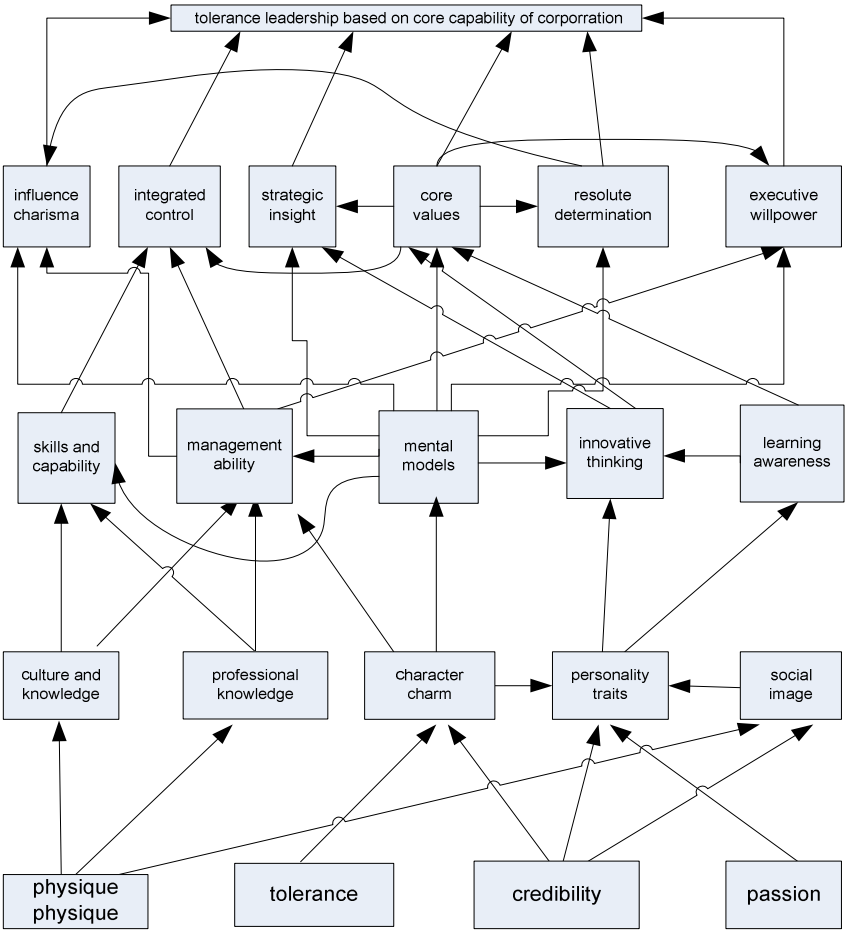


Fig. 1. the explaining structure model of tolerance leadership

5 The Systematic Analysis of Tolerance Leadership Based on Core Capability of Corporation

We can see that from figure 1, the tolerance leadership system based on core capability this article explores is a 5 levels hierarchical structure system.

The first is the final output level of the system, is the general destination that the operation of the system wants. To strategically promote influence charisma, integration of control, strategic insight, resolute decision-making, implementation willpower and the core values.

The second level is core element level, which is can also be called smart learning level, the tolerance leadership is a dynamic system whose core elements combine organically, effect mutually. The formations of the core elements are finished through action learning process, on this foundation, integrate every core element into smart knowledge through strategic learning. Among elements in each level:

1) Influence charisma={ skills and capability(Technology, information, norms, risk, conflict, and position)+management capability(Relations, interests, communication, power, experience)+mental models(tolerance, confidence, focus, scientific and reasonable)}* core values
2) Integrated control={ management capability(relations, interests, passion, communication, power, experience, and confidence)+ skills and capability(technology, information, norms, risk, conflict, and position•} * core values
3) strategic insight={ mental models(tolerance, confidence, focus, scientific and reasonable)+ innovative thinking(curious, environment, business, development and mode)+learning awareness}* core values
4) Resolute determination={ mental models(tolerance, confidence, focus, scientific and reasonable)}* core values
5) Executive willpower={management ability (relations, interests, communication, power, experience)+ mental models (inclusive, confidence, focus, scientific and reasonable)} * core values

It's visible that in this level the formation of core values and every element are all product relations, we all it times application device, a key element of this level.

The third level is general ability level, which mainly includes the leaders' function ability, the foundation of the formation of more superior ability. General ability are acquired by practicing study, while the formation of elements in general ability are acquired by conventional leading.

Among them:

1) Skills and capability= (cultural knowledge + professional knowledge)+ Mental models (inclusive, confidence, reputation, focus, science);
2) Management ability = {cultural knowledge + professional knowledge + personality quality/character charm (sincerity, credit, passion, integrity and fairness, tolerance)} * mental models (inclusive, confidence, focus, scientific and reasonable);
3) Innovative thinking = personality quality/character charm (sincerity, credit, passion, integrity and fairness, tolerance)} * mental models (inclusive, confidence, focus, scientific and reasonable) * learning awareness

It's visible that in this level the formation of mental models and every element are all product relations, we all it times application device, a key element of this level.

The fourth level is basic element level, which is the foundation for formulation of accumulative knowledge, among them:

(1)Cultural knowledge and professional knowledge is information knowledge, most exists visibly , cultural knowledge mostly belong to the public knowledge, professional knowledge can at least be divided into two levels, the first level is professional scientific knowledge, the second level is the special knowledge of the organization where the leaders work in.Conventional learning gets cultural knowledge and professional knowledge, but the ways are different, cultural knowledge and the special scientific knowledge most are from the public way outside the organization, is a kind of passive learning in a deep way.While the special professional knowledge in the second level are mainly from a special inner way of the organization, basically a kind of initiative study. The IQ of leaders plays a leading role in the learning process.Individual personality and character charm exist invisibly, a small part of it is born, but most are formed subtlety and unknowingly by cultivation, and understand, intuition, sensitive, psychological suggestion, self-awareness, and self-motivated, in this process, EQ of the leader plays a leading role.

The fifth level is the basic quality, which is also called the basic elements level and the smallest elements level; comprehensively support the leader's various ability, training and learning. So, a healthy physique, and an active, dynamic and ambitious passion, scientific and reasonable tolerance, and a justice and sincere good reputation, these are the foundation of cultivating, promoting and maintain the tolerance leadership of core capability of enterprises.

6 Epilogues

From the structure levels of system explanation model, transfer from the basic elements in the low level to general ability in a higher level through conventional study. Then integrate the ability into smart knowledge represented by core elements through action learning. At last ,the knowledge will grow into the tolerance leadership through strategic studying.During this process, credit is the cornerstone of tolerance leadership, core values is the internal motivation for the operation of the system, the most key element of the system, the improvement of the culture of enterprises and the inner environment of the organization are the lubricant. Learning mechanism is the basic way for the operation of the system. The leader who has the non-positional power and influence is an expert or a thinker with noble character.

Generally, tolerance leadership based on core capability of enterprises is intellectual, global, predictable and unique, inclusive, difficult to model and of high value is inner wisdom and knowledge that the leader has, most exist in the form of invisible knowledge, original from scientific core value system, healthy mind pattern, the rich knowledge of science and technology, proprietary data and information, the leader's accumulation of decision making experience , innovative ability and excellent analysis ;In short, tolerance leadership based on core capability of enterprises can make enterprises find a pattern can create a first class outcome and competitive advantage.

References

1. Liu, F., Jian, Z.: Sustainable competitive advantage: based on the perspective of dynamic ability. Sicentific Management Research 28(3), 51–55 (2010)
2. Parehalad, C.K., Gary, H.: The core competence of the corporation. Harvard Business Review 68(5/6), 79–91 (1990)
3. Wu, J.: The research on the mechnism of formation and application of core capability of enterprises. Southeast University Press, Nanjing (2006)
4. Rong, H.: The strategic management based on core competence. Shanxi Economy Press, Taiyuan (2007)
5. Huang, X.: Leadership model and the delevopment of leadership. Beijing University of Posts and Telecommunications Press, Beijing (2008)
6. Hollenbeck, G.P., McCall Jr., M.W., Silzer, R.F.: Ledership compantency modes. The Leadership Quarterly 46(3), 466–480 (2008)
7. James, K., Barry, P., (translated by Li, L., Zhang, Z., Yang, Z.): Leadsership. Publishing House of Electronics Industry, Beijing (2009)
8. The group of studies on technology leadership from Chinese Academy of Sciences. The research on 5 leaderships. The Sicentific of Leadership (9), 20–23 (2006)
9. Robinson, D.A., Harvey, M.: Global Leadership in a Culturally Diverse World. Management Decision 17(4), 398–413 (2006)
10. Collins, R.: Knowledge Leadership: The art and science of the knowledge-based organization. The Learning Organization 12(3), 299–315 (2005)
11. Wang, Y.: Systems Engineering. China Machine Press, Beijing (2010)

A Multichannel QoS-MAC Protocol for Two-Layered WMSNs

Shi Bo[1] and Chen Kesen[2]

[1] Department of Computer Science, JiNing University,
QuFu XingTan Road No. 1, ShanDong Province, P.C. China, 273155
mengboshi@139.com
[2] Department of Environmental Engineering, Shandong Water Polytechnic
Rizhao Xueyuan Road No. 677, ShanDong Provice, P.C. China
cks111@126.com

Abstract. Rapid penetration of small customized wireless devices and enormous growth of wireless communication technologies have already set the stage for large-scale deployment of wireless sensor networks. Offering precise quality of service (QoS) for multimedia transmission over sensor networks has not received significant attention. However offering some better QoS for wireless multimedia over sensor networks raises significant challenges. In this paper, we propose an adaptive Cross-Layer multi-channel QoS-MAC protocol to support energy-efficient, high throughput, and reliable data transmission in Wireless Multimedia Sensor Network (WMSNs). Our proposed protocol uses benefit of TDMA and CSMA/CA to adaptively assign channels and timeslots to active multimedia sensor nodes in clusters. Simulations show that the proposed system achieves the performance objectives of WMSNs with increased network throughput at the cost of a small control and energy overhead.

Keywords: Wireless Multimedia Sensor Networks, MAC, Multichannel, Adaptive Cross-Layer, Cluster-QoS.

1 Introduction

The main component of wireless multimedia sensor network, are the sensor nodes, which are small in size, capable of self-organizing, sensing, processing data and communicating with other nodes. The availability of inexpensive hardware such as CMOS cameras and microphones that can ubiquitously capture multimedia content from the environment has fostered the development of Wireless Multimedia Sensor Networks [1], i.e., networks of wirelessly interconnected devices that can retrieve video and audio streams, images, and scalar sensor data. The major objectives behind the research and deployment of sensor networks [2] lie in the following two aspects:

1) Event detection and possible data acquisition by sensing, data processing and communication through node coordination and data transmission [3,4] to the sink or to the interested user.

2) Conservation of energy [5] to maximize the post deployment, active lifetime of individual sensor nodes and the overall network. The reason is that replenishing the

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 793–801.
springerlink.com

energy of sensor nodes by battery-replacement is clearly not feasible for a large network consisting of hundreds of nodes. Moreover, wireless sensors are often deployed in an area which is inapproachable to humans and away from any sustained power-supply.

This ever-increasing popularity of multimedia applications has already started penetrating the domain of wireless sensor networks-thereby giving birth to the new terminology wireless multimedia sensor networks [6]. A quick look into the existing MAC protocols for sensor networks reveals that lack of standardization and application-specific diverse requirements has deprived wireless sensor networks from having a single de-facto standard MAC protocol. Most of the existing MAC protocols for wireless sensor networks can be divided into two categories: 1) time division multiple access (TDMA)-based and 2) carrier sense multiple access (CSMA) based with (possible) collision avoidance (CA) [7].

TDMA protocols have a natural advantage of collision-free medium access; CSMA/CA protocols have a lower delay at varying traffic loads. However, transmitting multimedia applications with QoS offers significant new challenges over these energy-constrained sensor networks. Design of an efficient sensory MAC protocol, satisfying QoS requirements, is one major step in end to end QoS provisioning over WMSNs. Current sensor nodes, such as MICAz and WINS, already support multiple channels for communication, for example, 40 channels in WINS [1]. Thus, by developing a multichannel MAC protocol, which can effectively utilize the available channel capacity through the cooperative work from other sensor nodes, we can achieve a better support for multimedia applications which demand for high data rates [8].

This motivates us to look for QoS-based, yet energy-aware, MAC protocols for WMSNs. The objective of this work is to develop a new QoS-based, energy-aware MAC protocol for WMSNs. In this paper, we propose an adaptive cross-layer multichannel protocol for MAC layer in WMSNs. This protocol use benefits of TDMA and CSMA/CA techniques in one MAC protocol.

The rest of the paper is organized as follows. Section II reviews existing works in sensory MAC protocols. Subsequently, in Section III we explain our proposed MAC protocol at some sub sections. Simulation results in Section IV corroborate the efficiency of the protocol in achieving the desired throughput and delay. Section V concludes the paper.

2 Related Work

A good survey of major MAC protocols for wireless sensor networks is provided in [9]. Provisioning QoS in MAC layers for wireless cellular and local area networks [10] is an active research area, QoS-based MAC protocol for wireless sensor networks have received relatively less attention. While both TMAC [11] and DSMAC [12] attempt to reduce the latency, little of the other MAC protocols are developed with an objective to optimize (or improve) some application-specific quality of service (like delay, through-put etc).

Protocols, like SPEED [13], cluster-QoS [14] and delay-constrained least cost routing [15] discuss the QoS- routing issues in wireless sensor networks.

Unfortunately, all these works attempt to optimize QoS in the sensor- routing from higher layers only. However, end-to-end QoS in WMSNs cannot be satisfied without designing an efficient QoS-aware MAC protocol. Unfortunately only a handful of works exist for QoS-MAC in wireless sensor networks. These include Q-MAC [16], PQ-MAC [17], and RL-MAC [18]. To the best of our knowledge COM-MAC [8] is the first one that use multi channel techniques for MAC protocol in WMSNs. But it uses static time slots at control channel and doesn¡¯t propose any mechanism for nodes that do not have data for sending at start of intervals or for nodes that start sending data between an interval that these could increase delay and degrades throughput of the network.

3 Design of the Proposed Protocol

3.1 Network Architecture

As shown in Figure 1, a WMSN consists of several more powerful nodes as cluster heads that located at the center of different monitoring area, a number of identical and stationary multimedia sensor nodes surrounding each cluster head and a remote data sink which stores the multimedia content locally for later retrieval. Each sensor node can communicate directly with its cluster head and cluster head can communication directly with the data sink using an out-of-band channel. But, if direct communication is not available, multi-hop routing is also employed.

3.2 Our Assumptions

We make some assumptions with relation to the configuration of the network. These assumptions are: 1）Topology of network is cluster based. 2）There are N different channels available for use and all channels have the same bandwidth except one that has lesser bandwidth than others and use as reserved channel, namely channel-R. 3) Cluster heads can transmit or receive on N channels at the same time. Cluster heads will have sufficient power supply and more processing capacity than other sensor nodes. 4）All multimedia sensor nodes in a cluster can transmit or receive on three channels, namely channel-1 and channel-2 and channel-R. Channel-1 is a contention based channel that assigned by cluster head at first phase of network deployment. Channel-2 is contention free and dynamically assigned by cluster head. Channel-R is a contention based reserved channel that is share between a cluster members.5）Sensor nodes are able to switch among channels dynamically. The channel switching time is less than 224 μs according to [1]. 6）The working of a cluster of sensor nodes is synchronized to the cluster head and each sensor node can communicate directly with its cluster head.

3.3 Proposed QoS-MAC Protocol

We assume that the clustering process has been completed by performing a clustering protocol, the assignment of sensor nodes to clusters can be handled by existing clustering techniques [2]. Within each cluster, all tasks are done in time intervals (ΔT),

which are dynamically changed. ΔT as Time interval can be varying depending on application and traffic load of the network.

We suppose three type nodes in a cluster as illustrated in Figure 2. These nodes are: cluster head, active nodes which are nodes that have data for sending, and passive nodes which have not data for sending at present. First of all, when the network is initially deployed, channel-1 allocation phase begin in each cluster and doing only one time. Channel-1 will be used as control channel for sending request message from multimedia sensor nodes to cluster head. As mentioned before the number of channels at cluster head is limited. So, it may happen that a channel assigned to more than one sensor nodes in a cluster. N-1 channels of cluster head could be assigned at this phase as channel-1 of multimedia sensor nodes. One remained channel from N channels of cluster head will be used as channel-R. That is share between all of nodes in a cluster.

The usage of channel-R will be expressed later. The operations of a cluster on ΔT are organized in three sequential phases: request phase, scheduling phase and data transmission phase. We now explain our MAC protocol details in three phases.

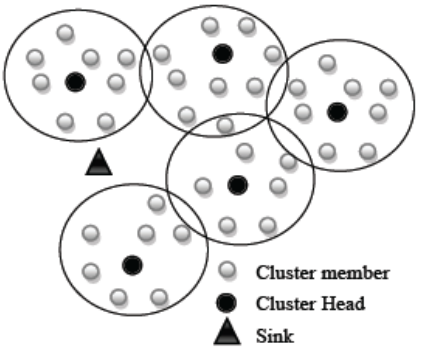


Fig. 1. Network architecture

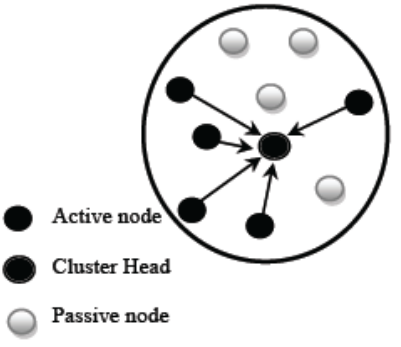


Fig. 2. Node types in a cluster

3.3.1 Request Phase

After channel-1 assignment phase, that runs only one time, the network operations begin at ΔT intervals. At the start of each ΔT, the network layer of each multimedia sensor node determines that whether information exists for sending or not in a cross layer manner. Then nodes that have information for sending, active nodes, start request phase on channel-1, and send a request message (REQ) to the cluster head.

The REQ message includes QoS requirements, such as amount of multimedia data to be transmitted, maximum delay, priority information and traffic class (streaming video, Non-Real Time (NRT), Best Effort (BE)), and Packet Error Rate (PER). Because channel-1 may be assigned to more than one node in a cluster, so, adaptive contention window protocol [19] can be used for better performance on this channel. When active nodes received acknowledgement of its REQ message, they go to standby mode and waiting for scheduling message from cluster head. Request phase time (ΔTr) determined dynamically, based on event rate, traffic load and average of previous ΔTr periods.

3.3.2 Scheduling Phase

After request phase, cluster heads gather REQ messages and then start scheduling phase. Each cluster head calculate an appropriate schedule, based on priority and other QoS requirements that specified in REQ messages, to coordinate the data transmissions of active nodes. Then cluster heads broadcast scheduling messages through all N-1 channels.

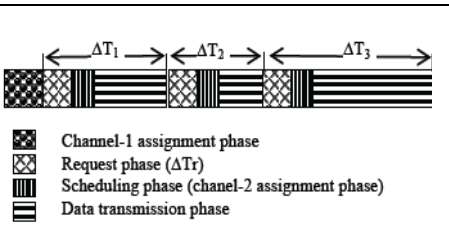


Fig. 3. A cluster operations for three intervals

Scheduling algorithm of Proposed MAC Protocol

1. *Sort "REQs based on Priority in descending order"*
2. *Sort "equal priority REQs based on amount of data in descending order"*
3. **Repeat**
4. *Find "first minimum occupied channel time"*
5. *Assign "channel or timeslot in channel to REQ(i)"*
6. **Until** *REQ (i) exists, i = 1, 2,..., k*
7. *New_ΔT = Find "maximum occupied channel time"*

Fig. 4. Scheduling algorithm

In scheduling message, cluster head assign a channel as channel-2 to each active node for data transmission. If the number of active nodes in a cluster is more than N-1, then a channel should assign to more than one active node. In fact, a time slot in a channel may be assign to an active node as channel-2. Therefore, the scheduling message includes a channel and probably a time slot in it as channel-2 for each active node. Moreover new ΔT that calculated based on REQs is included in scheduling message. New ΔT specifies end of this interval and start of next interval indeed. A cluster time intervals have been illustrated in Figure 3 for three intervals.

Lengths of time slots that assign to active nodes are depending on amount of data that specified in REQ messages and times need for sending acknowledgment messages if needed. Some traffic types may not require acknowledgement message from cluster head, that it should be declared in REQ message.The priority of REQs is essentially based on its traffic class, low priority for best effort traffic, medium priority for non-realtime and high priority for streaming video traffics. Figure 4 shows the pseudo-code of scheduling algorithm that assigns channel-2 and timeslots to active nodes in a cluster. Also announce new_ΔT for next interval. The-reafter, scheduling message broadcasts for sensor nodes in the cluster at network.

If some nodes that were passive get active after this phase, send their REQ messages to cluster head on channel-R. Then, if there is enough unused time on channels, it could be assigned to these nodes. Otherwise, if there is not enough time for some of these nodes, then these nodes only will be announced with new_ΔT. So unsuccessful nodes will go to sleep mode until starting of next request phase. The network throughput in an interval is given by

$$\tau_j = \sum_{i=1}^{k} P_i / ((N-1)T_j C) \quad (1)$$

where k is the total number of REQ messages in jth interval, Pi is amount of bits of data requested for transmission in ith REQ message. Maximum occupied channel time in jth interval is Tj seconds. In other word Tj is equal to new_ΔT at jth interval that used in Figure 4.N-1 is the number of contention free channels available at cluster head. And capacity of each these channels are C bps. The total throughput of network in m time intervals is the average of for j = 1, 2,…m.

3.3.3 Transmission Phase

After receiving a scheduling message by an active node, it could send its data on assigned channel. If a time slot in a channel is assigned to an active node, it could go to sleep mode until its time slot for sending data approaches.

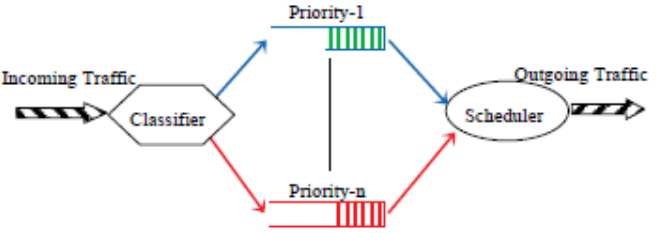


Fig. 5. Traffic differentiation and priority queuing in WMSN

As mentioned before, some passive nodes may get active at this phase, and send its REQ messages at channel-R. So, if free channels or time slots in channels has been assigned to such nodes, they could send its data at scheduled time. When cluster head receives packets from its cluster members, it classifies traffics based on its priority then schedules it for sending toward the sink. Such framework illustrated in Figure 5.

Nodes receive acknowledgement messages (ACK) for proper sent packets, if specified before at REQ message. Nodes could request unused timeslots, if exist, in an interval for lost packets. As mentioned earlier the streaming video traffic is assigned the highest priority and the best effort traffic is assigned the lowest priority. We will now analyze the average delay incurred in each of this traffic class. The mean waiting time of a type i customer is denoted by E(Wi) and E(L iq) is the number of type i customers waiting in the queue. Further let's assume the processing time of traffic class i is μi, with mean E(μi) and residual processing time (Ri), with mean E(Ri). Then the traffic intensity of the system is given by: ρi=λi E(μi) [12]. Hence, for the highest priority streaming video traffic it holds that

$$E(W_1) = E(L_1^q)E(\mu_1) + \sum_{j=1}^{r} \rho_j E(R_j) = \frac{\sum_{j=1}^{r} \rho_j E(R_j)}{1-\rho_1} \tag{2}$$

where r is the number of different traffic classes whose service is in progress during the arrival of the highest priority traffic class. And the mean waiting time for lower priority traffics[7]could be estimated as

$$E(W_i) = \frac{\sum_{j=1}^{i} E(L_j^q)E(\mu_j) + \sum_{j=1}^{r} \rho_j E(R_j)}{1-(\rho_1 + \rho_2 + ... + \rho_{i-1})} \qquad \forall i : 2 \le i \le n \tag{3}$$

4 Performance Evaluation

In this section we show different simulation results demonstrating the efficiency of this proposed MAC protocol. We have developed a discreteevent object oriented packet-level simulator in C++. In the simulations presented in this section, the considered packet size is 25 bytes. We assume 3 data channels at sink that each channel capacity is 250 kbps. Experiments are repeated 10 times. The performance of our algorithm is compared with COM-MAC [8] and a baseline protocol, the multichannel TDMA (M-TDMA) protocol. For M-TDMA, the cluster head first evenly distribute the cluster members on the available channels. Then, the cluster head generates a TDMA schedule on each channel and allocates a fixed slot to each cluster members.

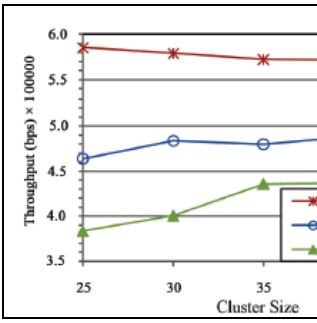


Fig. 6. Throughput performance for various cluster sizes up to 45

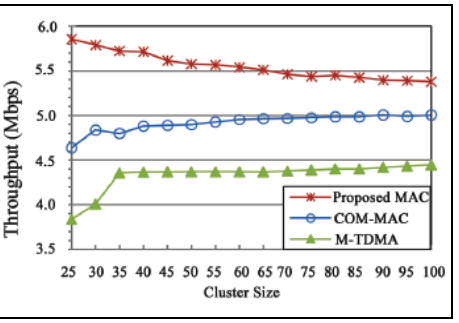


Fig. 7. Throughput performance for various cluster sizes up to 100

Figure 6 compares the network throughput performance of proposed MAC protocol with COM-MAC and M- TDMA for different cluster sizes. Our protocol works well. But, when the number of nodes in cluster increases, the proposed protocol throughput decreases.

To find the reason we repeat simulation up to 100 nodes in a cluster, Figure 7, our protocol throughput reach to a steady state and it has better throughput than other protocols. This is because the channel tends to be saturated when more nodes are trying to utilize the channel. As expected, proposed MAC protocol outperforms other two protocols. This is because our protocol is designed to maximize the network throughput by adaptively changing intervals and using unused channels and time slots for passive nodes that get active after request phase.

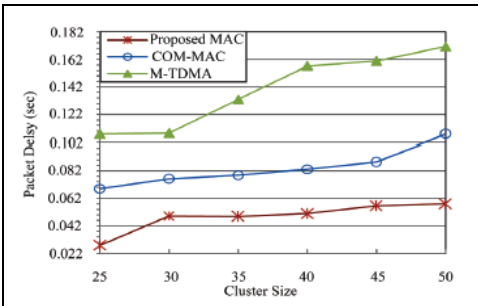


Fig. 8. Packet delay performance for various cluster

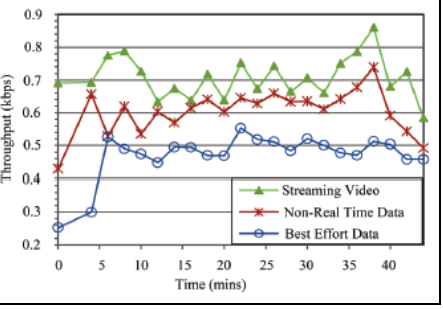


Fig. 9. Differentiated MAC-throughput

Figure 8 shows the delay performance compassion of three protocols as cluster size increases. We see that our proposed MAC incurs lower delay when compared to COM-MAC and M-TDMA. It is because that in our protocol, when a passive node gets active, it could send request from channel-R to use unassigned space on channels for data transmission. So it incurs lower delay than other two protocols. With COM- MAC and M-TDMA such nodes should wait until next intervals thus this increases packet delay. We also notice that the delay performance increases as cluster size increases. This is because that larger cluster size will lead to heavier network load so that a packet has to wait longer to be transmitted.

Figure 9 explains the throughput-dynamics for different traffic classes. The novelty of our MAC protocol is that it first classifies the traffic into different classes depending on the type of service (ToS) then schedules it for data transmission. The streaming video traffic is given the highest priority, the NRT traffic is given the second priority and the BE traffic attains a lowest throughput.

5 Conclusions

In this paper we have developed a cross layer multichannel QoS-MAC protocol for wireless multimedia sensor networks which classifies the wireless traffic into different class, and adaptively assigns channel to various traffics. In our proposed protocol nodes get active only when the network layer specifies that there are data for sending. We verify the advantages of our protocol through network simulation, in terms of network delays, throughput and differentiated throughput of different traffic classes. We see that our proposed MAC protocol provides better energy-efficiency, high-throughput, and data reliability support in WMSNs.

References

1. Akyildiz, I.F., Melodia, T., Chowdury, K.R.: A Survey on Wireless Multimedia Sensor Networks. Computer Networks 51(4), 921–960 (2007)
2. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A Survey on Sensor Networks. IEEE Communications Magazine 40(8), 102–114 (2002)

3. Heinzelman, W., Kulik, J., Balakrishnan, H.: Adaptive Protocols for Information Dissemination in Wireless Sensor Networks. In: Proceedings of the 5th Annual ACM/ IEEE International Conference on Mobile Computing and Networking, Seattle, WA, pp. 174–185 (August 1999)
4. Yao, Y., Gehrke, J.: The COUGAR Approach to In-Network Query Processing in Sensor Networks. ACM SIGMOD Record 31(3), 9–18 (2002)
5. Mini, R.A.F., Machado, M.D.V., Loureiro, A.A.F., Nath, B.: Prediction-Based Energy Map for Wireless Sensor Networks. Ad Hoc Networks 3(2), 235–253 (2005)
6. Akyildiz, I.F., Melodia, T., Chowdhury, K.R.: A Survey on Wireless Multimedia Sensor Networks. The International Journal of Computer and Telecommunications Networking 51(4), 921–960 (2007)
7. Saxena, N., Roy, A., Shin, J.: Dynamic Duty Cycle and Adaptive Contention Window Based QoS-MAC Protocol for Wireless Multimedia Sensor Networks. Computer Networks 52(13) (2008)
8. Li, C., Wang, P., Chen, H.-H., Guizani, M.: A Cluster Based On-demand Multichannel MAC Protocol for Wireless Multimedia Sensor Networks. In: IEEE International Conference on Communications, Beijing, May 13 (2008)
9. Demirkol, I., Ersoy, C., Alagoz, F.: MAC Protocols for Wireless Sensor Networks: A Survey. IEEE Communications Magazine 44(4), 115–121 (2006)
10. Kuhn, T.: A QoS MAC Layer for Ambient Intelligence Systems. In: Proceedings of the 4th International Conference on Pervasive Computing, Dublin, pp. 69–72 (2006)
11. Dam, T.V., Langendoen, K.: An Adaptive Energy-Efficient MAC Protocol for Wireless Sensor Networks. In: Proceedings of the 1st International Conference on Embedded Networked Sensor Systems, Los Angeles, pp. 171–180 (2003)
12. Kleinrock, L.: Queueing Systems, Theory. John Wiley & Sons, New York (1975)
13. Hea, T., Stankovica, J.A., Lub, C., Abdelzahera, T.: SPEED: A Stateless Protocol for Real-Time Communication in Sensor Networks. In: 23rd IEEE International Conference on Distributed Computing Systems, Rhode Island, USA, pp. 1–10 (May 2003)
14. Tang, S.S., Li, W.: QoS Supporting and Optimal Energy Allocation for a Cluster Based Wireless Sensor Network. Computer Communications 29(13-14), 2569–2577 (2006)
15. Gao, Q., Blow, K.J., Holding, D.J., Marshall, I., Peng, X.H.: Radio Range Adjustment for Energy Efficient Wireless Sensor Networks. Ad-Hoc Networks 4(1), 75–82 (2006)
16. Paek, K., Kim, J., Song, U., Hwang, C.: Priority-Based Medium Access Control Protocol for Providing QoS in Wireless Sensor Networks. IEICE Transaction Letters on Information Systems E90-D(9) (2007)
17. Sabena, N., Roy, A., Shin, J.: Dynamic Duty Cycle and Adaptive Contention Window Based QoS-MAC Protocol for Wireless Multimedia Sensor Networks. Computer Networks 52(13), 2532–2542 (2008)

The Application of Improved GA-BP Algorithms Used in Timber Price Prediction

Hong-e Ren, Yan-mei Ma, and Ben-zhi Dong

Information and Computer Engineering College, Northeast Forestry University, Harbin, Heilongjiang 150040, China
mym08@yahoo.cn

Abstract. Data mining technique is an effective tool used to obtain desired knowledge from massive data. BP Neural Network and genetic algorithms(GA) are widely used in the application of data mining. The paper proposes an improved GA-BP mix algorithms according to feature of GA global optimization and advantage of improving convergence speed when used additive momentum and adaptive learning in BP algorithm and then used in data mining on timber price prediction. The results show that it can improve convergence speed and achieve a high precision within finite step with the improved GA-BP algorithms.

Keywords: Neural Network, genetic algorithms, price prediction, data mining.

1 Introduction

Data mining is a process of nontrivial extraction of implicit, previous, unknown and potentially useful knowledge from a large amount, incomplete of noisy, fuzzy and random data. The extracted knowledge represent as the forms of rules and regularities and patterns and so on. To speak more exactly data mining is a Decision Support process which primarily based on artificial intelligence, machine learning and statistics. Data mining highly automatically analysis original data of enterprise and makes the reasoning conclusion, digs out the potential patterns and predicts the customers behavior, and then helps decision-makers adjust the market strategies to reduce risks and make the right decisions.

We can get the timber market classification forecast with data mining techniques used in the timber sale and price statistics which focused on classification prediction about demand and supply total and composition changes of timber species. Timber sale and price classification prediction provide scientific and comprehensive and reliable basis for arranging wood cutting strategy, inspecting the production and sale policy and price, strengthening forestry market management and analyzing the effect, trends and rules of timber sale and price changes.

Although the traditional methods used widely and the effect is good, it is difficult to deal with complex nonlinearly relations and many complicated influencing factors. Artificial Neural Network(ANN) has good nonlinear mapping and self-learning abilities and it can stabilize the complicated interactions among various factors which

T. Zhang (Ed.): Mechanical Engineering and Technology, AISC 125, pp. 803–810.
springerlink.com

influence demand for example the basic requirements, cyclical factors, the market and season random factors and so on in the form of weight value to the internet by learning, which improve the prediction accuracy. Therefore, this paper uses ANN technology to study the timber price prediction.

2 Back Propagation Neural Network and Genetic Algorithm

ANN is a complex network system widely interconnected by vast and meanwhile very simple processing unit(be called a neuron) and is a highly complex nonlinear dynamic system. Neural Network has the abilities of distributed storage and treatment, massively parallel, self-organizing, self-adaptive and self-learning and is especially used to process the information processing problems which need to consider many factors and conditions as well as inaccurate and fuzzy. The M-P model proposed by McCulloch and Pitts is the earliest and most influential neuron model. It is the basic of most Neural Network model. As shown in the figure 1 below.

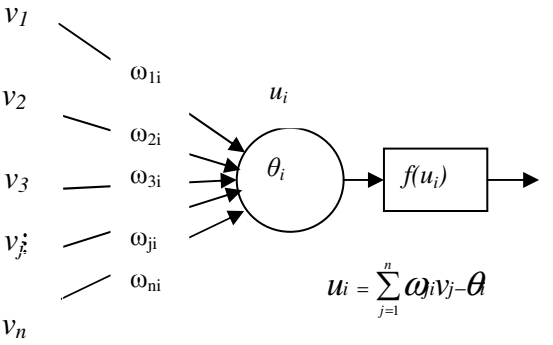


Fig. 1. M-P model

For the input of the *i*th neuron: $v=[v_1, v_2, \cdots, v_n]$, the weight value connected with the *i*th neuron is $\omega_i=[\omega_{1i}, \omega_{2i}, \ldots \omega_{ni}]$.

The threshold of neuron itself isθ_i and the output Y can be expressed as $Y = f(u_i) = f(\sum_{j=1}^{n} \omega_{ji} v_j - \theta_i)$ in which u_i represents the *i*th neuron's active value namely neuron's state and the function f() expresses the input and output characteristic of neuron. In M-P model f() is defined as Step Function: $v_i = f(u_i) = \begin{cases} 1, u_i > 0 \\ 0, u_i \le 0 \end{cases}$, that is to say output and input have the excitation and inhibition two states which excitation is 1 and inhibition is 0.

2.1 Back Propagation Neural Network Model

Back Propagation (BP) is one of the most widely used Neural Network model, which is a multiple layers feedforward network reverse propagated according to error. It can learn and store large numbers of input-output model mapping relations without the beforehand revealed and described mathematical equation. BP is composed of input layer, output layer and hidden layer. Every layer is composed of parallel operational simple neuron. The neuron among layers is full interconnected network and the neuron in the same layer is not mutual connected. Its structure is shown in the figure 2 below.

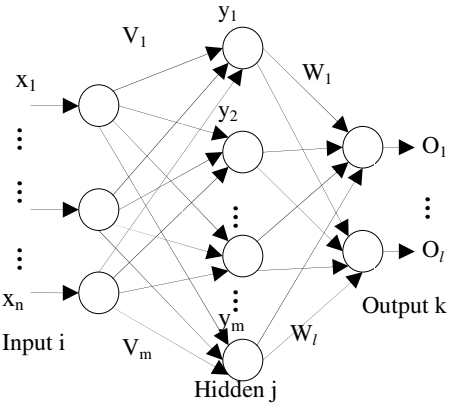


Fig. 2. Three layers BP network structure

BP algorithm is a supervised learning algorithm. Its main idea is that for input learning samples: *X1,X2,..., Xn,* we know their corresponding output sample: *D1,D2,...,Dm* and the aim of learning is to modified the weight value according to the differential value between actual output *O1,O2, ..., Om* and target vector *D1,D2,...,Dm* so as to make the input *O* as close as possible to the expected *D* which means that let sum of square error in network layer be least. It approaches the target by continuously calculating network weights and bias in error function slope downward direction.

2.2 BP Algorithm and Its Improvement

The standard BP algorithm has the following prime defect: (1)Slow convergence speed.(2) Easy to fall into local minimum. (3) Difficult to determine the network structure. Structural parameters (including the number of hidden layers, the number of hidden neurons) lack of theoretical guidance, usually only based on experience to choose.

In practical application, the standard BP algorithm is difficult to do, so many improved algorithms are proposed. To improve BP algorithm two ways are mainly adopted, one is heuristic learning method and the other is using a more effective optimization algorithm. This paper uses the strategy combined momentum with adaptive learning rate to increase the learning speed and the algorithm reliability.

Momentum method reduces the sensitivity the network toward error curve local details and effectively inhibits the network into the local minimum. Standard BP algorithm is actually a simple steepest descent static optimization algorithm and we only correct weights w(k) in accordance with negative gradient direction of the *K*th step without taking the previous accumulated experience(the former gradient direction) into account, which often makes the learning process oscillate and convergence slow. Momentum weight adjustment algorithm is[1]: w(k+1)=w(k)+ α[(1-η)D(k)+ηD(k-1)].

In which α is the learning rate and η(0 < η< 1) is the momentum factor. The momentum added in equivalent to damping, which reduces the oscillating trend in the learning process so as to improve the convergence. Another important reason for the slow convergence of standard BP algorithm is that the choice of the learning rate is undeserved, too small the convergence too slow and too big possibly correct too much leading to oscillation even divergence. Self-adaptive learning rate weights adjustment algorithm is: w(k+1)=w(k)+α(k)D(k).

$$\alpha(k)=\begin{cases}1.05\alpha(k-1) & E(t)<E(t-1)\\ 0.7\alpha(k-1) & E(t)>1.04E(t-1)\\ \alpha(k-1) & other\end{cases}$$

When the gradient of two successive iterations are in the same direction, means falling too slowly, we can increase the learning rate; when reverse means dropping too much, then we can correspondingly reduce the learning rate. Combined the above two methods together we get momentum-self-adaptive learning rate BP neural network weight value correction algorithm: w(k+1)=w(k)+α(k)[(1-η)D(k)+ηD(k-1)].

2.3 Genetic Algorithm and the Improved BP Algorithm

Genetic Algorithm(GA) is a random search and optimization algorithm based on heredity theory of Mendel and evolutionary of Darwin[2].This algorithm simulates the natural selection and genetic mechanism and takes survival of the fittest as operational criterion and searches for optimal solution in the global. In the process of searching optimal solution, GA begins with an initial individual group and gradually finds out the optimal solution until meeting the convergence criterion or a predetermined iterations number.

Genetic algorithm has the prime following characteristics:

(1) Using multiple search point search information in the same time. GA begins to search for the optimal solution from an initial population composed by many individuals rather than from a single individual.
(2)Taking the target function as search information directly. The genetic algorithm uses only the fitness function value transformed by the objective function value without the need of objective function derivative value and some other auxiliary information to determine the further search direction and search range.
(3) Using a probabilistic search technology. The search process has a good flexibility because of the selection, crossover and mutation operations being in a probabilistic way.
(4) Perform an efficient and heuristic search, not blindfold or completely random search in the solution space.

(5) Unlimited to optimization function, which not require continuous function or a differentiable function. It can be an explicit function expressed by mathematical formula, also the mapping matrix even the implicit function in neural network.
(6) Features of parallel computing. It is suitable for large-scale and complex optimization problems.

GA has the global search and population optimization characteristics. Combined GA with neural network, it not only uses the ability of neural network generalization mapping, but also makes the neural network overcome the defect of the slow convergence rate and easily falling into local minimum. It is a very vital significance to make full use of their respective advantages and overcome their disadvantages to improve the ability of neural network solving the problem. This paper combined GA with the improved momentum-self-adaptive learning rate BP algorithm (the improved GA-BP algorithm) to predict timber price, which improves the convergence speed and makes the network away from being trapped in local minimum. Figure 3 is the algorithm flow chart GA optimizes BP network.

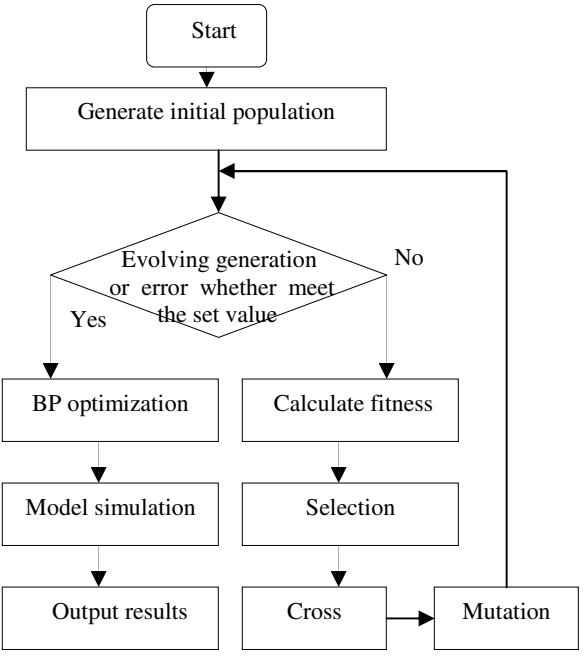


Fig. 3. The algorithm flow chart GA optimizes BP network

3 Timber Price Prediction Model Based on Improved GA-BP Algorithm and Instance Analysis

3.1 Determine the Nodes Number of Input Layer, Output Layer and Hidden Layer

In BP network topology, the input node and output nodes are decided by the problem itself which lie in the number of hidden layer and hidden layer nodes. The node

number of input layer and output layer is respectively the number of input vector and output vector. The node number of input layer is the sum of the number of selecting historical data and the influencing factors. The factors influence timber price and historical data this paper selects are M(compared with last month)PI(price index), Y(compared with the same month last year) PI(price index) and CPI (consumer price index). There are 4 input vectors(4 input layer nodes). The output vector is timber price in prediction time. It is a one-dimensional vector so the output layer node number is 1.

The increase of the number of hidden layer can improve the BP network nonlinear mapping ability, but if it exceeds a certain value the network performance will decrease. Theoretical study shows that, for any continuous function in closed interval can be approached using the single hidden layer BP network. It means one 3 layers BP network can complete the arbitrary n-dimensional mapping to m-dimension. This paper adopts a single hidden layer of 3 layers BP neural network to establish the prediction model. According to the Kolmogorv theory (input layer node number is K, then the hidden layer node numbers is 2k+1) we set the node number of the hidden layer is 9 (from the previous analysis the value of K is 4) [3].

3.2 The Sample Data and Instance Simulation

The data in Table 1 is the price list of Larch logs which timber length is 4 meters and diameter class is 24 to 28cm from HanJiayuan Forestry Bureau in January to October 2010(from Internet and being arranged). Data in January to September is used to train the BP network and in October is used to test the trained BP network. Data in table 2 is simulation value. Figure 4 and figure 5 are respectively the standard BP algorithm and improved GA-BP algorithm simulation results in which abscissa is training step and the ordinate is the expected target error value(0.000001 is set in this instance).

Table 1. The data of instance

Month	Price (yuan)	Price index		CPI
		YPI	MPI	
1	806	98.53	99.19	100.0
2	814.3	99.16	100.63	101.0
3	825	98.24	99.08	99.6
4	835	99.12	100.90	99.6
5	845	100.99	101.89	99.5
6	926.7	103.19	101.81	99.6
7	850	103.74	100.53	100.0
8	950	104.48	100.71	100.1
9	850	104.66	100.18	100.4
10	850	104.26	100.35	100.8

Table 2. The simulation data

algorithm	Mean square error	Predictive value	training step
standard BP	0.0052	885.9951	1938
improved GA-BP	0.0041	839.6862	321

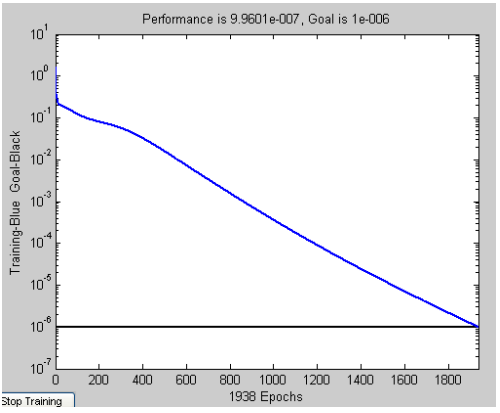


Fig. 4. The standard BP algorithm simulation results

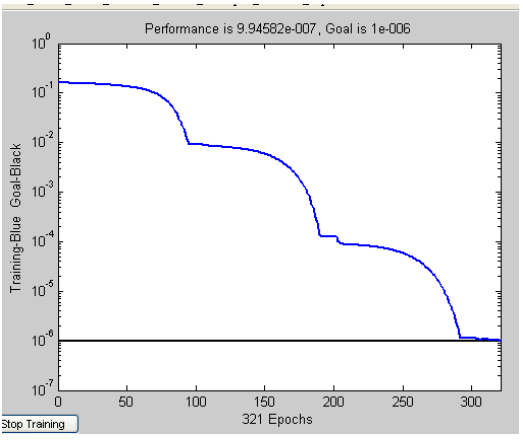


Fig. 5. The improved GA-BP algorithm simulation results

The comparison of Figure 4, figure 5 and table 2 shows that using the improved BP algorithm the prediction error was 0.0041, predictive value was 839.6862 (actual value was 850), training step was 321, while the standard BP algorithm for after training the results were 0.0052, 885.9951, 1938. Therefore, the improved GA-BP algorithm in convergence speed and prediction accuracy is obviously better than the standard BP algorithm.

4 Conclusion

This paper uses the improved GA-BP algorithm and established a timber price prediction neural network model. This model can well adapt the nonlinear relationship among the influence factors and had a higher predictive precision. The improved GA-BP algorithm can simultaneously select and optimize multiple points within the solution space and search after finding out the optimal point according to the negative gradient direction. Not only avoid the BP algorithm being trapped in local minimum and slow convergence but overcome the defect of long searching time and speed slow caused by GA in exhaustive form to find the optimal solution. The experiment results show that the improved GA-BP algorithm is greatly improved in computational precision, generalization capability and convergence speed, which has a certain practical value.

Acknowledgements. This study is supported by national forestry bureau projects of "948" (2010-4-05).

References

1. Jin, R.-J.: Momentum-Adaptive-Learning-Rate BP Algorithm Compared With the Standard BP Algorithm and Its Application. Microcomputer Applications 17(7), 30–32 (2001)
2. Xiao, G.-R.: Application Research of Fund Net Value Prediction Based on BP Neural Network. Computer Simulation 28(3), 373–376 (2011)
3. Xu, J.-B., Qi, Z.-G., Hang, W., et al.: Research on Application of Improved BP Algorithms in Price Prediction for Highway Engineering Main Material. Journal of Highway and Transportation Research and Development 25(4), 62–65 (2008)
4. Li, X.-Z., Zhang, Y.-W.: Application of Improvement Self-Adaptation Genetic Algorithm in BP Neural Network Learning. Journal of Tianjin University of Science & Technology 25(4), 64–67 (2010)
5. Deng, K., Zhao, Z.-Y.: Research and Simulation of Stock Price Prediction Model Based on Genetic Algorithm BP Neural Network. Computer Simulation 26(5), 316–319 (2009)
6. Doganis, P., Alexandridis, A., Patrinos, P., et al.: Time series forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. Journal of Food Engineering, 196–204 (2006)
7. Satyadas, A., Chen, H.: An Application of Intelligent Neural Network to Times Series Business Fluctuation Prediction. In: IEEE International Conference on Neural Networks, pp. 3641–3645 (2004)

Author Index