Automatic System for Grading Paper Based Multiple Choice Exams

By

Rose Mohammad AL-Zuo'bi

Thesis Submitted in partial fulfillment of the requirements for the degree of

Master of Computer Engineering specialty of Embedded Systems

Advisor

Prof. Hussein AL-Zoubi

Yarmouk University, Hijjawi Faculty for Engineering Technology

December, 2017

Automatic System for Grading Paper Based Multiple Choice Exams

By

Rose Mohammad AL-Zuo'bi

Thesis Submitted in partial fulfillment of the requirements for the degree of

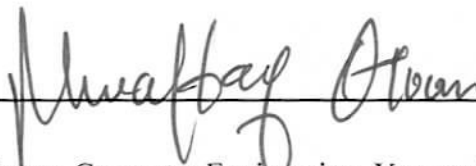Master of Computer Engineering specialty of Embedded Systems

APPROVED BY

Prof. Hussein AL-Zou'bi ——————————————— , **Chairman**
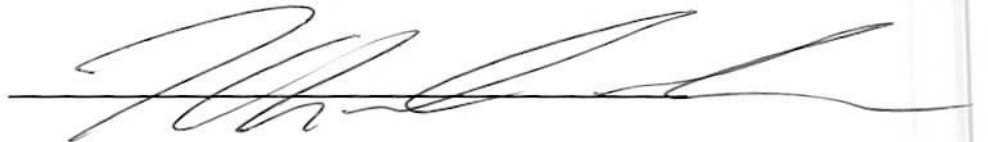
Professor, Computer Engineering, Yarmouk University

Dr. Mwaffaq Al-Otoom ———————————————

Associate Professor, Computer Engineering, Yarmouk University

Prof. Khalid Darabkeh ———————————————

Professor, Computer Engineering, University of Jordan

DECLARATION

Plagiarism is the breach of copyright or using another person's work and pretending that is one's own.

I, **Rose AL-Zuo'bi**, recognize what plagiarism is and I hereby declare that this thesis, which is submitted to the department of **Computer Engineering** at Hijjawi Faculty for Engineering Technology, for the partial fulfilment of the requirements for the degree of Master of Science in Embedded Systems Engineering, is my own work. I have not plagiarized from any sources. All references and acknowledgments of sources are given and cited in my thesis. I have used the conventional citation and referencing. Each significant contribution to and quotation in this report from work of other people has been attributed and referenced.

Eng. Rose Mohammad Al-Zuo'bi   December 30, 2017

ACKNOWLEDGMENTS

DEDICATION

My very special thank go to the candles whom bum themselves to light my way my **parents**, to my **sisters**, my **brother** for all their love and support as well as for the understanding they show it to me in every circumstance.

I would also like to deeply thank to the one I can't smile without him my lovely husband, **Hazem AL-Qateshat**, your great love, support, and encouragement lets me find hidden energy during my Master studies and for all the life.

My heart beats daughter **Juman** and son **Ali** for happiness they bring to our life and my newborn son **Ahmad** for allowing me to finish my experimental work and write the thesis manuscript late into my pregnancy.

My soul mate, **Maiosh**, all the words in the world couldn't describe the feelings of being your sister is your soul mate too, sister you love her like no other, she loves you like no other. Simply according to my proposed voting algorithm …You are the Winner.

My lovely family, love you all

Rose . . . .

ABSTRACT

Rose AL-Zuo'bi, Automatic System for Grading Paper Based Multiple Choice Exams, Master of Science in Computer Engineering, department / Embedded Systems, Department of Computer Engineering, Yarmouk University,2017,(Advisor: DR. Hussein AL-Zoubi)

Nowadays, many computer-based grading of multiple-choice exams are available worldwide. Despite this strive towards automating multiple-choice exams, pencils and papers are still widely used especially when the number of examinees submitting the exam exceeds the number of computers.

Manual-grading for a large number of multiple-choice exams is a tiresome and an error prone process. Therefore, this is a field where machines would prove more capable than humans. Employing an automated system for grading reduces grading time, human effort, as well as reducing the probability of grading errors.

In this work, a novel approach for developing an automatic, efficient, fast and offline fixable system is introduced that can be used for grading scanned multiple-choice exams. This system guarantees high accuracy based on a similarity matching algorithm.

The scanned answer sheets are processed by robust algorithms for noise removal and feature extraction. After which they are passed through edge and line detection, right answer detection, and scores accumulation.

We focused on multiple marked cases of irregularities that are observed by exam takers and accounted for them in our system; whereas existing automatic exam grading systems failed to address them. The proposed grading system relies on four voting functions for similarity measurement. In addition to providing an automatic grading for each answer sheet, a confidence ratio is also provided as well as list the questions that require manual inspection.

This proposed grading system does not require any special hardware or any special requirements. The provided experimental results are based on actual exam sheets from several semesters and various graduate and undergraduate courses and diverse student levels showing superior performance compared to existing solutions.

Table of Contents

List of Figures

www.manaraa.com

List of Tables

X

Acronyms

MCQ	Multiple Choice Questions
OCR	Optical Character Recognition
OMR	Optical Marked Recognition
ANNs	Artificial Neural Networks
AI	Artificial Intelligence
IDE	Integrated Development Environment
GIS	Geographic Information System
EPR	Exam Paper Reader
GUI	Graphical User Interface
ROI	Region of Interest
CCF	Cross Correlation Function
MSE	Mean Square Energy
MAD	Minimum Absolute Difference

CHAPTER 1

Introduction

1.1 Topic Overview

The classical multiple-choice question (MCQ) where a student chooses one answer option from many answering options supplied (one of five choices labelled as A, B, C, D and E). Figure 1 shows a classical format of answer matrix for MCQs. MCQs answer matrix consist of the question number and option cells provided for choosing the right answer. The provided answers include the correct answer in the list of options in addition to distracted choices or the wrong answers [1].

Figure 1: A classical format of answer matrix for MCQs [1].

Careful tailing of multiple choice tests guarantees that MCQs exams can tests the student level as essay exams do [2]. For effective exam three important content characteristics should be achieved validity, clarity and reliability. The exam's validity means that the exam measures the student knowledge, skillfulness, and efficiency that supposed to be obtained over the period of the course. Exam clarity is an evaluation of the question's quality that covers the course subjects. Exam reliability based on general grading consistency and getting reliable class evaluation curve for students of differing performance levels. Well-designed multiple-choice exams can be more effective, clear and reliable than essay exams because they cover most of material subjects, differentiate between student's levels and grading consistency is achieved [2].

For effective multiple-choice exam content characteristics should carefully tailed There are some of the general features of MCQs are summarized by [3] [4]:

1

- Coverage of grate part of the syllabus via compulsory questions. This makes the test efficient.

- MCQs are objective that means there must be one answer.

- Multiple-choice exam is easy to grade, and that feature attracts teachers to choose this type of exams. Where there are online and offline specialist software for grading process ,or anyone with the key answers sheet can accomplish grading mission; sometimes students can grade for each other's.

- Feedback Provision can be effectively targeted in compare off essay exams because of the limited possibilities in which students can have gone wrong.

- MCQs check the quality of exam construction. That's mean, if all the class (the superior and the weak students) answer equally right (or equally wrong) on a question, this means that the question is not tailed professionally.

1.2 Motivation

Multiple choice questions (MCQs) exams are widely used for examination, especially when there is huge number of students, where they are marked quickly with minimal training or preparation, MCQs are high reliable, and provided an effective way of assessing factual knowledge. But manually grading hundreds of multiple choice exams is a tedious work and is prone to errors. Therefore, it is a task more suitable for a machine than for a human. So, using of automated grading systems saves time, enables faster processing of exams and reduces the probability for grading errors.

Due to MCQs importance and widely usage, many researchers have presented automated systems for grading MCQs to reduce time that the teachers waste in the manual grading. But these systems have some limitations and restrictions. Therefore, in this thesis, a robust, and reliable automatic MCQs grading systems presented.

2

1.3 Contribution

The main challenges in developing automatic MCQs grading systems are the standardized answer sheets template, how students can annotate the answers in the answer matrix, and what he has to do if he changes his mind or annotate the wrong answer.

The effectiveness proposed system that it:

- Used a low cost answer sheet template that can be printed by any printer and these avoid using of bubbles answer sheets or OMR sheets.

- The system is not based on handwriting character recognition classification, so respect to the best of our knowledge other related works doesn't support any annotation mark for answering the question. In our system although the exam instructions states to use the (X) mark for annotate the right answer and darken the cell to revoked the wrong answer but in real data set students annotated the answers in the answer matrix by any character, Figure8 shows samples of used marks, and they revoked mistakes by any way of correction like darken the answer cell, scratching the mark or erase it the system dealt with all special cases and recognize his last answer.

- Applying the proposed four voting systems to recognize the approved student answer in cases where the student has more than one answer for the same question obtained high confidence of results, where the choice get the highest voting result is the winner.
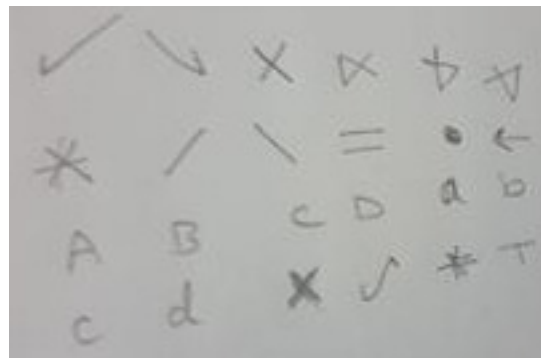
Figure 2: Examples of answers characters or marks by students.

The proposed system has a possibility to do the following activities:

- Read the scanned answer sheets.
- Recognize student answer for each question.
- Compare student answers with key answer, which is already stored in the system database.
- Apply Scoring Rules
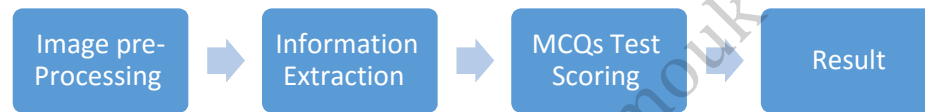
Figure 3: The application structure.

1.4 Thesis Outline

The rest of thesis is recognized as follows:

Chapter 2: In this chapter we briefly reviewed all required concepts, which are related to the proposed work like Sobel edge detection, Hough Transform, cross correlation and absolute differences.

Chapter 3 presents the proposed methodology followed to design and implement the proposed grading system.

Chapter 4: Presents experimental results obtained proposed grading system and compares the obtained results with existing methods.

Finally, **Chapter 5** concludes this thesis and provides some ideas for a future work that could guide us or others to improve this work.

Chapter 2

Literature Review

Nowadays and because of the high expansion in technologies, automatic grading of multiple choice exams is an important need for university courses which have large number of students. Many available systems provide automatic grading using simple form for answer sheets easy to use.

Early Optical Mark Recognition (OMR) systems required OMR scanners, readers, and OMR forms, and that infrastructure needs a sizable budget. In [5] the optical recognition commits many restrictions on the quality of paper, the layout of the answer sheets, and even restrictions on the used ink. These sheets are then marked by comparing them with a key sheet. Finally, the final score is obtained after applying grading scheme. Although OMR applications are useful in automate the grading for a large number of exams sheets with 100% accuracy rate. But the cost of the used equipment (like an optical scanner) and the data input time for these systems (time for scanning the sheets) limit their use to high enrollment exams, where the time and cost deficiency are considerable.

In this thesis our contribution is to overcome OMR limitations. Where our proposed system doesn't require OMR forms, any restrictions on the type of paper, OMR scanners and readers or using pencil for answering to choose only one answer. And it's also low cost system with high accuracy rate of 99.8%.

Authors in [6] presented a scanner-based optical mark recognition technology called ''low-cost'' OMR solution, where exam forms can be printed using any printer or copier and software tools used to process the results. Soon Low-cost OMR tools are widely used in educational courses. Although requiring some time and effort in the scanning phase for adjustments, these tools recently integrated with conventional learning management systems.

Authors in [7][8] presented their own open-source desktop application called *FreeFormScanner*, *which* is desktop application developed by using Java that can communicate with scanners to

digitalize the input answer sheets. The system scans a predefined answer sheets format as seen in Figure4 on which the correct answer cell is marked by the student. Then, information is collected from the scanned answers-sheets.

Figure 4: Answer sheet template of FreeFormScanner [7].

The system proposed is not fully automated, as it requires human help in some cases for example, while question is normally have only one answer and student selected more than one answer. This case happens often when a student annotates an option by mistake, and then black another one, or when the student makes an annotation error, i.e. decides to change his answer. In similar cases, the student is should annotate a special error column (where there more than answer although only a one answer is expected) and handwrites the letter refers to the correct answer in a dedicated cell. When the system discovers that there are multiple answers, or the error column had been annotated, a dialog box will appear to the human operator, who is required to review the answer sheet and the handwritten letter and feed it back into the system. Hence, a certain amount of supervision is required during the automated grading process.

In [9], authors launched updated version of their version in [7][8] . And adhere to the general automated scoring system depends on finding image edge distances and using random forest classifier, and propose a limited OCR for recognizing the labels of multiple choices of automated grading. Their algorithm was specifically designed to the set of supported characters. It is classifier designed so as to detect handwritten answer by measuring properties characteristic to predefined classes. The customized answer template shown in Figure3. Answer table consists of rows of empty bubbles, where bubbles represent multiple choices of the question in the same row. The

student should darken the right bubble. In the case that a student decides to revoke his answer, the student should annotate the bubble in the "Error" column for the same question or darken the correct answer in the same row, and write the alphabet label of his final answer in the rectangle near the appropriate row. Also illustrated in Figure3, when the system detects that the "Error" cell is marked, or that the question is multiple marked, they used computer vision technique for handwritten letter recognition.
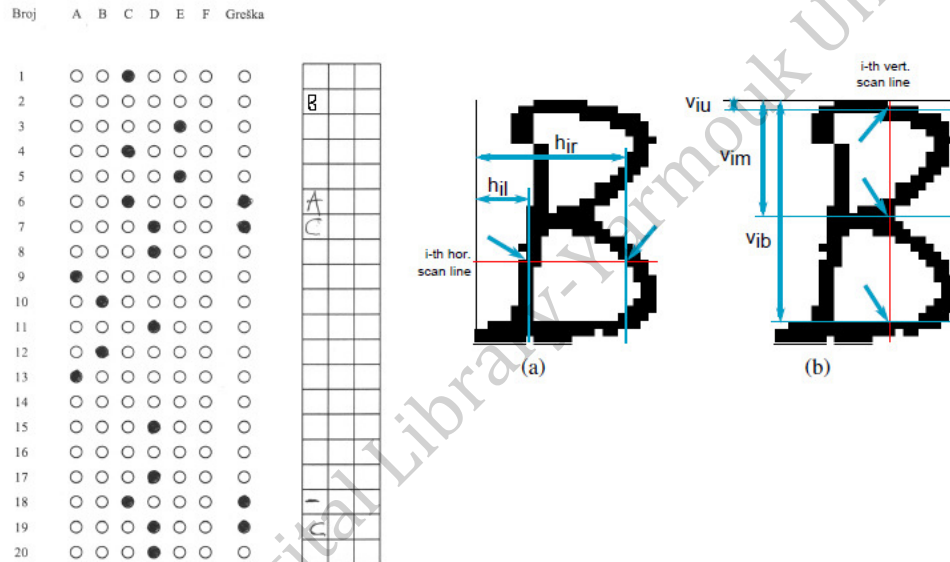
Figure 5: Definition of character features [9].

Our full automated system achieved higher accuracy rate (99.8%) than FreeFormScanner rate. Where its random forest classifier can only classify correctly 96.5% of samples from the test set (3821 of 3960).

In [10], Kim et al. used various features represented in around feature vector of 120 dimension, applied to a very difficult problem of handwritten recognition of a limited cursive handwritten words set. Minimum Distance between the boundaries of the rectangle and image pixels, along the scanned horizontal and vertical lines were utilize. By applying a heterogeneous classifier constructed of two multi-layer perceptions and a hidden Markov model, their results recognition accuracy is 92.7% for 32 classes of legal words. They also used distances of the image elements from the predefined minimum boundaries of a whole word. Their results recognition accuracy is

7

of 92.7%. They get this high accuracy by using a heterogeneous classifier constructed of two multiple layer perceptron's and a hidden Markov model.

Similarly, in [11] Surinta et al. used the distance values between the closest black pixels and predefined hotspots in four or eight different directions to represent characters. They reported a recognition rate of 90.1%.

The program in [12] was written in MATLAB programming language to grade a specially designed MCQ exam scanned papers with 15 questions with four choices for each question, where the student can choose only one answer per question. The correct way to answer is shown in Figure4. The square must be completely filled with dark heavy mark so that we cannot see the letter inside the square, light or partial marks may not be read properly by the scanner.
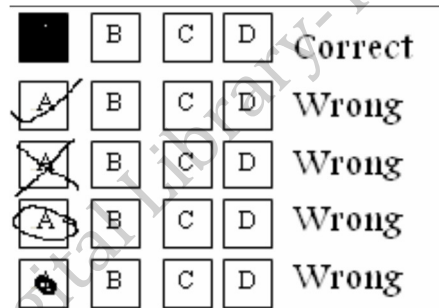
Figure 6: The correct way for annotating the right answer [12].

In the preprocessing stage, the image is converted from a grey image to a binary image. The binary (Black and White) image contains all the information necessary to discern the object's outline. In image registration stage, the image is aligned to the base image. After making the registration, it is only required to cut the center of the base test paper and the test paper (where the important information is located) by using the crop function. In evaluation stage, ANDing process was applied between the inverted base image and answer image. This will result in an image that contains only the correct answers, where the correct answer may be defined as marking the same square in the base and answer sheets after the ANDing process, the squares that resulted from the ANDing process and these squares are considered as the correct answers. By calculating the percentage of the correct answers to the total answers we can get the final score.

The experiment conducted on a group of students and its result is 100%. But doesn't achieved all student satisfaction, where some of the students complained that the shapes of the answers (squares) are difficult to be filled (especially on the edges) so that it is better to change from the square shape to another shape. So our contribution was to support any way of answering and support revoking the first answer and choose any other answers.

Development of an unconventional method were discussed in [13]. The method used instead of usual alphabets (A, B, C, D and E) a numerical arithmetic as choices of answers options, a successful technique for administering short and simple MCQs was used as a tool of estimating student levels in a focus area of the subjects of the course. In [13] an MCQs test was designed, where in the five options (n = 5) of answer to a question are labeled with different numerical values according to the multiplication of numeral (n + 1) and not by the letters A, B, C, D and E as usually. When the exam finished, the candidate will overall the student choices and the test result can then be obtained by exporting the total choice value in a simple spreadsheet program. The spreadsheet program will decode that number and accumulate the number of correct student answers. In the last algorithm for grading MCQs instead of using alphabets letters as labels for choices of answer, a numbering system was proposed. Therefore, in place of A, B, C, D and E, the labels of options will be designated as 1, 2, 3, 4 and 5 for the first question and 6, 12, 18, 24 and 30 for the second question. And continue numbering the rest of the questions in the same way. These numbers which refers to the options of the questions answer were derived, based on the multiplication (n + 1) = 6, where the number of choices to choose from for each question is n = 5. A simple equation will run using Excel spreadsheet and read input values to calculate the total value of the student's choices in the specified column, the program will decode the input value and accumulate the number of correct answers annotated by the student for that MCQs test.

Eftimie, Iordache and Rughiniş for example in [13] developed an application that not only automatically evaluates the paper based multiple choice tests, but also provide an environment for test generation, questions management, storage and other features. Where the system selects a certain number of questions from examination database which contains a large number of questions to generate a test paper that can satisfy the requirements of examination, and the formation of questions should be so reasonable to reflect different examination targets.

The application proposed in [14] not only can process individual scanned multiple choice tests, but also can process series of multiple choice scanned tests. The first stage of the application is to process the scanned images. This has to be done at the resolution of 200 dpi. It is not relevant if the image is colored or not, because the original image will be transformed so as it will end as a binary one. Basically, the image enhancement can be done either through adjusting the intensity of the pixels, or by the histogram equalization. The two operations have been implemented in the GUI of the application.

The clipping operation is performed to produce two separate images. The first one is the upper left corner that contains the information about the person that filled in the test, while the second one contains the area with the choice rectangle. The operation of test automatic evaluation is done by binarization of the region of interest with small squares. Then complement of the image and filtering the noise are performed by determining the checked square to generate the vector of answers and compare it with the correct solution vector.

The idea behind the character recognition is that a template with characters 0-9 and A-Z has been created. The template is a matrix of 36 bitmap (the 10 figures and the 26 capital letters). With OCR limitations the success rate in the process is at least 95%. Using of template matching instead of OCR achieved higher accuracy of (98.8%).

The proposed grading system in [15] uses a scanner to capture answer sheets rather than using a costly OMR system. The proposed system grades a specially designed MCQ exam test and feedback-analysis paper. Also students can give the feedback about the lecturer who conducts lectures during the semester. Their application hacks the answer sheet scanner, detect responses for each question, stores the scanned data in a database and compare student responses with reference answers which would be pre stored in a master database. At the end of the test, the test result can be obtained by inputting the total choice value in a simple spreadsheet as illustrated in Figure5.This project used C# from Microsoft, and SQL (Structured Query Language) as a relational database management system (RDBMS).
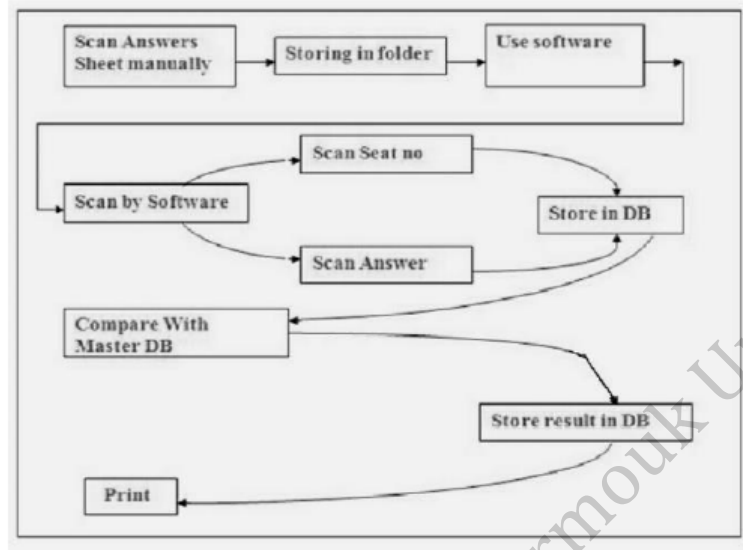
Figure 7: System block diagram [15].

Eyegrade in [16] is a system for automatic grading of multiple choice exams. Eyegrade based on using off-the-shelf webcam to offers a truly low-cost solution. Additionally, performs both optical character recognition and mark recognition, to avoids using the bubbles in the answer matrix. The system is based on techniques of computer-vision. After the exam has taken place, and examinee annotate the appropriate cells by crossing them from corner to corner, where they may choose to clear a cross by perfectly filling the cell entirely with ink. The system ignores these cells allowing students to change his mind a question to ''un-answered''. The webcam captures a stream of images, and the system detects the properly framed answer sheet. then extracts it from the image, stored and shown to the instructor, who can review and correct them. When the system incorrectly detects an answer (either a false positive or false negative), the user interface allows the user to correct it by simply clicking in the appropriate cell.

Once an image is captured from the webcam, the color mode is transformed. The straight lines are detected by using the Hough transform [16]. The geometry of the answer cells is detected by intersecting the lines detected in the last step. The position of the corners of every cell in the answer tables is defined . Then the system decide, for each cell, whether it has been marked or not. The system expects cells to be marked with a cross sign in order to detect these marks; the cell is masked by a thick cross placed where the mark is expected as shown in Figure6. The percentage

of marked pixels behind the mask is computed. If it is higher than a given threshold, the system decides that the cell is marked.
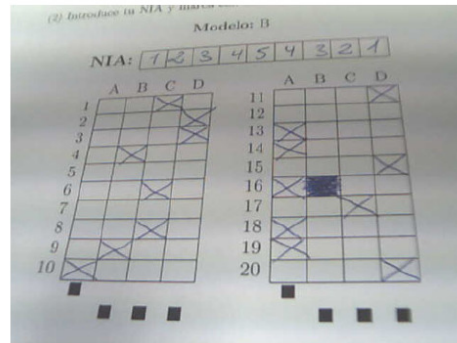
Figure 8: Example capture of an answer sheet [16].

Eyegrade precision in detection of the automatic system correct answers is 97.0 % without user intervention. The remaining 3 %, most of the erroneous decisions were due to students not following the instructions to mark their answers. The flexibility of our system our system raises our precision in detection to 98.8%. Where our system is fixable enough to accept any way for marking the question, and any way for revoking the answer. 29 answer sheets (0.6 %) of Eyegrade dataset failed despite being properly marked due to weaknesses in the detection system. Our system doesn't require special hardware setup of tray or webcam, and using any scanner doesn't require accurate alignment of test paper.

A new solution approach was suggested in paper [17] for developing an automatic system capable of recognizing student ID number and his corresponding marks from answer sheet and storing them automatically. This system attached to a web camera, which capture sheet images of the cover page of each answer script-A cover page template has been proposed for easy extraction as shown in Figure7. Boxes for Student ID number and Marks have a red boundary.

www.manaraa.com

Figure 9: Template for answer sheet of the suggested EPR [17].

A formula used to process the image and transfer it from the RGB space to the HSI space. The motivation of this transform is the that red color can be extracted efficiently. Also, the mathematical representation is much easier in the HSI space. Once the red color is extracted, red box corners can be found out. The region of interest (ROI) is defined from the coordinates of the corner. Then, the image is digitized into a binary one, so the processing time decreases. The image obtained after noise reduction is skeletonized, finally use simple OCR engine for the recognition .

Accuracy of the suggested system depends on the sample space size of OCR system. The developers trained the OCR with 50 samples of numerals set (0-9).and got only 8 1 % average accuracy. Using of template matching instead of OCR achieved higher accuracy of (98.8%), and detecting answer table lines without coloring restrictions makes our system more reliable and fixable.

CHAPTER 3

Background and Theoretical Framework

3.1 Template Matching

Template matching is a modern vision method used to match part of a figure to predefined template. It also presents a digital image methodology to recognize matching splits in the image with a predefined template. The main goal is to match the predefined template to an image under processing wherever the template is a sub of that image und get the best match. Templates are usually used to recognize characters, numbers, and other symbols. It can be used for edge detection of images, in quality controlling of industrials to navigate PLCs and robots. Template Matching is the process of discovering parts of an image which matches (are indistinguishable) a template image (patch). That's required two crucial segments. [18]

Source image (I): The picture under investigation which include a match to the template image.

Template image (T): The patch pattern that the input image compared with to discover similarity. Matching technique finds the similarity measure however it computes the accuracy rate among images by calculating the Mean Squared Error (MSE) metric [18].

$$MSE(Mean\ Squared\ Error) = \frac{\Sigma\Sigma(Temporary(u,v) - Target\ (u,v)}{Number\ of\ pixels} \tag{1}$$

Wherever, Temporary (u, v) and Target (u, v) are the strength values of Template image and the input image respectively.

Template matching approaches generally categorized into featured-based approach, area based approach or template based approach.

14
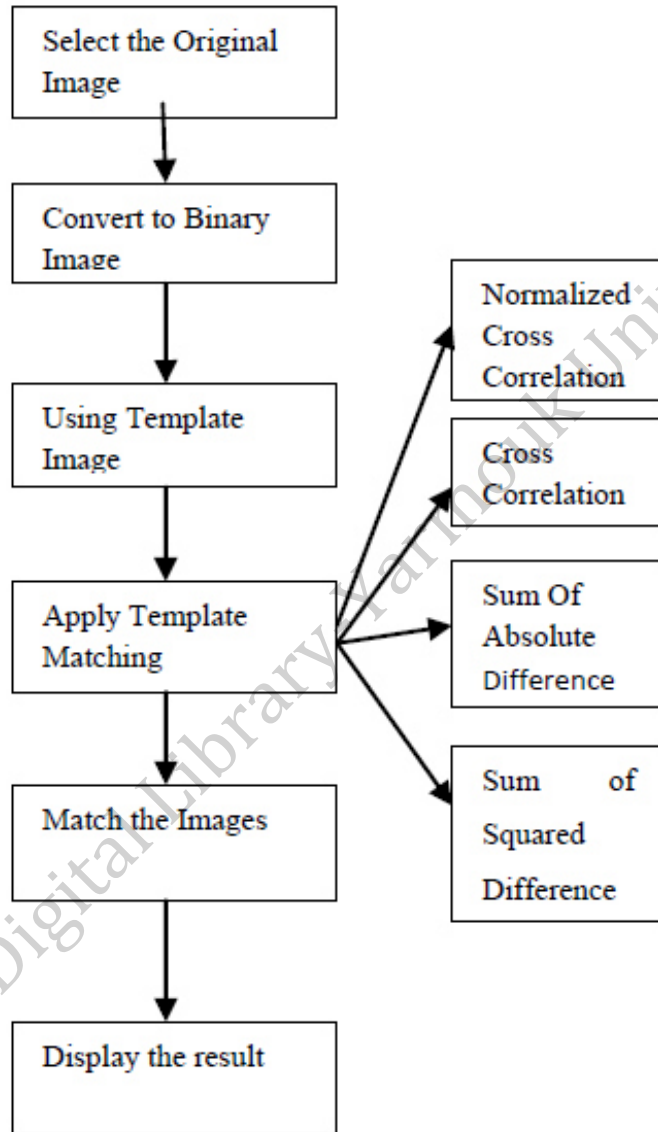
Figure 10: Template matching flow chart [18].

3.2 Sobel Edge Detection

Edge detection methods assume that the edge occurs when the intensity function cuts out when a very gradient steep of intensity occurs at the image. Based on this assumption, Sobel detection takes points when the derivation of intensity values are maximum to locate edges. The gradient vector measures how pixel values are changed with distances in x-y direction rapidly [19] [20].

Gradient components are computed using the Equations (2) and (3) [20].

$$\frac{\partial f(x,y)}{\partial x} = \Delta x = \frac{f(x + dx, y) - f(x,y)}{dx} \tag{2}$$

$$\frac{\partial f(x,y)}{\partial x} = \Delta y = \frac{f(x, y + dy) - f(x,y)}{dy} \tag{3}$$

Where dx measures the distance along the x-direction and dy measures the distance along the y-direction. At the discreet images, dx and dy can consider a number of pixels between two different points [20].

$$\Delta x = f(i + 1, j) - f(i, j) \tag{4}$$

$$\Delta y = f(i, j + 1) - f(i, j) \tag{5}$$

1 (pixel spacing) present the point when pixel coordinates are (i, j) thus,

In order to detect the presence of a gradient discontinuity, Sobel calculates gradient changing at (i, j) to detect gradient discontinuity using the following magnitude measure [20].

$$M = \sqrt{\Delta x^2} \tag{6}$$

Where the gradient direction θ is found by [20].

$$\theta = \tan^{-1}\frac{\Delta y}{\Delta x} \tag{7}$$

3.3 Hough Transform

Hough transform is one of the most popular techniques that are used to isolate features in an image. These features are being specified in a parametric form. The Hough transform is commonly used for regular curve detection like lines, circles, parabola, ellipses, etc. Hough transform employed in applications where a detailed analytic description of image feature(s) is needed [21]. Usually classical Hough transform is used more than the general Hough transform because the computations of General Hough transform algorithm are more complex than the classical Hough transform. The most important advantage of the Hough transform algorithm is that it is indulgent of gaps in boundary descriptions at the image features and is also unaffected by noise and occlusion in the images.

Hough transform technique is used to compute a global feature's description, given local measurements. The main idea in the Hough technique for line-detection is that each input measurement presents a contribution indicator for a globally consistent solution. Figure11 shows some detection forms.
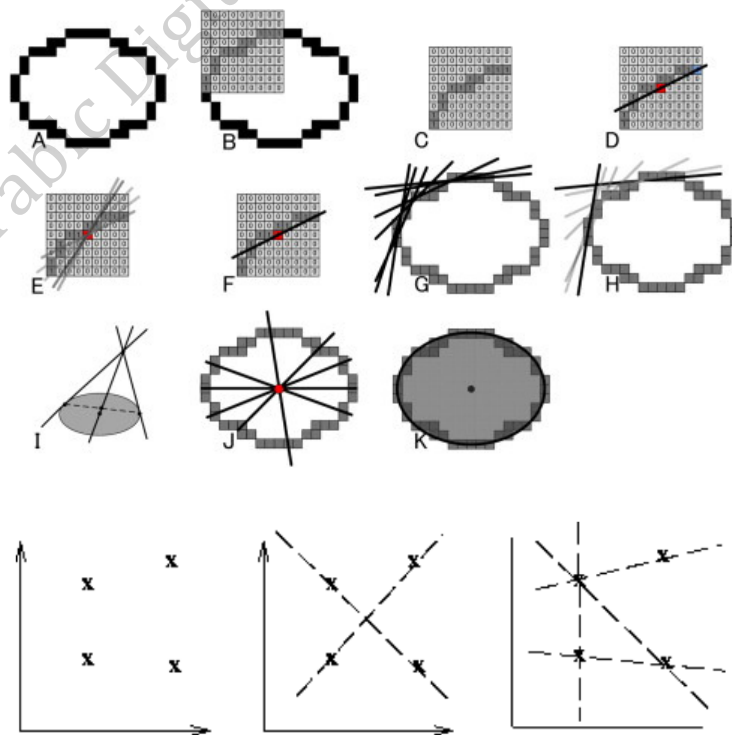

Figure 11: Coordinate points and possible straight line fittings [21].

We can describe any shape in a number of analytical forms. Table below describes a set of parameters and equations which are used in the Hough transform [21].

Table 1 : Shapes Analytics Forms

Analytical Form	Parameters	Equation
Line	r, q	$X \cos q + y \sin q = r$
Circle	x_0, y_0, r	$(x-x_o)^2 + (y-y_0)^2 = r^2$
Parabola	x_0, y_0, r, q	$(y-y_0)^2 = 4r(x-x_o)$
Ellipse	x_0, y_0, a, b, q	$(x-x_o)^2/a^2 + (y-y_0)^2/b^2 = 1$

3.4 Cross-correlation

Cross-correlation is a similarity measuring of two series as a relative function of the displacement which is also known as a sliding dot product or inner-product [22]. Cross-correlation is used for searching a long signal for a shorter and to find known features to serve several applications such as [23]:

- Pattern Recognition
- Single Particle Analysis
- Electron Tomography
- Averaging
- Cryptanalysis
- Neurophysiology.

For continuous functions f and g, Cross-correlation is defined as [22].

$$(f * g)(\tau) \stackrel{\text{def}}{=} \int_{\infty-}^{\infty} f * (t)g(t + \tau)dt, \tag{8}$$

For discreet function f and g, the cross-correlation is defined as [22].

$$(f * g)[n] \stackrel{\text{def}}{=} \int_{m=\infty-}^{\infty} f * [m]g(m + n) \tag{9}$$

Autocorrelation is the cross-correlation of a signal with itself, where always there is a peak at a lag of zero with size equal to the signal energy.

In statistics and probability terms, the cross-correlations is defined between two random vectors X and Y and the correlations of a random vector is considered to be the correlations between the vectors itself [23].

3.5 Sum of absolute differences

Sum of absolute differences (SAD) is a similarity measurement between image-blocks in the digital image processing. Absolute difference is calculated between each pixel with corresponding pixel at the original block for comparison. Then all differences are summed to create a similarity metric of block. Which may be used for several purposes [24]:

- Object recognition
- Generate disparity maps for stereo images
- Estimate motion for video compression.

The sum of absolute differences presents a simple method to find the objects inside an image in an automated way, but it could be effected by some contextual factors like shape, size, color, lighting, or viewing direction. The sum of absolute differences is used in conjunction with other recognition methods like cross correlation for results' reliability improvement. [24]

$$SAD = \sum_{(i,j)\in R} (f(i,j) - g(i,j)) \tag{10}$$

19

3.6 Mean Squared Error (MSE):

The mean squared error is a similarity measurement of a regression line to a set of points. This calculated by finding the distances from the points to the regression line (these distances are the "errors") and squaring them. The squaring is necessary to remove any negative signs. It also gives more weight to larger differences. It's called the mean squared error as you're finding the average of a set of errors.

Calculation of the mean squared error for a set of temporary and target values by subtracting temporary value from the target values to get the error, square the errors, then add up the errors, and finally find the mean. [18]

Mean Squared Error (MSE) metric [18].

$$MSE(Mean\ Squared\ Error) = \frac{\Sigma\Sigma(Temporary(u,v) - Target\ (u,v)}{Number\ of\ pixels} \qquad (11)$$

Wherever, Temporary (u, v) and Target (u, v) are the strength values of Template image and the input image respectively.

3.7 Pattern Recognition

Pattern recognition is one of machine learning branches that focuses on analyzing similarity and patterns in data either using a predefined training data or for discovering previously unknown patterns. Pattern recognition get the attention of researchers due to its wide application areas like speech recognition, face recognition, biometric recognition, classification of handwritten characters OCR, medical diagnosis, GIS and many more [25].

Pattern recognition consists of mainly three processes as illustrated in Figure 12. First stage is data acquisition or the process of converting the form of input data (character, voice, images etc.) into computing device acceptable form for further processing by using sensors, optical machine and scanners. Second stage after data acquisition is data analysis. During this stage information is

20

collected of the available pattern classes in set of data. This information or knowledge about the data is used for further processing. Pattern recognition third stage is classification. It's the stage of decision making and that decides the class of new data depending on knowledge received from learning data [26].
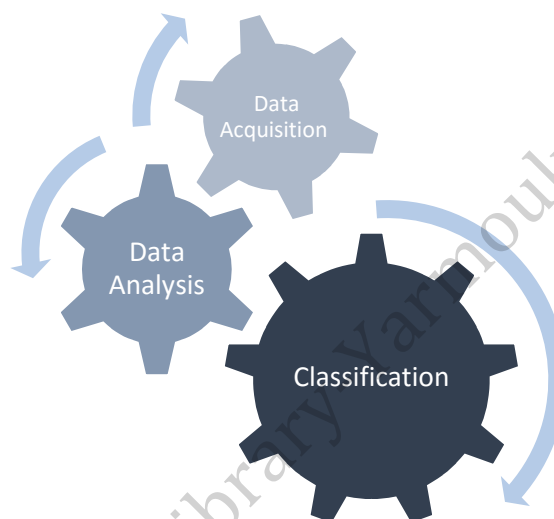

Figure 12: Processes of pattern recognition.

Data set presented to a pattern recognition system is divided into two portions: the first for training and the rest for testing purpose. Training set is used to make the system learn the patterns and testing is set for checking the efficiency of system. To develop a pattern recognition system with high performance and high capability of handling massive amounts of data. Three main elements should be considered (i) amount of data (ii) technology used (method) (iii) designer and the user. There are various models for pattern recognition like: Statistical Techniques, Structural Techniques, Template Matching, Neural-Network based techniques, Fuzzy models and Hybrid Models [27].

3.8 Optical Character recognition (OCR)

Optical Character recognition (OCR) has attracted researcher's attention for a long period of time. Character recognition can be defined as the process for recognizing any handwritten character of any language from human world to computer operating system environment. But the big challenge

21

still the Recognition of handwritten text, due to diversity in shapes and intensity of different people handwriting [17].

Hardware used in any OCR system shows a notable improvement in the speed of insertion, fast retrieval and reduce the possibility of human error. Figure13 shows the logical components of any OCR system. Camera or scanner with OCR software and an output interface between the user and the machine language.
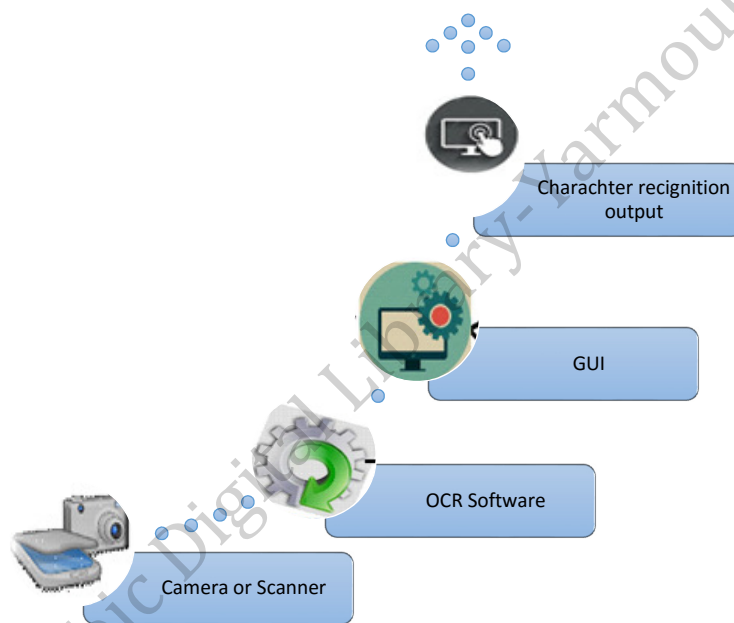

Charachter recignition output

GUI

OCR Software

Camera or Scanner

Figure 13: Logical component of OCR system.

CHAPTER 4

Proposed Approach

4.1 Methodology

An automated multiple-choice scoring system is proposed in this thesis to replace manual scoring or limited existing systems.

Unlike many existing systems, our application is designed from scratch to handle any handwritten marks to offer more flexibility and accuracy.

The power of this work comes from handling student's answers regardless if they used alphabetic character or any other signs in their answers to overcome all limitations related to the previous work. So, we didn't use any specific recognition methods like OCR in this work.

Our execution framework includes the following main components

- Custom printed answer-sheet template; that can be reproduced using any photocopier.
- Scanner.
- MATLAB Application for scoring process.
- Microsoft Excel; for imported results.

The Microsoft word answer-sheet template is designed to accommodate up to 50 standard multiple-choice questions with five possible answers cells (A, B, C, D and E) as placeholders for the correct answers. Which is printed using laser or inkjet printers on white A4 paper that could be reproduced using any white/black photocopier.

Scanner is used to the collected answered papers from students and store all scanned images to the input folder.

Image scanning done by any scanner with low resolution of 200 dpi. While the original image doesn't have much gradient in color between the white background and black table lines, and it will be enhanced to binary one by the histogram equalization.

23

MATLAB code is developed using MATLAB R2015/Mathwork Platform from scratch to read scanned images, apply all enhancement procedures as the preprocessing procedures to produce suitable images for scoring and return the scores into the excel sheet for testing and validation.

As shown in Figure14 our answer-sheet template includes the following objects

1) Header: General information section including the University, faculty and exam information.

2) Student information section including the student name and ID.

3) Exam instructions.

4) Answers matrix that can accommodate up to 50 exam items with 5 possible choices.

5) Answers' signs (letters, or any used symbol).

6) Footer: page number section.

Figure 14: Answer-sheet template.

Speed is a very important factor for any application. Hence, the computation time is enhanced by processing only the important parts. In this work, only the ROI is processed representing part 3 for

localization and scoring, which is defined by finding the middle of the image (half way in the image Horizontal and Vertical) and counts the number of white pixels found there, then scans left to right, top to bottom, right to left, bottom to top to find 98% of that original middle pixel count .The proposed application ignores regions 1, 2, 3 and 5. This step saves time and reduces the total computation operations to have more system that is efficient.

In addition, this chapter elaborates on the building of the table detection, ROI's segmentation and mark-signs recognition modules, and concludes the used dataset with a few remarks.

4.1.1 Proposed Scoring Algorithm

The proposed methodology is illustrated using the following algorithm.

Proposed Algorithm

Input   : Scanned Images answer sheet format
 	
Output : Final scores, list of multiple marked exported to excel sheet
 	
Main ( )
 	
  // Read Input Image.
  Call Read_Image ( );
  // enhance image (Noise, Redlines, Alignment).
  Call Image_Quality ( );
  // Find Answer Matrix
  Call QABoxes ( );
 	
  While (Q_n <= 50)
 	
      // Find Q_Type: Multi_Mark, Single_Mark or Not_Mark
     Call QABox_Test ( );
     If (QABox == Single_Mark)
     Call Scoring ( );

25

If (QABox == Multi_Mark)

    Call Voting ( );

    If (QABox == Not_Mark)

    QABox_Grade = 0;

Return Final_Scors ( );

Pre_Processing ( )

  Call RGB_To_Gray ( );

  Call RIO ( );

  Call Noise_Reduction ( );

  Call Readlines_Mask ( );

QABoxes ( )

  Call Sobel ( );

  Call Hough ( );

Scoring ( )

  If (QABox_Position == AnswerKey_Position)

    QABox_Grade = 1;

  Else

    QABox_Grade = 0;

26

```
Voting ( )

    Call CCF (  );

    Call SAD (  );

    Call MAD (  );

    Call MSE (  );

    Return QABox_Wiener;

    Call Scoring (  );

    Write Multi_Marked_List;
```

4.2 Research Data-Set

4.2.1 Images Data-Set

This section provides a general description about the image data set, which is used in this research.

Our data-set includes more than 800 images of 50 questions real exams held at Yarmouk University, those images are taken over a long time by Dr. Hussein AL-Zoubi over the years 2011 to 2016, covering 9 semesters and 5 main courses of computer engineering department. That's improve the system by processing various images patterns of handwritten marks, hard special cases for multi marked samples. Different color-modes, and noise types.

4.2.2. Data set Formats

The dataset is divided into two main formats:

Original Images

The input dataset contains grayscale and true colored Images RGB

Table 2: Original Input Dataset specifications

Input Dataset Specifications	
Color Representation	Grayscale, RGB
Resolution	1716*2464
Size	< 900 KB
Bit depth	8, 16, 24 and 32 bits
Vertical/horizontal resolution	200 dpi

Grayscale Images

These input images are prepared during the pre-processing stage to convert colored images into grayscale mode with simplest format, which provides just the important details at the image for fast computing issues.

Table 3: Processed Input Dataset specifications

Processed Input Dataset Specifications	
Color Representation	Grayscale
Resolution	516*507
Size	< 100 KB
Bit depth	8, 16, 24 and 32 bits
Vertical/horizontal resolution	96 dpi

4.2.3 Dataset Contributions

The data set contains more than 800 images, which are collected from several sources:

- Real data set from Yarmouk University, those images are taken over a long time by **Prof.Hussein ALZoubi** over the years 2011 to 2016, and presents more than 800 students answering patterns.

  These data set covering many semesters and many courses in computer engineering department:

  - CPE-462 Data communication- Spring of 2012.
  - CPE-562 Computer network course – Summer of 2012.
  - CPE-350 Object oriented molding and application course –Fall of 2013.
  - CPE-562 Computer network course – Fall of 2014.
  - CPE-350 Object oriented molding and application course –Spring of 2014.
  - CPE-562 Computer network course – Summer of 2014.

29

o CPE-562 Computer network course – Fall of 2015.

o CPE 452 Soft Ware Engineering -Spring 2015.

o CPE 568 Multimedia systems and Technology-Spring 2015.

Each course has four forms of the exam (A, B, C and D), all images grouped according to the forms with their key answer sheet.

My personal collection images which taken during the testing stage to enrich dataset with many unexpected handwritten marks like Figure15.
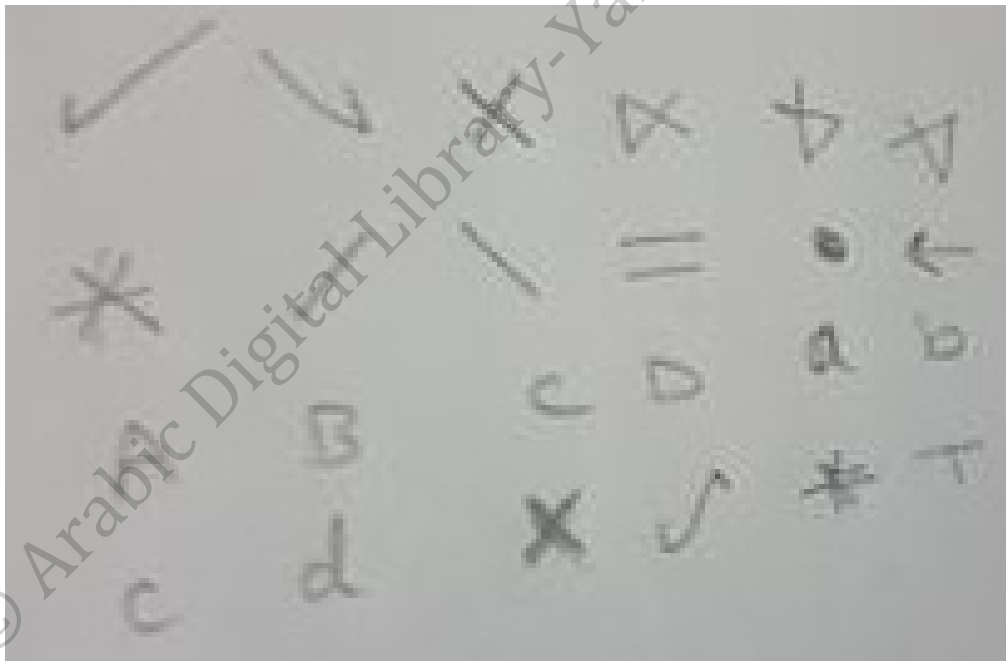

Figure 15: Answers signs examples.

4.3 Implementation

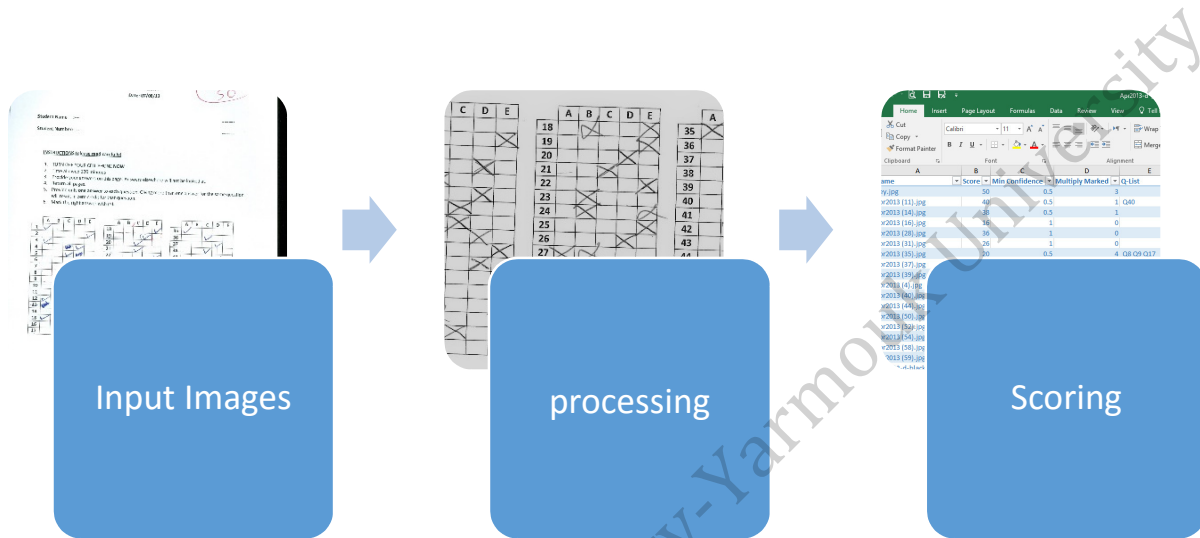The high level stages of the proposed methodology is illustrated in Figure16.

Figure 16: High Level Block Diagram of the automated multiple-choice scoring application.

4.3.1 Research Framework

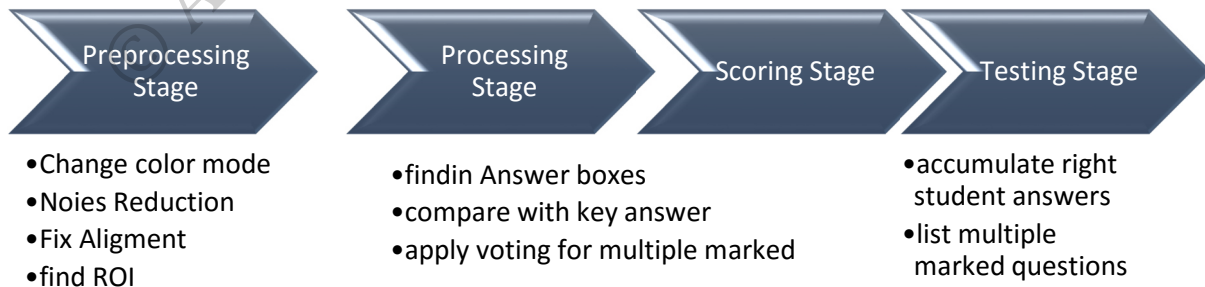In this section, we outline the framework for scoring multiple-choice exams in an efficient way as shown in Figure.17

Preprocessing Stage
- Change color mode
- Noies Reduction
- Fix Aligment
- find ROI

Processing Stage
- findin Answer boxes
- compare with key answer
- apply voting for multiple marked

Scoring Stage
- accumulate right student answers
- list multiple marked questions

Testing Stage

Figure 17: Framework stages.

The basic framework is divided into four main stages. Below the flow charts in Figure 19 to 22 and descriptions for each stage in details.

Preprocessing Stage: In this stage

- o Input image (colored image) is converted to grayscale image by using MATLAB function *rgb2gray.*

- o Noise is redacted by using Morphological MATLAB function clean.
- o Not aligned images is aligned and finding the rotation angle by Hough transform.

32
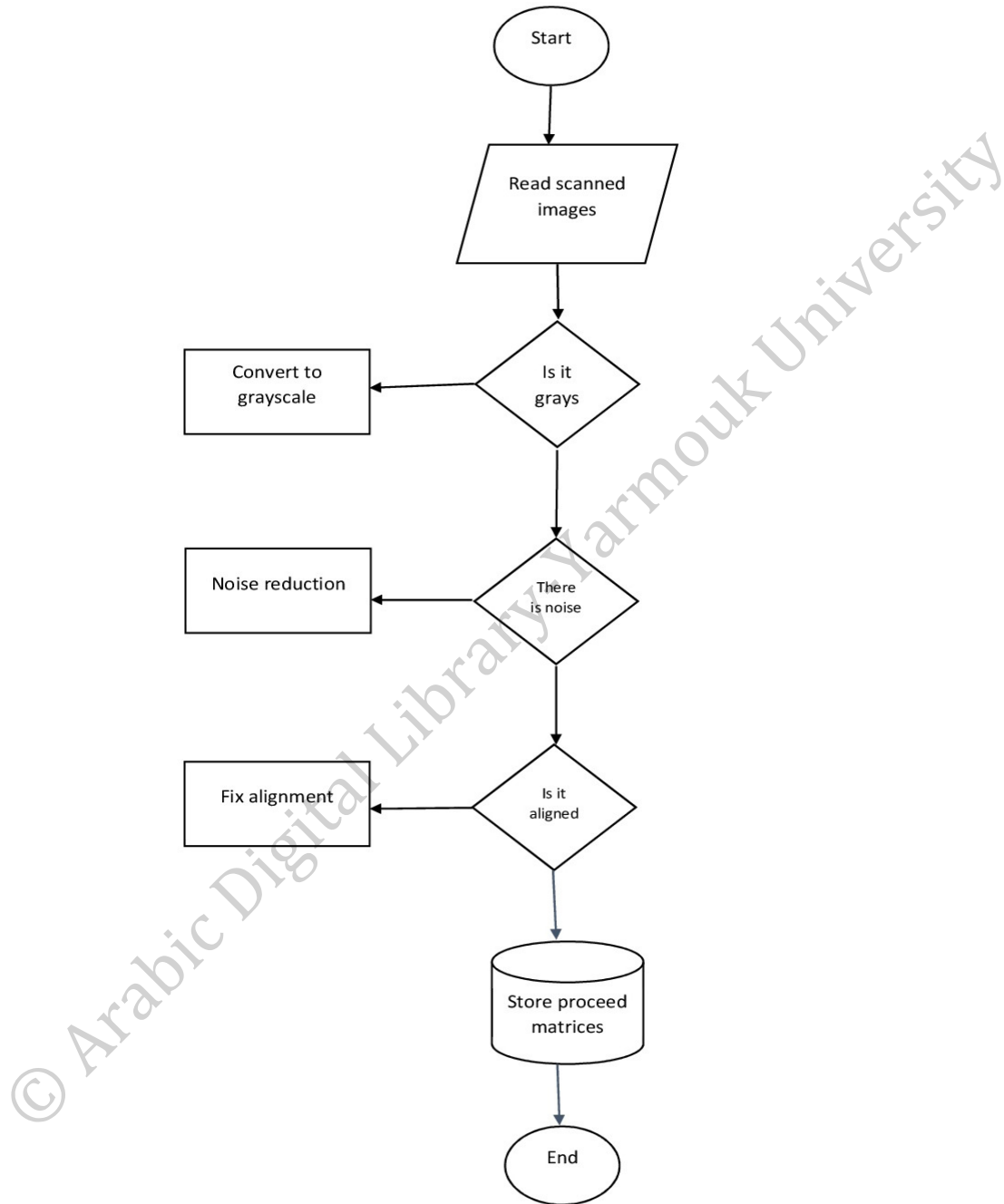
Figure 18: Preparation stage procedure.

The Figure 18 shows the flowchart for the preprocessing stage.

33

Processing Stage: At this stage main objects are detected and handwritten pattern is recognized.

Edge detection by using Sobel edge detection, the edge occurs when the intensity function cut out when has a very gradient steep intensity at the image. That's mean there are two pixels for every one-pixel wide line: one going into edge (black to white) and one coming out of edge (white to black). [24]

Gradient component found using the Equf (12) and Equ (13) [24].

$$\frac{\partial f(x,y)}{\partial x} = \Delta x = \frac{f(x+dx,y) - f(x,y)}{dx} \qquad (12)$$

$$\frac{\partial f(x,y)}{\partial x} = \Delta y = \frac{f(x,y+dy) - f(x,y)}{dy} \qquad (13)$$

Where dx measure the distance along the x-direction horizontal answer table lines and dy measure the distance along the y-direction vertical answer table lines.

The output Sobel function – BW=edge (gray input, 'Sobel') - is binary image.

- Line detection by using Hough transform, where the output binary image of Sobel edge detection or an edge detected image is the input of Hough transform. The MATLAB function Hough (BW) generates a 2D histogram, the first axis the angle θ of a line from the origin and the other axis is the distance from the origin [21].

$$r = x \cos\theta + y \sin\theta \qquad (14)$$

- Gridding the answer table to answer boxes after calling Hough ( ) function twice, first to find vertical line distances and second to find horizontal line distances. By using crop MATLAB function.
- Flag the boxes if it's red marked from the manual teacher grading, to distinguish teacher grading mark.
- Check if the box is marked by student, by applying thresholding.
- Check the question is single or multiple marked, and flag it if its multiple marked.

- Applying the four voting system to recognize which mark is the winner.
- The voting system consist of four functions measuring similarity of the multiple marked question under investigation to the single marked questions which traded as reference template for approved answer mark. Comparing student mark with its last mark increase processing speed and confidence.

Voting System for Scoring

This procedure consists of four voting methods to handle all special cases (i.e. revoked answer, multiple answers). Each marked QABoxes matched with all single marked QABoxes are averaged to get the maximum similarity and to detect the final student answer for scoring.

The proposed voting system applies the most four used methods for template matching and similarity measurement, and returns four measurement results to vote for student's approved answer.

1. Maximum Normalized cross correlation voting system (CCF)

The two terms convolution and cross-correlation are implemented in a very similar way (convolution is correlation with different order of one sequence forward correlation backward convolution, so using MATLAB function conv2(QueryBox,TemplateBox,'same') is like correlation in 2D), but cross correlation is used for similarity measuring of two series , also known as a sliding dot product. A block is multiplied by another block pixel-wise and summed.

The cross-correlation is defined as [22].

$$(f * g)[n] \stackrel{\text{def}}{=} \int_{m=\infty-}^{\infty} f * [m]g(m+n) \tag{15}$$

Where f is the mark under investigation (one of multiple answers) and g is reference pattern of single marked question. Sliding over f pixel by pixel to measure similarity.

2. **RAW Mean Square Energy (MSE) [18].**

$$MSE = \sum Current\ sum(BW) - singlmark\ sum(BW)^2 \qquad (16)$$

3. **RAW Minimum Absolute Difference (MAD)**

MSE and MAD stand for Mean Square Energy and Minimum Absolute Difference.

The MSE /MAD is for the sum of BW pixels, Sum (BW) of mark under investigation is compared with Sum (BW)) of reference pattern of single marked question [18].

$$MAD = \sum ABS|(Current\ sum(BW) - singlmark\ sum(BW)) \qquad (17)$$

4. **Minimum shifted difference voting system (DIF)**

For measuring similarity of two images calculate difference between each pixel in the mark under investigation block image and the corresponding pixel in the reference pattern of single marked question block image. Two blocks are subtracted (one from the other). This is done for various relative shifts (up/down & left/right).

All selected measurements are mathematically computed based to find the similarity degree between the possible student's answer patterns, our reference is the student's single-marked questions to decide the most likely intended answer and apply scoring procedure on it.
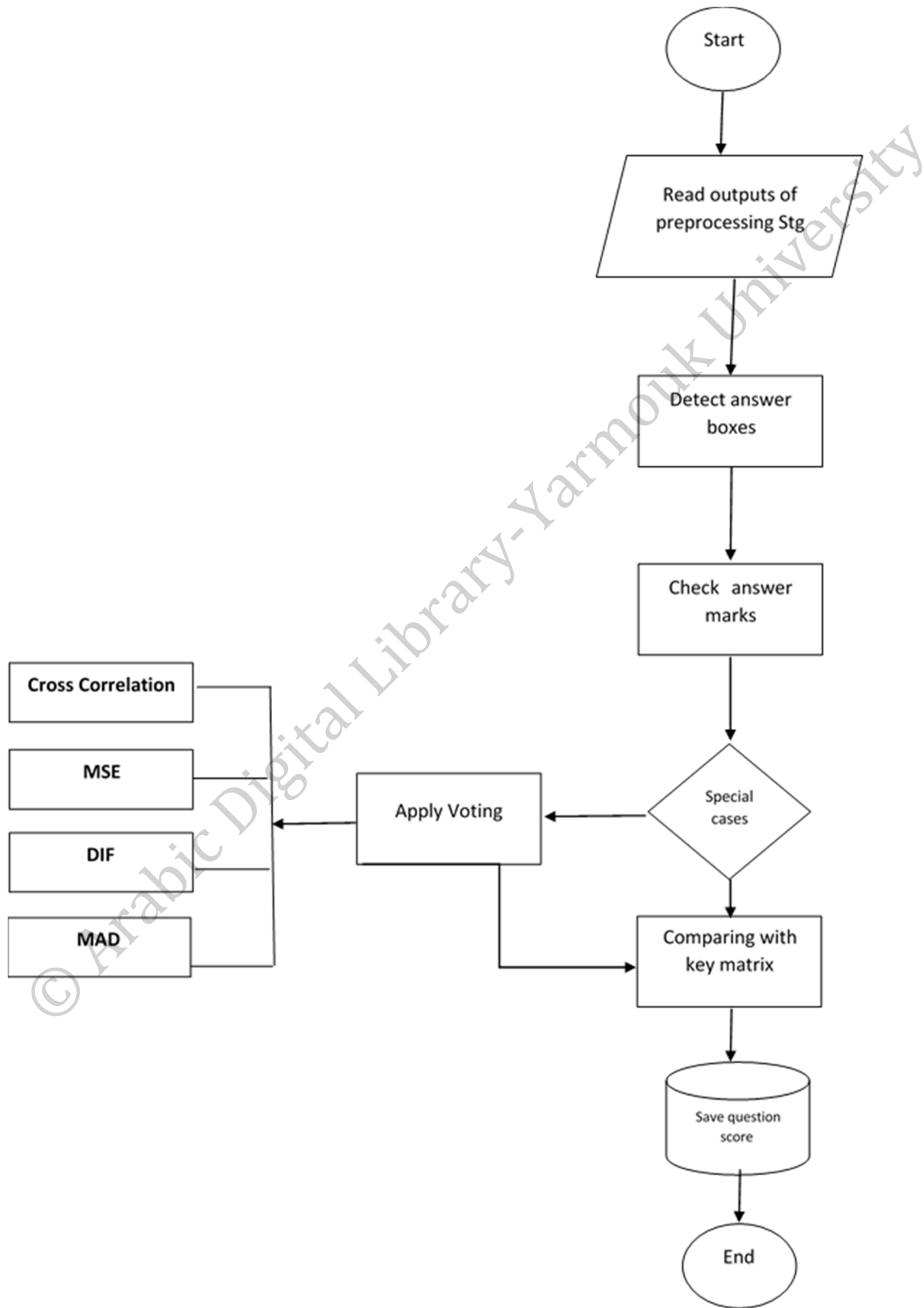
Figure 19: Processing stage procedure.

Scoring Stage: after preparing all required matrices the final scoring process is done at this stage to return all students marks to excel sheet as shown in Figure 20.
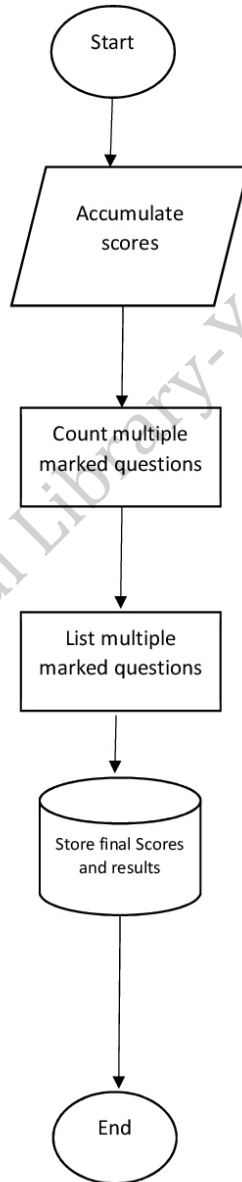

Figure 20: Scoring stage procedure.

38

CHAPTER 5

Experimental Results and Analysis

This Chapter Provide some running environment details and experimental results to evaluate the true potential of the automated multiple-choice scoring application, proposed application had been implemented in MATLAB R2015.

5.1 Designing and Running Platform

This section presents all used tools for developments and simulation of the proposed application.

5.1.1 Tools and Software Used in this Project

The list of software and simulation tools used in our work is:

- General MATLAB Toolbox
- Image Processing Toolbox

5.1.2 Filtering Techniques

The proposed algorithm used the below filtering and similarity measurements techniques at the pre-processing and processing stages.

- Sobel
- Hough Transform
- Convolution
- Morphological

5.1.3 Platform Specifications

This work was developed and simulated using a personal laptop with the following specifications:

- Operating System: Windows 10 Enterprise
- Processor: Intel® Core i7-7500U CPU @ 2.7 GHz, 2.9 GHz
- RAM: 16 GB
- System Type: 64-bit Operating System, x64-based processor

5.2 Images Dataset

Testing and validation were done using several scanned images as input samples, input images are collected randomly which are scanned using several scanners to have variant image's resolutions, sizes and color modes.

Image scanning done by any scanner with low resolution of 200 dpi. while the original image doesn't have much gradient in color between the white background and black table lines, and it will be enhanced to binary one by the histogram equalization.

5.3 Samples of Handwritten Answer's Patterns

The main challenge in this work is to know how the student marks the correct answer in the answer table, and how we can detect if the student changes his/her answer. As shown in Figure(21,22), there are many possible answering patterns by students like alphabetical small and capital characters, √, ×, and other unexpected handwritten symbols.

40
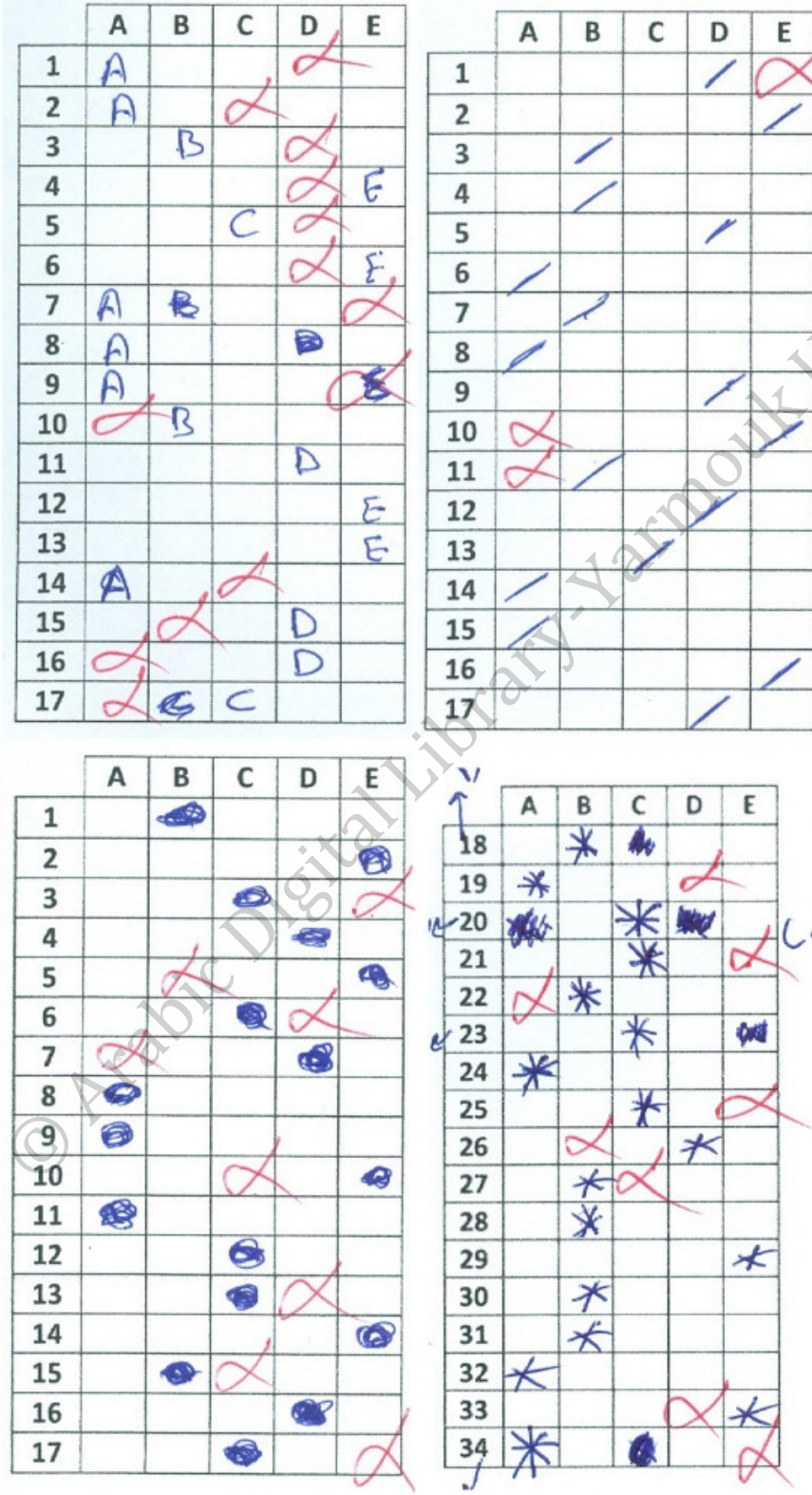
Figure 21: Examples of used answering patterns

Figure 22: More examples of used answering patterns.

5.4 Proposed Work flow

The proposed workflow involves the following steps:

1. Design and print answer-sheet templates with custom format using inkjet or laser printers with A4 paper size

2. Copy the answer-sheets using photocopiers for all students.

3. Collect the answer-sheets after exam.

4. Scan the collected answer-sheets using off-the-shelf scanner and save them as individual images.

5. Folding the images according to exam form with the form key answer sheet, where every exam has four forms (A, B, C and E).

6. Store all scanned images at the input folder path

7. The answers-sheets data are imported into matrices as well as the key answer-sheet for scoring.

8. Run the proposed application

9. Export scoring results into spreadsheet with annotated marks

5.5 Experimental Results

The performance factors, including execution time, CPU time, accuracy and power, are discussed in this section to show the efficiency of the proposed approach by providing experimental results of different scanned answer-sheets images.

5.5.1 Execution time results

The table below is a summary for time needed to process different samples of real input data.

Where *Tic/Toc* is the elapsed time required to run local functions, *CPU-Time* is the total execution time and *clock time* is the time measurement unit (internal system time measurement).

Table 4: Experimental Results

Image Size	Tic/Tocs	CPUtimes	Clock Time
778 KB	6.1122	6.42188	6.105
688 KB	6.40419	6.51563	6.405
491 KB	5.83147	5.875	5.828
14 images	34.57	36.42	34.62
50 images	310.7	317.5	311.9

5.5.2 Accuracy rate of results

The real data-set of more than 800 student's exam sheets used for testing the system. These real exams held at Yarmouk University, and taken over the period of the years 2011 to 2016, covering 9 semesters and 5 main courses of computer engineering department. That's improve the system by processing various images patterns of handwritten marks, hard special cases for multi marked samples. Different color-modes, and noise types.

Every exam sheet consists of 50 questions that's mean we have (800*50=40000) pattern to recognize. Our system solved 791 sheet and get right final score comparing to manual grading. Only 9 sheets get wrong result with (77 questions)

- Accuracy rate of sheets = 791/800%=98.875% .
- Accuracy rate of questions = 39923/40000=99.805% .

44

Table 5: Accuracy Rates

Method	Accuracy Rate	Notes
Optical Mark Recognition OMR [5]	100% (Conditionally)	*The right choice should be perfectly filled by the examinee with pencil ;No revoking ;Unaffordable for many institutes Not flexible*
MCQ exam scanned paper grading [12]	100% (Conditionally)	*Select only one choice; No revoking* *Darken one the right cell perfectly*
Eyegrade [16]	97%	*Answering restriction; answer with cross mark match cell corners only.* *Revoke by darken the cell; capturing images by webcam weak table detection.*
FreeFormScanner[9]	96.5%	*Using simple OCR ;OCR limitation* *Answer by darken answer bubble; have very restricted instruction for revoking.*
Method in reference[14]	95%	*Using simple OCR ;OCR limitation* *Only capital letters are accepted*
Our Scoring System	99,8	*Answers out of answer box;*

The reasons of wrong results summarized in:

- Writing answer out of answer box.
- Marking more than one cell by same symbols and didn't revoke one of them
- Using red lines for answering.
- Revoke all answers.

45

5.6 The Followed Steps

This section presents the main stages followed in the automated multiple-choice scoring system running
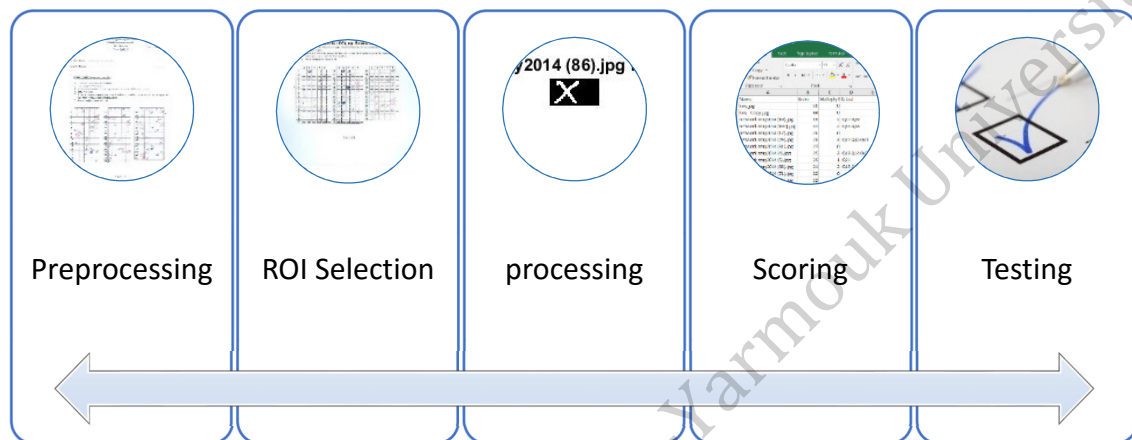
Figure 23: Application running Steps.

Step 1: Preprocessing stage

At the pre-processing stage, some of preprocessing techniques are applied to handle image enhancement, noise reduction, thresholding, skew detection and correction, define segmentation factors, extract important features, to produce suitable images in grayscale mode for next processing stages.

Figure24 shows a sample of a student answers sheet which is graded manually by the instructor. After that, the marks made by the instructor and flagged (to be masked) during the automatic grading.

www.manaraa.com

Figure 24: Original Test Image.

Figure 25 shows the answer key image after conversion of the Input image (colored image) to grayscale.
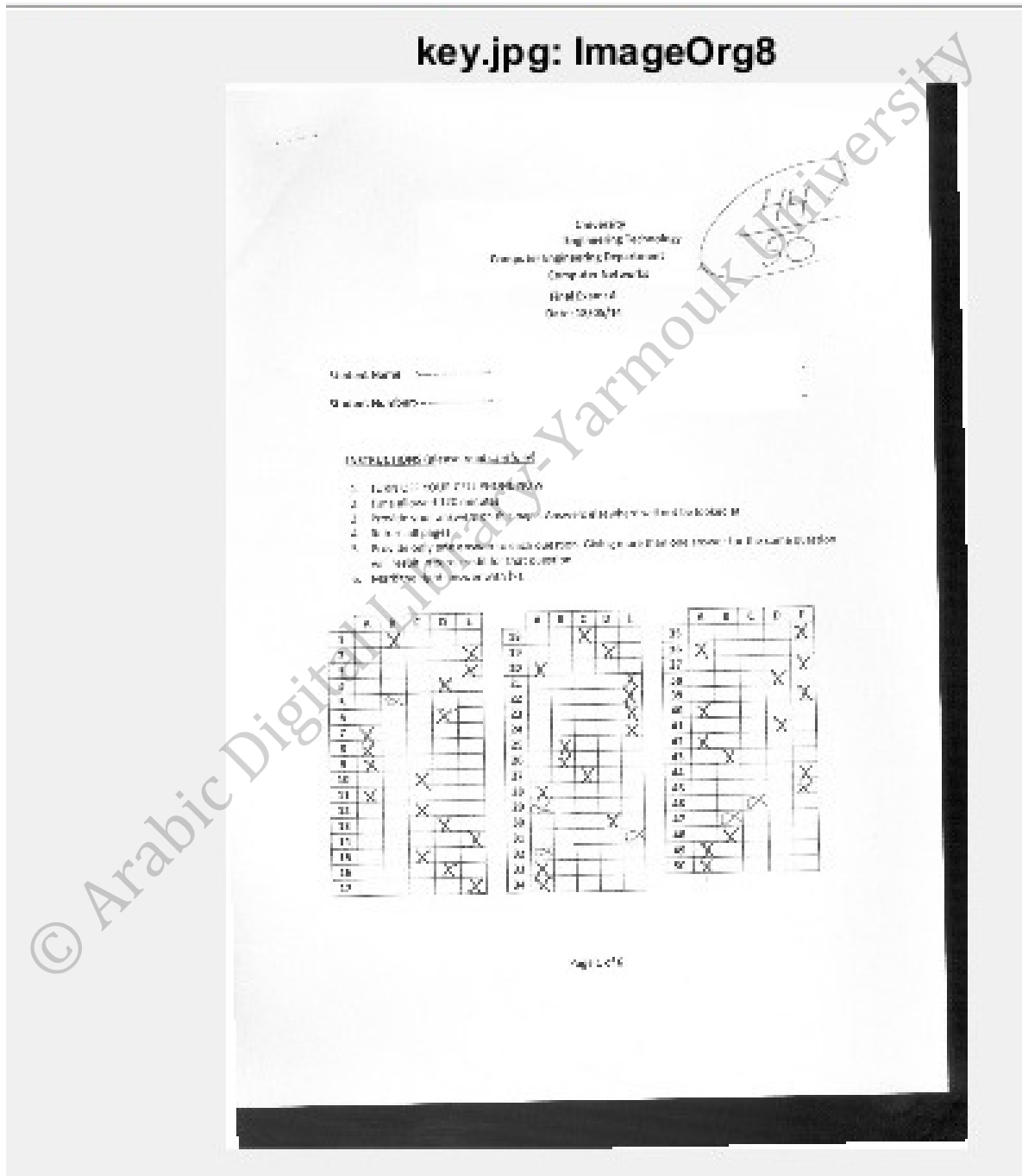

Figure 25: Image Converted To Gray Scale

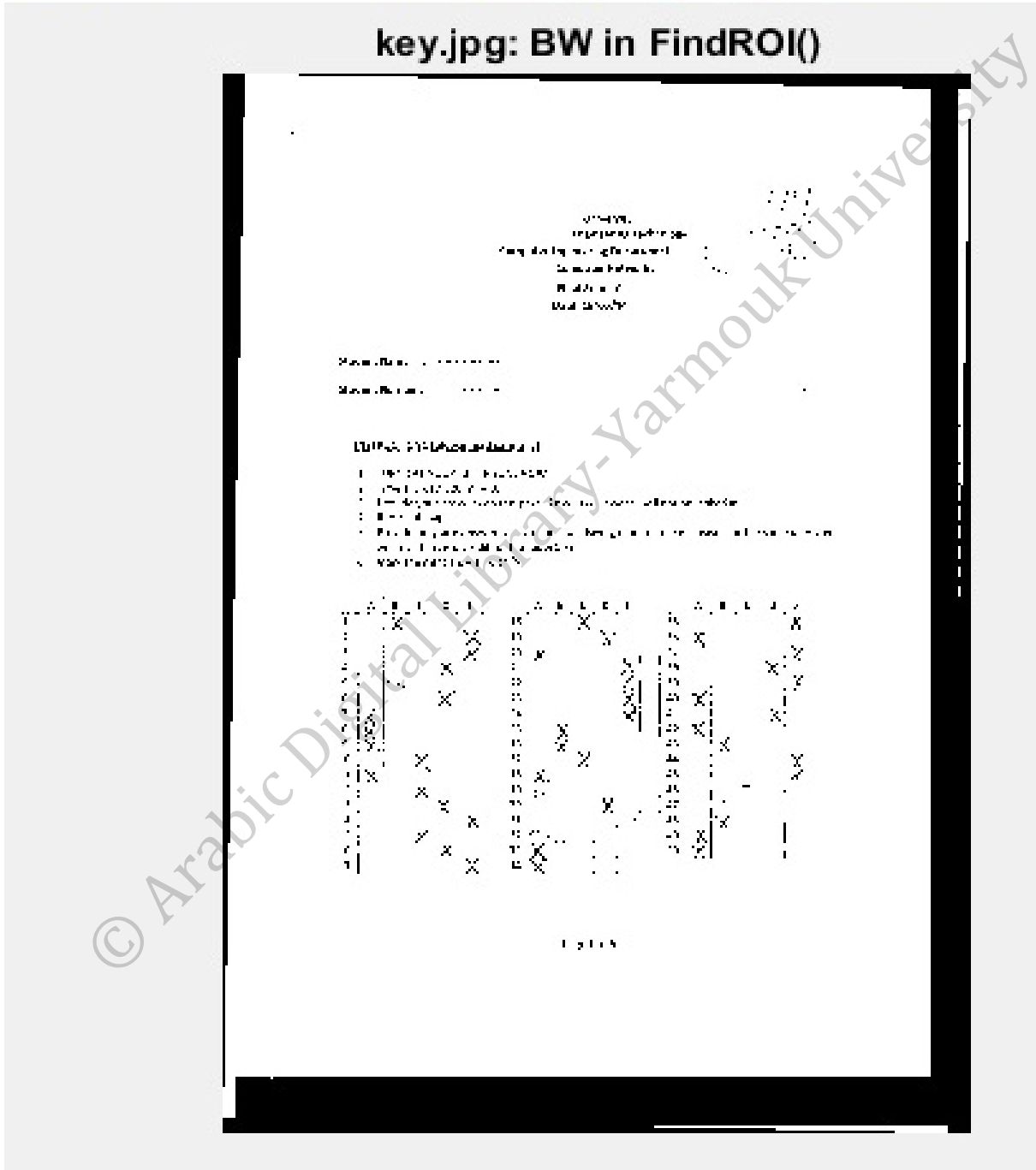Figure 26 shows the aligned image, by angle θ, which found by using Hough transform.

Figure 26: Rotation angle image adjustment.

Step 2: Region of Interest Selection Stage

The proposed approach detects the most important part in the preprocessed image and assigns it as a region of interest in the proceed image.

As shown in the algorithm, the proposed application scans input image from left to right; top to bottom; right to left; bottom to top (raster scan) to find 98% of that original middle pixel count and assigned it as a region of interest, which is a half way in the image vertical and horizontal paths.

In addition, as shown in Figure27 we defined new margins to skip the Exam-Instructions part 0.4*yN .
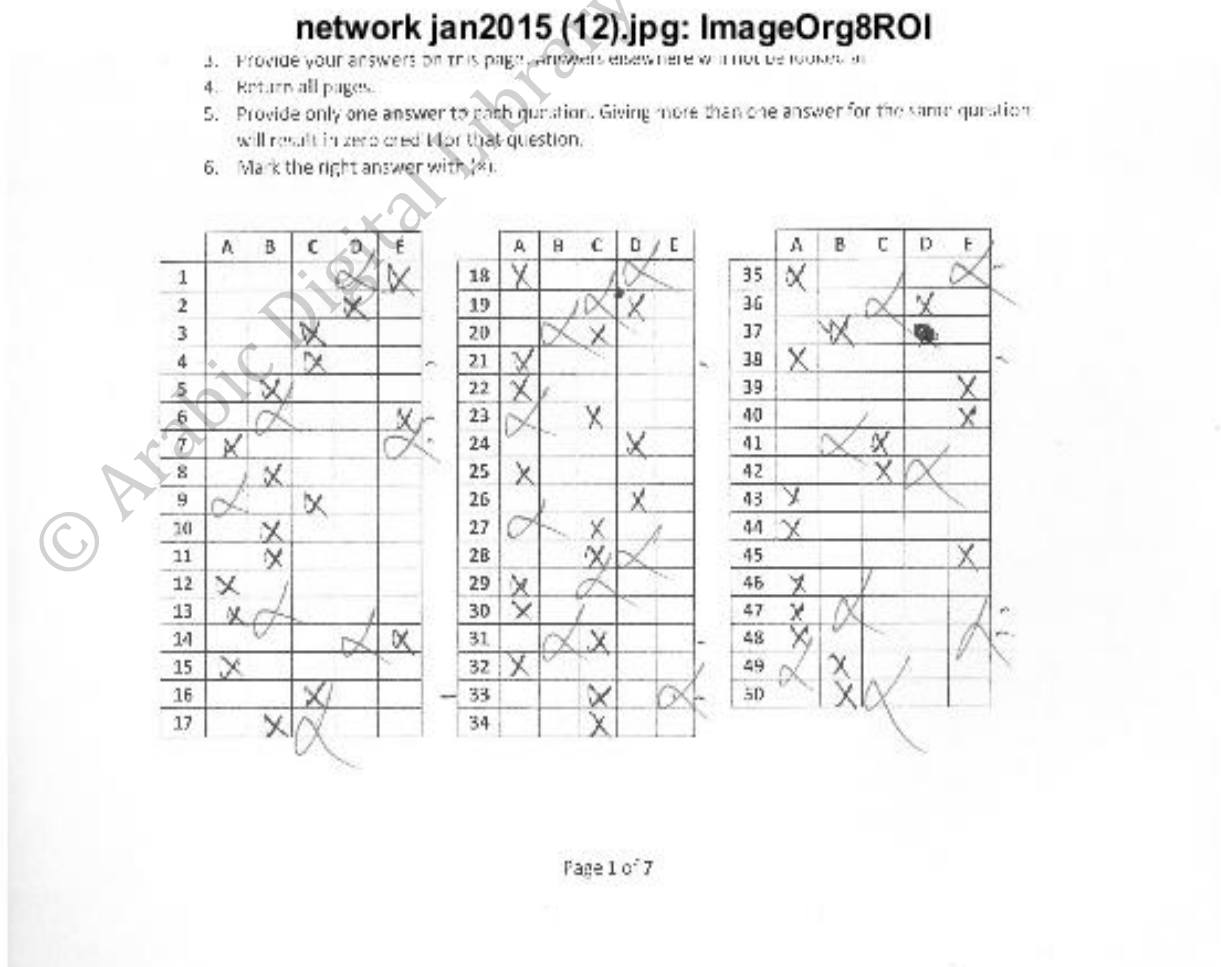
Figure 27 shows the cropped image with ROI

Figure 27: Cropped image (Region of interest).

Step 3: Segmentation and Localization

This stage consists of two processing layers. Firstly, we applied table detection techniques to detect the two or three main Answer's tables and answer's marks.

Sobel edge detection and Hough transform were applied together for edge detection and to find important image features for segmentation for Hough Transform, Theta Histogram and RhoHistogram, winV & winH, maximum and minimum Peaks were obtained, this was procedure applied twice for rotated (ignore theta) and un-rotated image to find exact x and y location for QABoxes. The whole histogram covers all possible pairings of lines at particular points (Rho, Theta); each element in the histogram is one pair of Rho, Theta.

Every pair describes a line (angle from x -axis at rho distance from the origin) according to the pixel coordinate system as shown in Figure28.
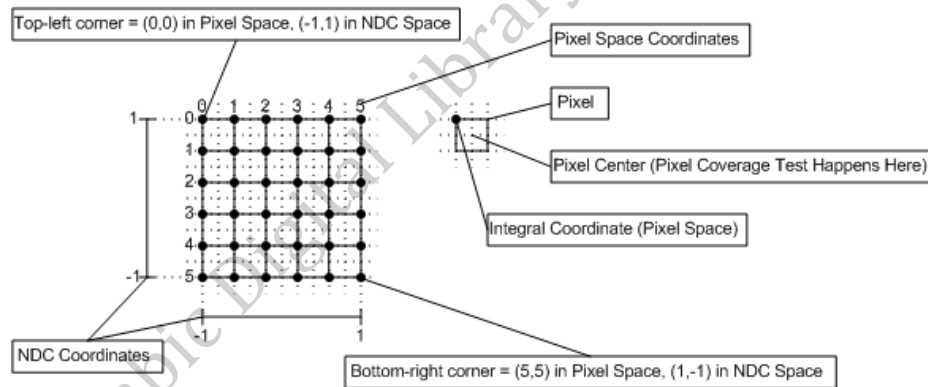

Figure 28: Pixel presentation at polar system.

The next step is finding a 2-D histogram as values for every Rho, Theta of 22 vertical and 19 horizontal input peaks. This illustrated by the Figure29.
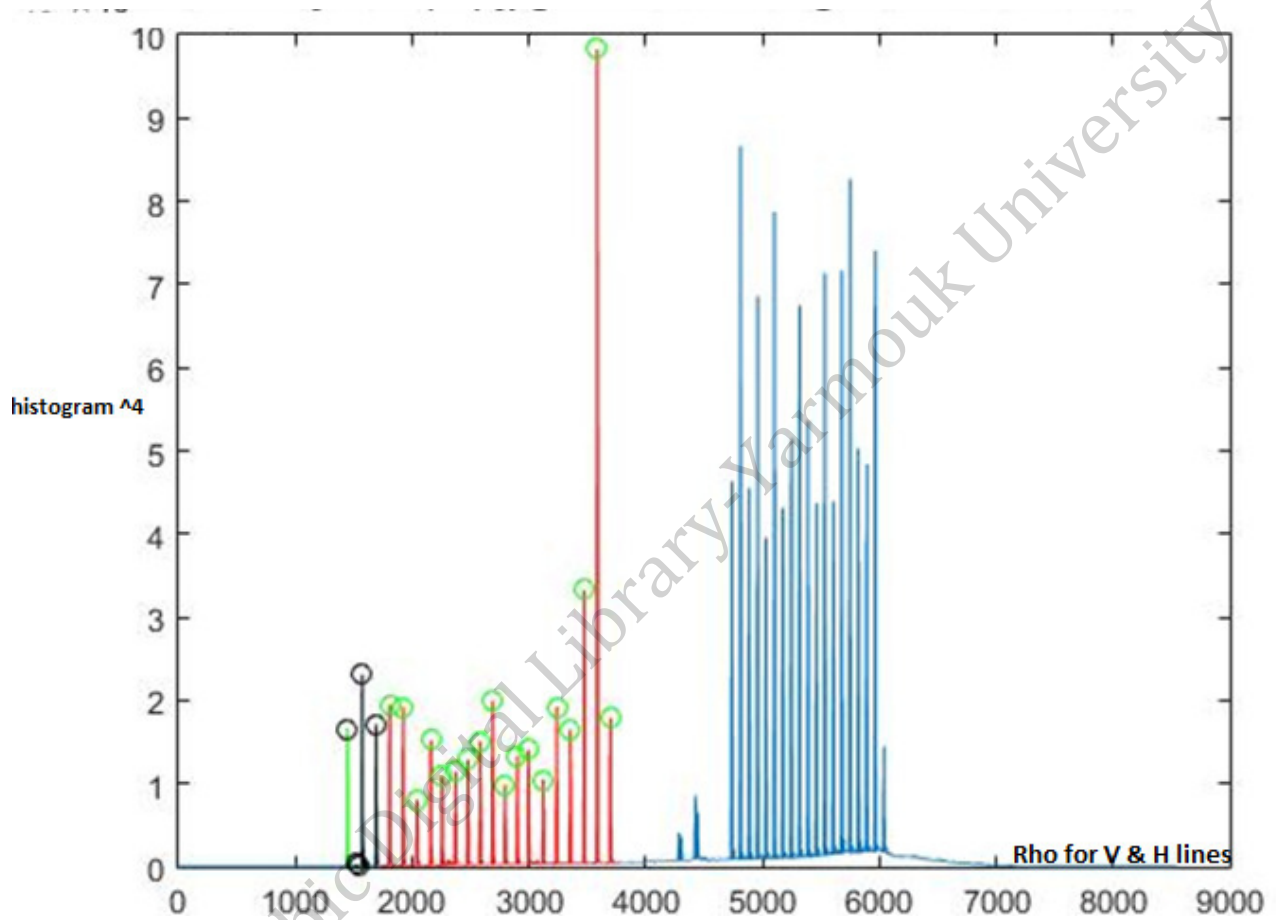
Figure 29 Find the histogram peaks average.

Each pair of points counts as a set of rho, theta pairings and Rho's give winV and winH and the Left and Top of the answer sheets. All these features represent localization mechanism in our application to find answer boxes and store student's answers into addressed matrix for each input image as well as for the answer key image. The results shown in Figure30.

© Arabic Digital Library-Yarmouk University

www.manaraa.com

Figure 30; finding answer boxes by edges and lines detection.

Step 4: Scoring Stage

For scoring issue, all QABoxes checked per question with the below possible cases:

- Non-Marked QABoxes (as shown in Question 9 of Figure31).

- Single-Marked QABox (as shown in Question 10 of Figure31).

- Multi-Marked QABoxes (as shown in Question 11 of Figure31).

Figure 31: QABoxes checking cases.

Firstly, we check the correct answer location if it is non-marked QABox, then the question-score=0. Else, we ignore all empty QABoxes from processing. Single-Marked QABoxes are flagged as a singly marked questions and checked if correct or not in comparison with the answer key matrix. When the application finds more than one QABoxes are marked and the correct answer one of them. -as seen in figures (32-37) Assigned this question is assigned as Multi-Marked QABoxes to be handled it using the proposed voting system to determine the most likely correct answer.

network jan2015 (12).jpg Raw Q37 A1

Figure 32: Cropped image to box of Q37, choice A.

54

network jan2015 (12).jpg Raw Q37 A2

Figure 33: Cropped image to box of Q37, choice B.

network jan2015 (12).jpg Raw Q37 A3

Figure 34: Cropped image to box of Q37, choice C.

network jan2015 (12).jpg Raw Q37 A3

Figure 35: Cropped image to box of Q37, choice D.

network jan2015 (12).jpg Raw Q37 A5
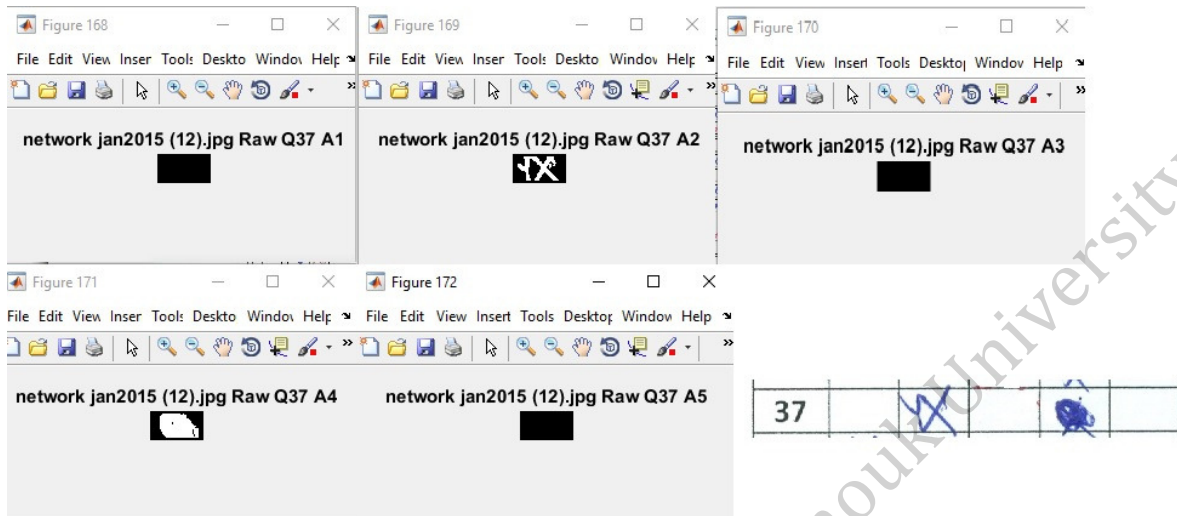

Figure 36: Cropped image to box of Q37, choice E.

Figure 37: Handling Q37, which has multiple answers.

Voting System for Scoring

This procedure consists of four voting methods to handle all special cases (i.e. revoked answer, multiple answers). Each marked QABoxes matched with all single marked QABoxes are avareged to get the maximum similarity and to detect the final student answer for scoring.

The proposed voting system applies the most four used methods for template matching and similarity measurement, and returns four measurement results to vote for student's approved answer.

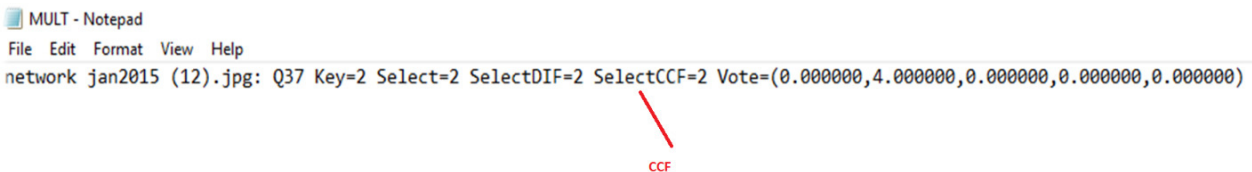1. **Maximum Normalized cross correlation voting system (CCF)**

Figure 38: Cross correlation voting system CCF result for Q37.

The two terms convolution and cross-correlation are implemented in a very similar way, convolution is correlation with different order of one sequence forward correlation backward convolution, so using MATLAB function conv2(QueryBox,TemplateBox,'same') is like correlation in 2D) ,but cross correlation is used for similarity measuring of two series , also known as a sliding dot product. A block is multiplied by another block pixel-wise and summed.

The CCF results in CCF.txt file as seen in figure38.
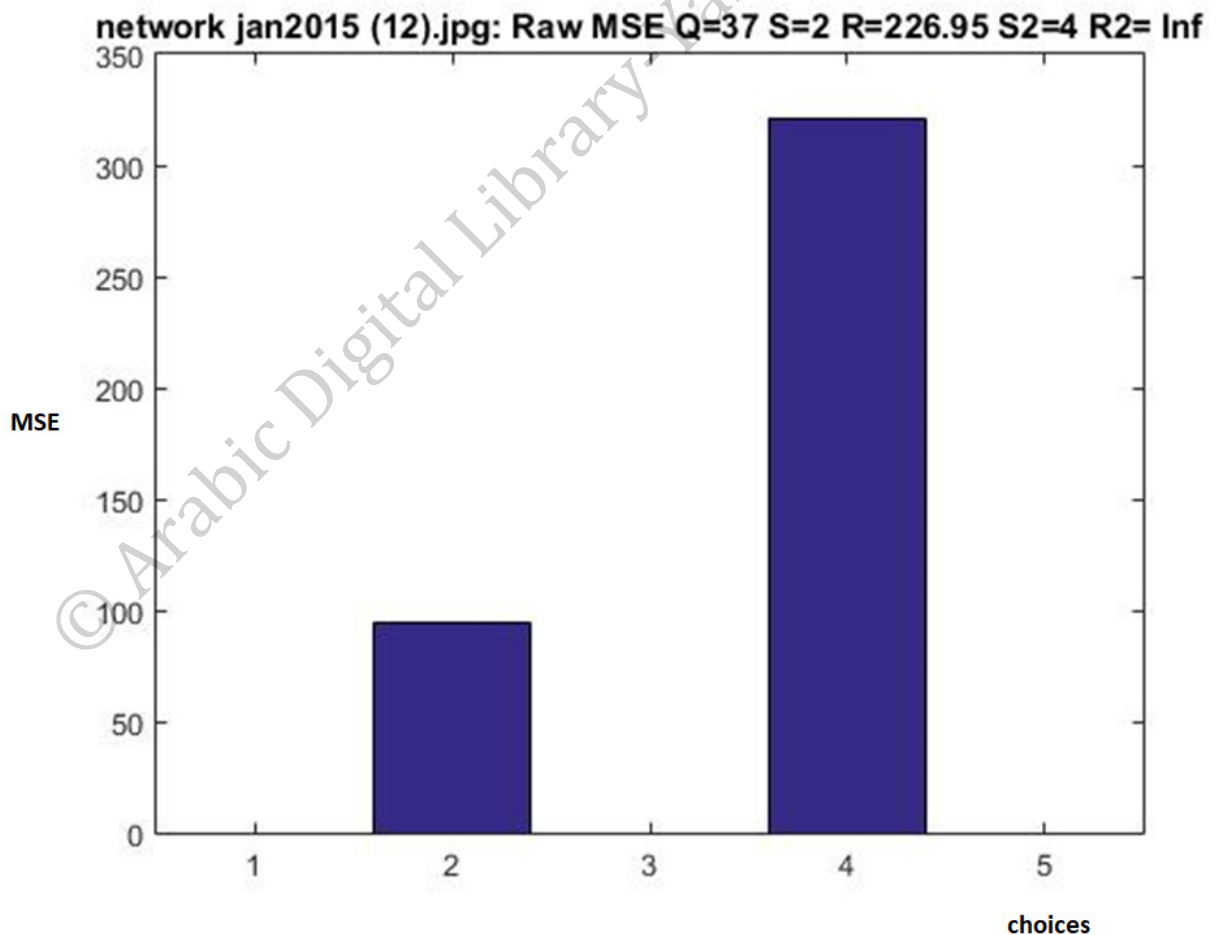
2. RAW Mean Square Energy (MSE)

Figure 39: MSE result for Q37.

3. RAW Minimum Absolute Difference (MAD)

MSE and MAD stand for Mean Square Energy and Minimum Absolute Difference. The MSE /MAD is for the sum of BW pixels, Sum (BW) of mark under investigation is compared with Sum (BW)) of reference pattern of single marked question.
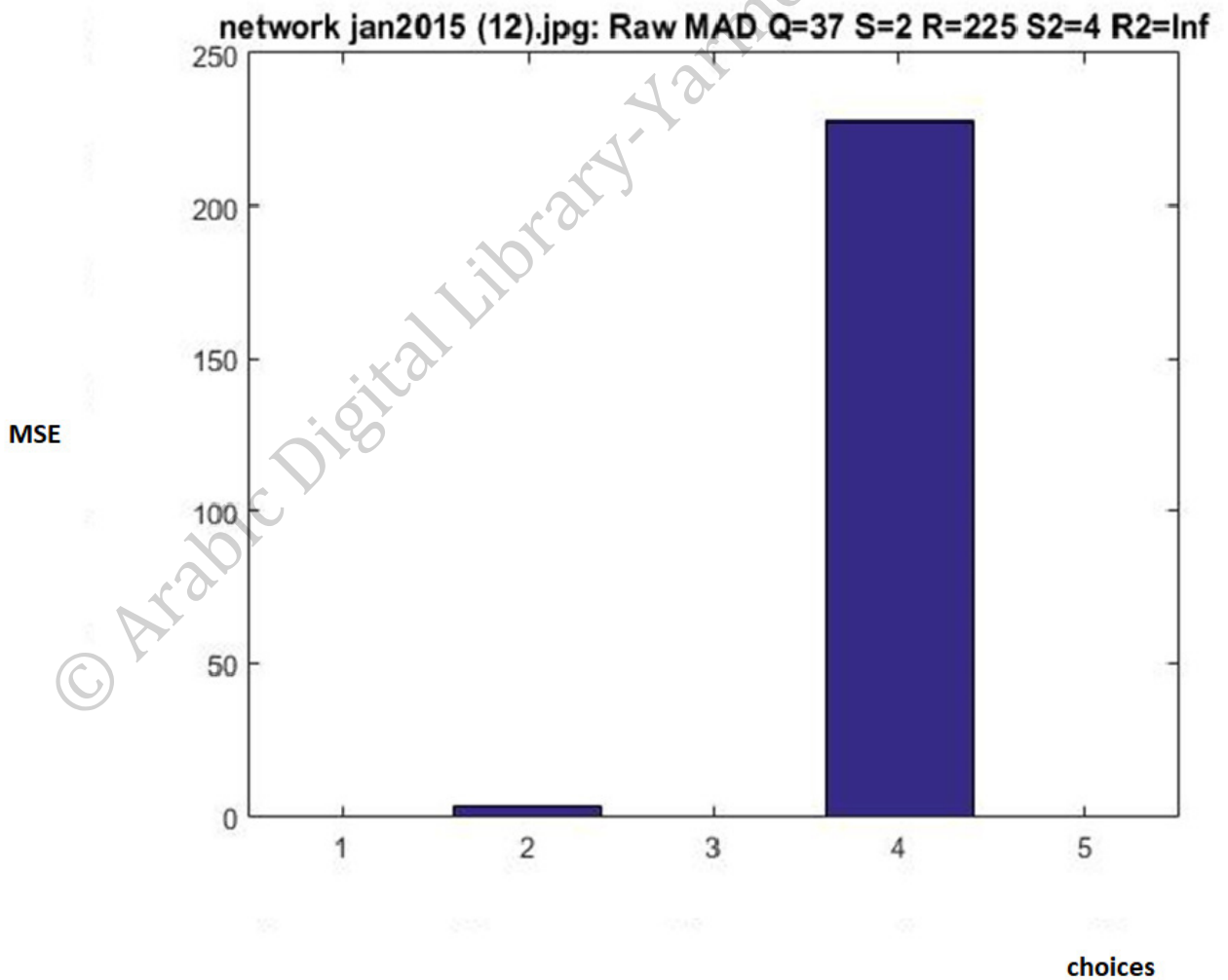

Figure 40: MAD result for Q37.

4. Minimum shifted difference voting system (DIF)


```
MULT - Notepad
File  Edit  Format  View  Help
network jan2015 (12).jpg: Q37 Key=2 Select=2 SelectDIF=2 SelectCCF=2 Vote=(0.000000,4.000000,0.000000,0.000000,0.000000)
                                    Shifted Difference
```
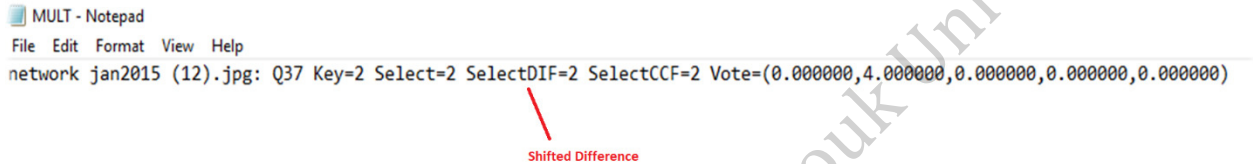
Figure 41: DIF result for Q37.

For measuring similarity of two images calculate difference between each pixel in the mark under investigation block image and the corresponding pixel in the reference pattern of single marked question block image. Two blocks are subtracted (one from the other). This is done for various relative shifts (up/down & left/right). All selected measurements are mathematically computed based to find the similarity degree between the possible student's answer patterns, our reference is the student's single-marked questions to decide the most likely intended answer and apply scoring procedure on it. Figure43 shows the voting result using DIF, CCF, MSE and MAD methods applied to question 29 where the third choice is the winner.
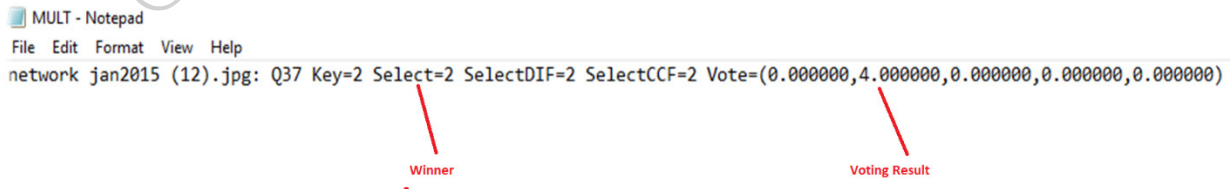


```
MULT - Notepad
File  Edit  Format  View  Help
network jan2015 (12).jpg: Q37 Key=2 Select=2 SelectDIF=2 SelectCCF=2 Vote=(0.000000,4.000000,0.000000,0.000000,0.000000)
                                    Winner                              Voting Result
```

Figure 42: Applying the voting system the winner is the 2nd choice

59

At the end of this stage, matrices are produced and compared with key answer matrix to find similarity, where all special cases are solved by the voting mechanism to accumulate the final scores. Moreover, the multiple marked question is counted and listed. Finally, the final scores are exported into excel sheet as shown in Figure44.

	A	B	C	D
1	Name	Score	Multiply Marked	Q-List
2	key.jpg	50	0	
3	network jan2015 (12).jpg	26	1	Q37
4				
5				
6				
7				

Figure 43: Final score of the student answer sheet.

	A	B	C	D	E
1	Name	Score	Multiply Marked	Q-List	
2	key.jpg	50	0		
3	key - Copy.jpg	36	0		
4	network jan2015 (1).jpg	20	1	Q29	
5	network jan2015 (10).jpg	21	3	Q29 Q32 Q40	
6	network jan2015 (11).jpg	24	1	Q13	
7	network jan2015 (12).jpg	26	1	Q37	
8	network jan2015 (13).jpg	13	1	Q16	
9	network jan2015 (14).jpg	32	0		
10	network jan2015 (15).jpg	17	1	Q42	
11	network jan2015 (16).jpg	40	1	Q45	
12	network jan2015 (17).jpg	35	2	Q13 Q27	
13	network jan2015 (18).jpg	20	0		
14	network jan2015 (19).jpg	36	0		
15	network jan2015 (2).jpg	17	0		
16	network jan2015 (20).jpg	18	1	Q46	
17	network jan2015 (21).jpg	33	1	Q6	
18	network jan2015 (22).jpg	10	0		
19	network jan2015 (23).jpg	22	0		
20	network jan2015 (3).jpg	31	1	Q36	
21	network jan2015 (4).jpg	36	0		
22	network jan2015 (5).jpg	31	0		
23	network jan2015 (6).jpg	25	0		
24	network jan2015 (7).jpg	25	0		
25	network jan2015 (8).jpg	35	0		

Figure 44: Final score of class answer sheets.

61

Testing Stage

All automated scoring results, where compared with manual results and special cases were tested to validate proposed application reliability and accuracy.

5.7 Case Study:

5.7.1 Examples of solved cases
The figures below show many cases the system could deal with it and get accurate final results.

- Figure 46 example where student use alphabet letters for answering.
- Figure 47 example where student used star mark for answering and just darken center of the star for revoke.
- Figure 48 example where student stroke over all answering cells with light font.
- Figure 49 example where student write many marks beside the answer cells but outside answer matrix, and the system ignored them.
- Figure 50 example where no big difference between right and wrong mark.
- Figure 51 example where student fill the paper with many notes.
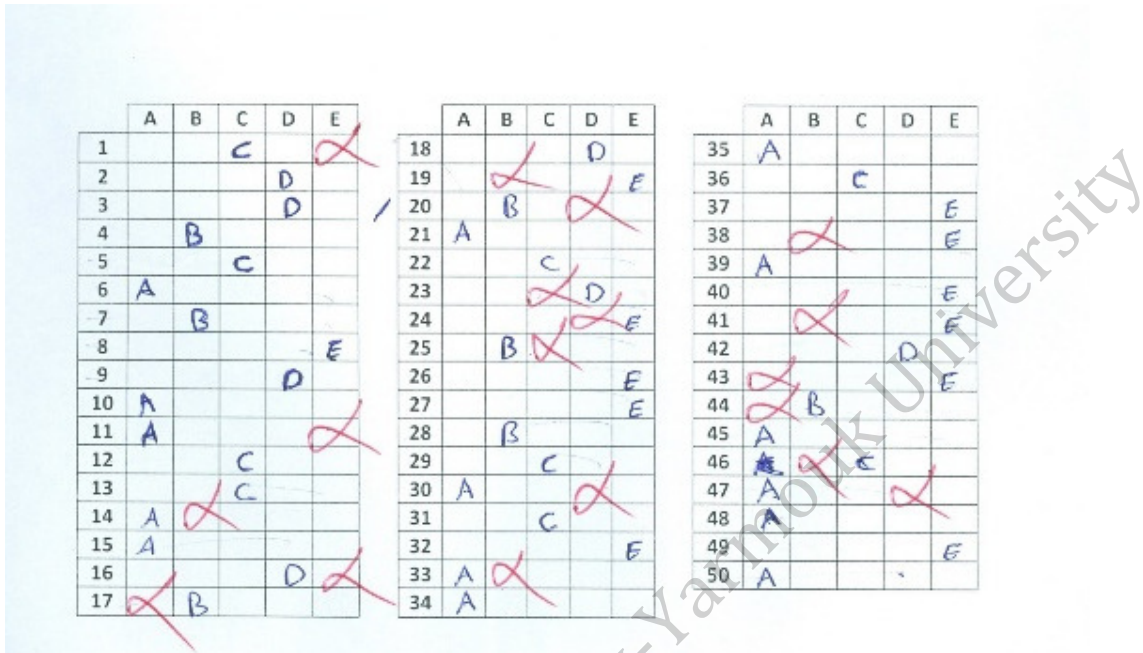
Figure 45: Example 1 of special cases.
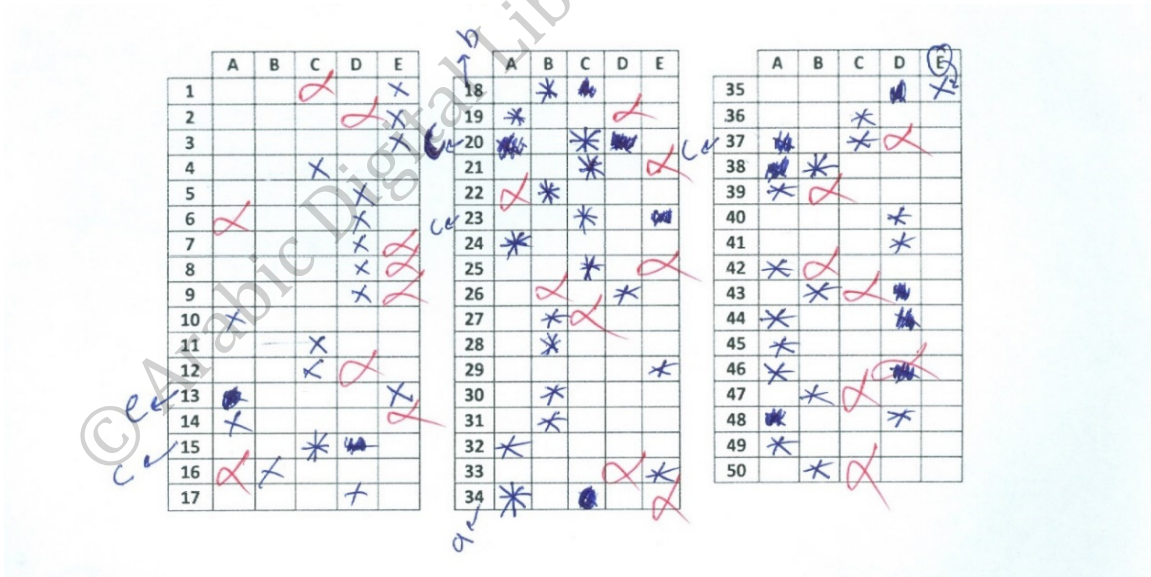
Figure 46: Example 2 of special cases.

Figure 47: Example 3 of special cases.

INSTRUCTIONS (please read carefully)

1. TURN OFF YOUR CELL PHONE **NOW**
2. Time allowed 120 minutes
3. Provide your answers on this page. Answers elsewhere will not be looked at.
4. Return all pages.
5. Provide only **one answer** to each question. Giving more than one answer for the same question will result in zero credit for that question.
6. Mark the right answer with (×).

Figure 48: Example 4 of special cases.

Figure 49: Example 5 of special cases.

Page 1 of 6

Figure 50: Example 6 of special cases.

5.7.2 Examples of not solved cases:

The figures below show many cases the system couldn't deal with it and get wrong result , but the system mention the Question number which its result not 100% accurate in the result excel sheet.

Figure 52 example where student stroke many answering cells.

Figure 53 example where student answer out of table cells.

Figure 54 example where student answer with red lines.

Figure 55 example where student answer out of table cells.

Figure 56 example where student get extra mark where the right answer is C (which is annotated) while he writes out of box his approved answer D.

Figure 51: Example 1 of not solved special cases.

Figure 52: Example 2 of not solved special cases.

6. Mark the right answer with (×).

Figure 53: Example 3 of not solved special cases.
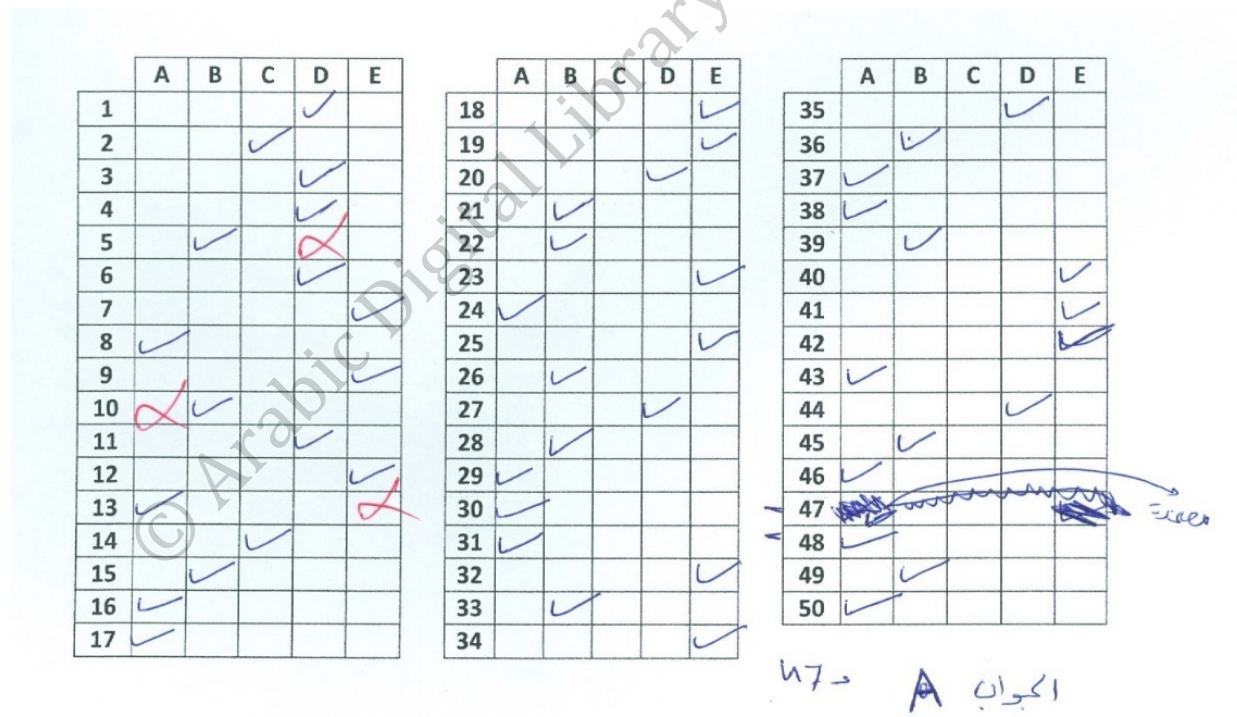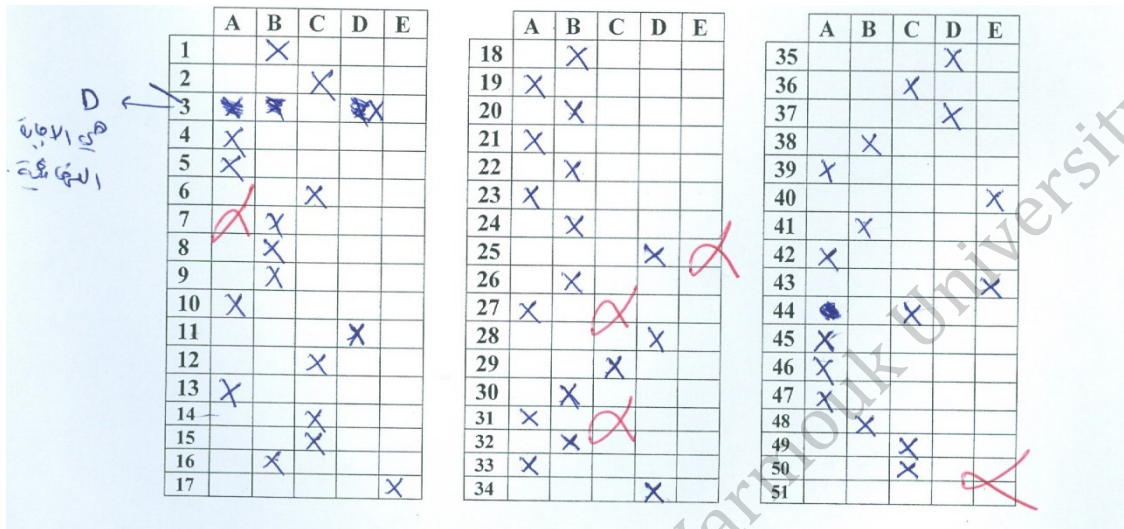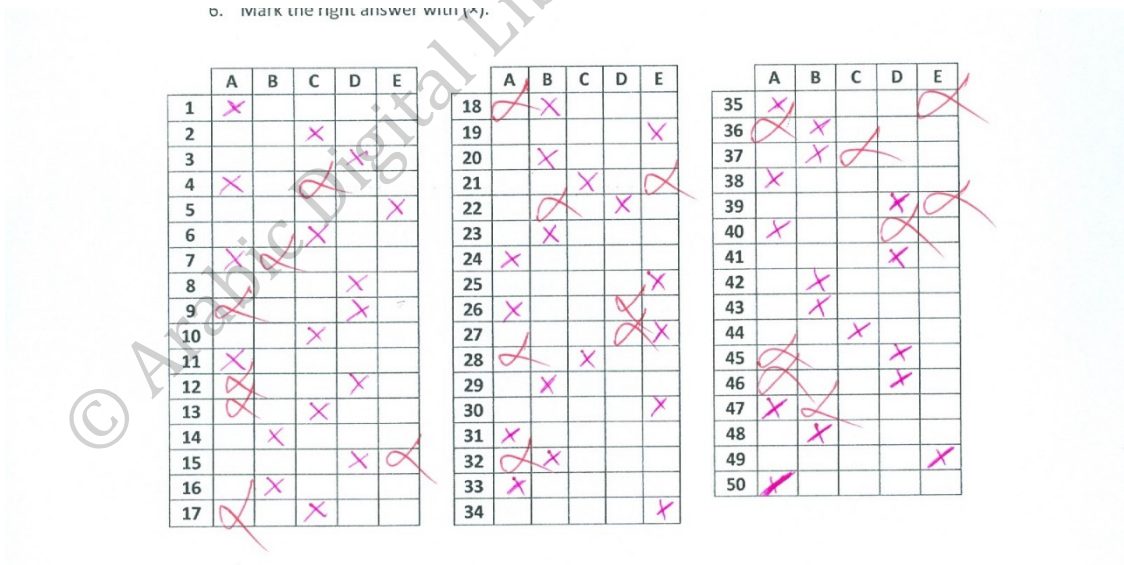
67

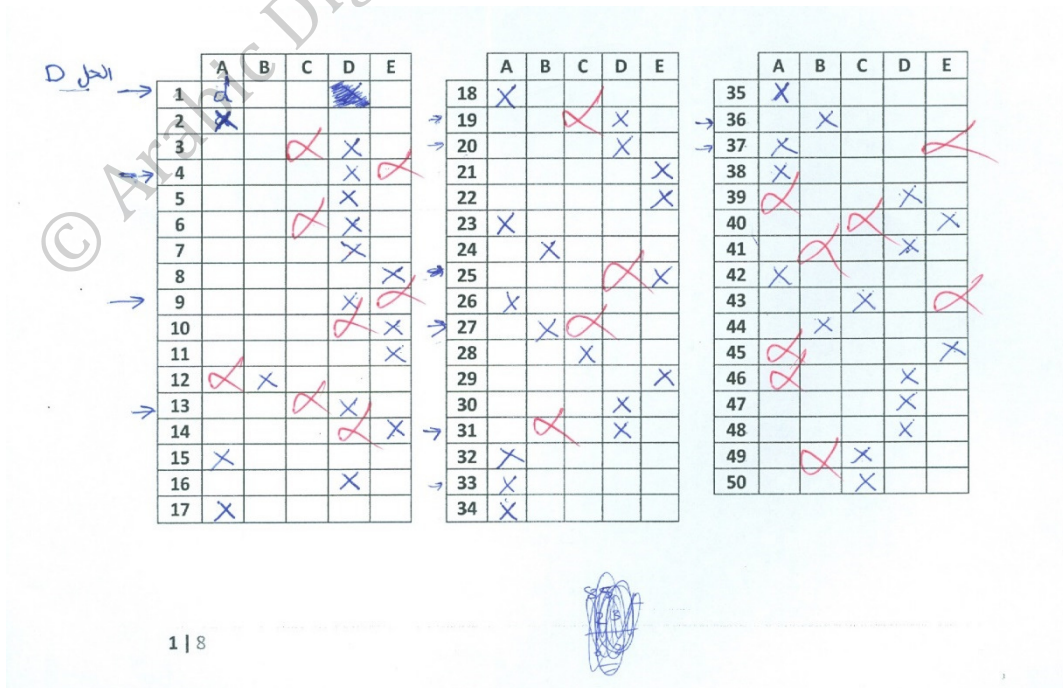Figure 54: Example 4 of not solved special cases.

68

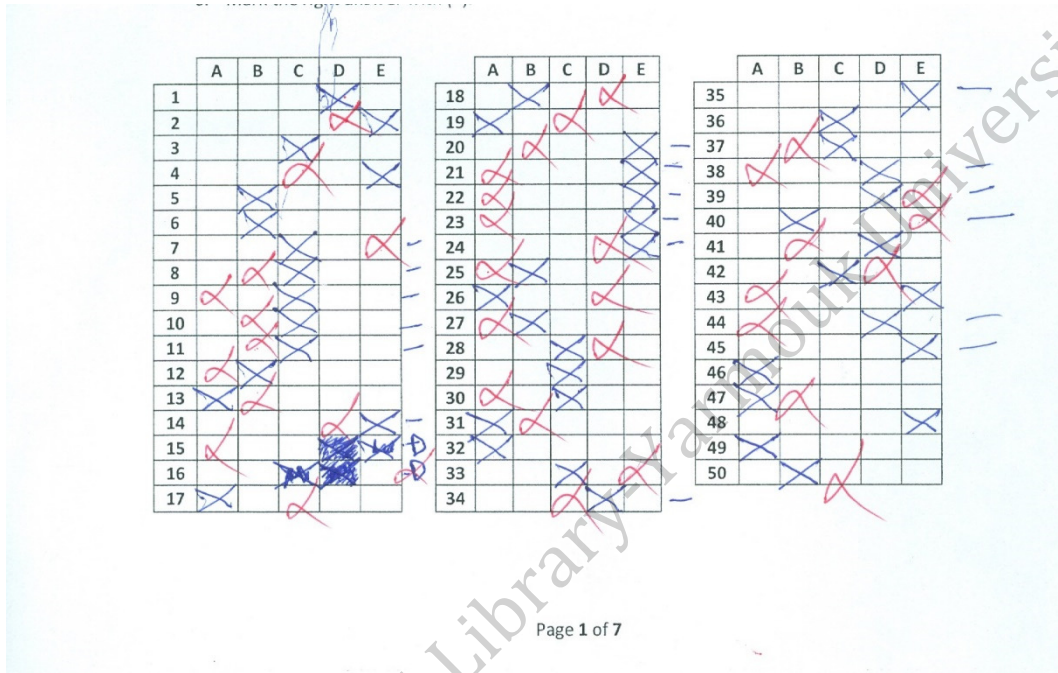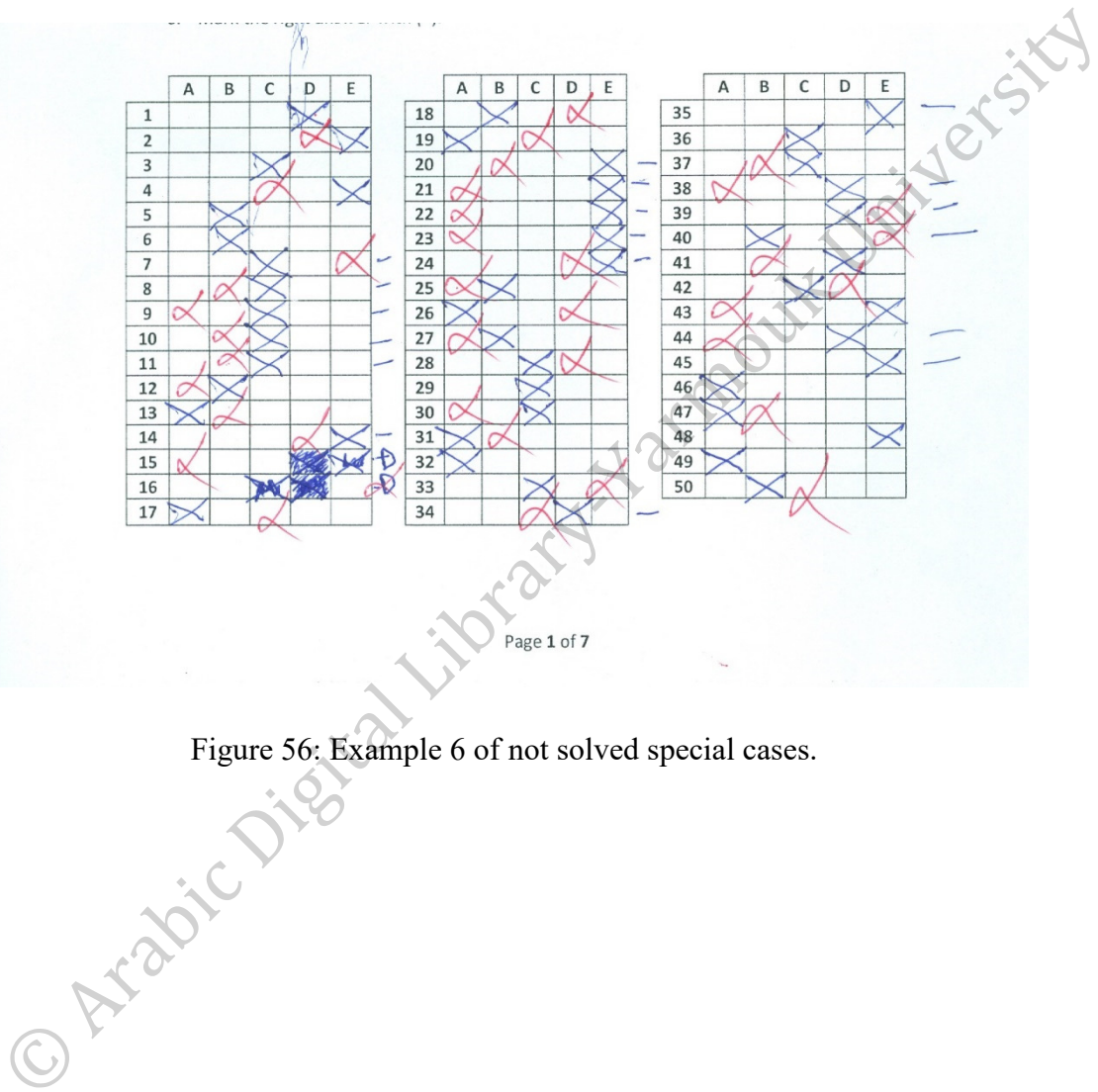Figure 55: Example 5 of not solved special cases.

Figure 56: Example 6 of not solved special cases.

69

CHAPTER 6

Conclusions and Future Work

6.1 Conclusions

In this thesis, the proposed framework is designed to be efficient, reliable, accurate, very fast, flexible and fully automated scoring multiple-choice exams to surpass the manual and limited existing systems, those are fixed, time and effort consuming. Furthermore, some existing applications are only applicable for custom answers methods (i.e. dark circles). In our approach we solved this issue by designing an application based on pattern recognition mechanism, which takes students answers (i.e. A, B, C, D, a, b, c, d or any other symbols like √, ×) as patterns. The proposed application takes the scanned answer-sheets as input images, these images may have noise, or they may not be aligned well. In our design we solved this issue by defining preprocessing stage to enhance the input images and crop the region of interest (ROI) by finding 98% of middle image sum of black pixels (Top, Bottom, Left, Right), that includes the answer-tables.

For more performance improvements, this work provided a localization mechanism to work in parallel with feature extraction, pattern recognition and segmentation mechanisms. All of these mechanisms and others are applied together in the processing stage to generate features matrix for the key answer image and a matrix for each input image (Answer-sheets). Those matrices represent input data for the scoring stage to find similarity and return the final scores.

This application/framework is defined to manipulate unlimited input images, which means we don't need any input parameters values. The proposed application reads all stored images in the input folder and process them during the running stage which makes it more flexible. For more flexibility, the proposed application is applicable with inputs annotated by any symbol.

Defining the region of interest, the proposed system manipulates only the important parts and ignore all unimportant parts, which reduces the internal storage memory as well as the total computational time. This application also uses localization tools, which help in reducing the processing time.

In addition, during the implementation stage we find theta using MATLAB functions and the Histogram^4 for peak-detection and for determining threshold values. This mechanism helps the proposed application to overcome some special cases (i.e. when the student annotates two answer cells). In this cases, the system applies a voting mechanism based on theta calculations and chooses the strongest possible answer.

The proposed system has several advantages over previously proposed models, for example our design can work on unlimited input images. Also it is applicable with several input patterns.

6.2 Suggestions for Future Work

Although the proposed application was improved the existing methods, there are some suggested improvements that could be made

In future, the proposed application may be further developed to:

- ✓ Detect student seat number.
- ✓ Detect form number.
- ✓ Handle questions that have more than one best choice.
- ✓ Be a mobile application for scoring.
- ✓ Deal with short-answer-questions exams.

References

[1] Carneson, John, Georges Delpierre, and Ken Masters. "Designing and managing multiple choice questions." (2016)

[2] Ali, Syed Haris, Patrick A. Carr, and Kenneth G. Ruit. "Validity and Reliability of Scores Obtained on Multiple-Choice Questions: Why Functioning Distractors Matter." Journal of the Scholarship of Teaching and Learning 16.1 (2016): 1-14.

[3] Wood, E. J. "What are extended matching sets questions?." Bioscience Education 1.1 (2003): 1-8.

[4] De Laet, Tinne, et al. "Assessing engineering students with multiple-choice exams: theoretical and empirical analysis of scoring methods." Proceedings of the 44th Annual SEFI Conference. 2016.

[5] Shrivastava, Vivek, and Navdeep Sharma. "Artificial neural network based optical character recognition." arXiv preprint arXiv:1211.4385 (2012).

[6] Saengtongsrikamon, Chatree, Phayung Meesad, and Sunantha Sodsee. "Scanner-based optical mark recognition."Information Technology Journal 5.1 (2009): 69-73.

[7] Cupic, Marko. "A case study: using multiple-choice tests at university level courses– preparation and supporting infrastructure." International Journal of Intelligent Defence Support Systems 3.1-2 (2010): 90-100.

[8] Čupić, Marko, et al. "Supporting automated grading of formative multiple choice exams by introducing student identifier matrices." IEEE International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). 2011.

[9] Supic, Marko, et al. "Automatic recognition of handwritten corrections for multiple-choice exam answer sheets." IEEE, 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). 2014.

[10] Kim, Jin Ho, Kye Kyung Kim, and Ching Y. Suen. "Hybrid schemes of homogeneous and heterogeneous classifiers for cursive word recognition." Proc 7th International Workshop on Frontiers in Handwriting Recognition, Amsterdam. Netherlands. 2000.

[11] Surinta, Olarik, Lambert Schomaker, and Marco Wiering. "Handwritten Character Classification using the Hotspot Feature Extraction Technique." ICPRAM (1). 2012.

[12] Abbas, A. A. "An Automatic System to Grade Multiple Choice Questions paper based exams." Journal of Al-Anbar University for Pure Science 3.1 (2009).

[13] Singh Ramesh ,Sidhu Manjit. "Exploring the Potential of Multiple Choice Questions in Computer-Based Assessment of Student Learning." Malaysian Online Journal of Instructional Technology .Vol. 2 , April 2005.

[14] Ionete, Bogdan, and Ionuţ Lambrescu. "Automatic Evaluation of Scanned Multiple Choice Tests." Petroleum-Gas University of Ploiesti Bulletin, Mathematics-Informatics-Physics Series 62.2 (2010).

[15] Patole, Soumitra, et al. "Automatic System For Grading Multiple Choice Questions And Feedback Analysis". International Journal of Technical Research and Applications e-ISSN: 2320-8163,2016.

[16] Fisteus, Jesus Arias, Abelardo Pardo, and Norberto Fernández García. "Grading multiple choice exams with low-cost and portable computer-vision techniques." Journal of Science Education and Technology 22.4 (2013): 560-571.

[17] Sharma, Deepak, et al. "Data extraction from exam answer sheets using OCR with adaptive calibration of environmental threshold parameters." IEEE, International Conference on Signal Processing and Communication (ICSC), 2013.

[18] Swaroop, Paridhi, and Neelam Sharma. "An Overview of Various Template Matching Methodologies in Image Processing." International Journal of Computer Applications 153.10 (2016).

[19] Jena, Kalyan Kumar. "Application of COM-SOBEL operator for edge detection of images." IJISET, Engineering & Technology 2.4 (2015): 48-51.

[20] Aslam, Asra, Ekram Khan, and MM Sufyan Beg. "Improved edge detection algorithm for brain tumor segmentation." Procedia Computer Science 58 (2015): 430-437.

[21] Hassanein, Allam Shehata, et al. "A survey on Hough transform, theory, techniques and applications." arXiv preprint arXiv:1502.02160 (2015).

[22] Sulaiman, Siti Noraini, et al. "An expert image processing system on template matching." International Journal of Computer Science and Network Security 7.7 (2007): 234-238.

[23] Cozzella, L., and G. S. Spagnolo. "Phase-only Correlation Function by Means of Hartley Transform." JSM Mathematics and Statistics 1.1 (2014): 1-8.

[24] Ulysses, J. N., and A. Conci. "Measuring similarity in medical registration." IWSSIP 17th International Conference on Systems, Signals and Image Processing. 2010.

73

[25] Safari, Pooyan. "Deep learning for sequential pattern recognition". MS thesis. Universitat Politècnica de Catalunya, 2013.

[26] Robert P.W. Duin, El_zbieta P, ekalska."Pattern Recognition:Introduction and Terminology ", 37steps, a tutorial." (2016).

[27] Bengio, Samy, et al. Guest editors. "introduction: Special section on learning deep architectures." IEEE transactions on pattern analysis and machine intelligence 35.8 (2013).

[28] C. Tappert, "Computer vision: algorithms and applications," Choice, vol. 48, (9), pp. 1733-1734, 2011.

[29] Gonzalez, Rafael, Richard Woods, and Steven Eddins. "Digital Image Processing Using Matlab Gatesmark Publishing." (2009).