ABSTRACT

Title of Document: MACHINE VISION TECHNOLOGY FOR FOOD QUALITY AND SAFETY INSPECTIONS

Fenghua Jin, Doctor of Philosophy, 2008

Directed By: Professor Yang Tao, Fischell Department of Bioengineering

With increased expectations for food products of high quality and safety standards, the need for accurate, fast and objective determination of these characteristics in food products continues to grow. Machine vision as a non-destructive technology, provides an automated and economic way to accomplish these requirements. This research thus explored two applications of using machine vision techniques for food quality and safety inspections.

The first application is using a combined X-ray and laser range imaging system to detect bone and other physical contaminants inside poultry meat. For this project, our research focuses on how to calibrate the imaging system. A unique three-step calibration method was developed and results showed that high accuracy has been achieved for the whole system calibration – a root mean square error of 0.20 mm, a standard deviation of 0.20 mm, and a maximum error of 0.48 mm.

The second application is separating walnuts' shells and meat. A backlight imaging system was developed based on our finding that the backlit images of walnut shells and meat showed quite different texture patterns due to their different light

transmittance properties. The texture patterns were characterized by several rotation invariant texture analysis methods. The uncorrelated and redundant features were further removed by a support vector machine (SVM) based recursive feature elimination method, with the SVM classifier trained concurrently for separations of walnuts' shells and meat. The experimental results showed that the proposed approach was very effective and could achieve an overall 99.2% separation accuracy. This high separation accuracy and low instrument cost make the proposed imaging system a great benefit to the walnut processing industry.

MACHINE VISION TECHNOLOGY FOR FOOD QUALITY AND SAFETY INSPECTIONS

By

Fenghua Jin

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:
Professor Yang Tao, Chair
Professor Hubert Montas
Professor Adel Shirmohammadi
Professor Nam Sun Wang
Professor Fredrick Wheaton

Acknowledgements

I am extremely grateful to my advisors, Dr. Yang Tao, for seeing my research potential and pushing me to complete my PhD degree. His confidence in my capability has enabled me to do a lot of creative work. I have learned tremendously as a direct result of working with him.

I would also like to thank Dr. Adel Shimohammadi, Dr. Fredrick Wheaton, Dr. Hubert Montas, and Dr. Nam Sun Wang for serving on my committee and for their time, support, and helpful advices during my research and academic development.

I also want to express my appreciation to all present and former members of the Bio-imaging and Machine Vision Laboratory at University of Maryland for their constant suggestions and friendship, including Bin Zhu, Lu Jiang, Xin Chen, Hansong Jing, Xuemei Chen, Abby Vogel, Angela Vargas, and Gary Seibel.

I would like to extend my deepest gratitude to my parents, who have always believed in me and knew I could accomplish the goal of being the first PhD in my family. And to my wife Lei Qin, thanks for the enduring love, encouragement, inspiration and support through all these years.

Last, but far from least, I want to express my thanks to all the people who have made this thesis possible and because of whom my graduate experience has been one that I will cherish forever.

Table of Contents

List of Tables

List of Figures

List of Symbols

λ	Wavelength
I_X	X-ray image
I_0	Incident X-ray intensity
μ	X-ray attenuation coefficient of a material
d	The length of the X-ray pathway through the material
f_{XL}	Transformation function between X-ray and laser range images
$D(x,y)$	Range image obtained from laser range imaging system
$I_{pX}(x,y)$	Pseudo X-ray image
$I_{tc}(x,y)$	Thickness compensated image
M	Point in the world coordinate system
(X_w,Y_w,Z_w)	Coordinates of a point in the world coordinate system
(X_c,Y_c,Z_c)	Coordinates of a point in the camera coordinate system
R	3-D rotation matrix
$r_1, r_2, \ldots, r_9$	Elements of the 3-D rotation matrix
T	3-D translation vector
T_x, T_y, T_z	Components of the translation vector in x, y, z axis directions
(x_u,y_u)	Undistorted image coordinates of a point
f	Effective focal length of the CCD camera
C	CCD camera center
p	Principle point of the CCD camera
(x_d,y_d)	Distorted image coordinates of a point
k_1	First order lens radial distortion coefficient
d_x	Center-to-center distance between adjacent CCD sensor elements in the row direction
d_y	Center-to-center distance between adjacent CCD sensor elements in the column direction
(x_f,y_f)	Image buffer coordinates of a point
(x_0,y_0)	CCD camera principal point coordinates in the image
dt	Camera lens tangential distortion
dr	Camera lens radial distortion
f'	Effective focal length of the X-ray line-scan camera
c'	X-ray focal spot
p'	Principle point of the X-ray line-scan camera
H	Homography matrix for X-ray line-scan camera

$h_1, h_2, \ldots, h_6$	Elements of the homography matrix H
s	An arbitrary scale factor
u_0'	X-ray line-scan camera principal point coordinates in the image buffer coordinate system
m'	Image projection of a model point in the X-ray camera image plane
θ	Angle between the conveyor belt and the detector image plane
K	Intrinsic matrix of the X-ray line-scan camera
U	Uniformity measure used by LBP operator
$hist_i$	ith bin of the normalized histogram of LBP values
$FeaVec_{LBP}$	Texture feature vector extracted using LBP operator
$FeaVec_{LBP\_VAR}$	Texture feature vector extracted using LBP and VAR operators
$g(x, y)$	Gaussian function
$G(x, y)$	Gabor function
$FeaVec_{GABOR}$	Texture feature vector extracted using Gabor filters
$FeaVec_{WAVELET}$	Texture feature vector extracted using wavelet transform
$FeaVec_{CS-GLCM}$	Texture features based on the circularly symmetric gray level co-occurrence matrix
NH_k	kth bin of the normalized image histogram
$FeaVec_{HIS}$	Texture features based on the image histogram
$Mean_{NH}$	Mean of the normalized image histogram
Var_{NH}	Variance of the normalized image histogram
Eng_{NH}	Entropy of the normalized image histogram
Ske_{NH}	Skewness of the normalized image histogram
Kur_{NH}	Kurtosis of the normalized image histogram

List of Abbreviations

ANN	Artificial Neural Network
AR	Autoregressive
ARMA	Autoregressive moving average
CCD	Charge-coupled device
CPU	Central processing unit
CS-GLCM	Circularly symmetric gray level co-occurrence matrix
eV	Electron volt
FOV	Filed of view of a camera
FFT	Fast Fourier transform
GLCM	Gray level co-occurrence matrix
LBP	Local binary operator
LUT	Look-up table
lb	Pound
MA	Moving average
MAX	Maximum
MB	Million bytes
MeV	Mega electron volts or million electron volt
MLP	Multilayer perceptron
NCH	Normalized cumulative histogram
RAM	Random access memory
RBF	Radial basis function
RFE	Recursive feature elimination
RMS	Root mean square
SFS	Shape from shading
SFT	Shape from texture
STD	Standard deviation
SOM	Self-organizing map
SVM	Support vector machine
SVM-RFE	Support vector machine based recursive feature elimination
TOF	Time of flight
USDA	United States Department of Agriculture
VAR	Variance measure in LBP local neighborhood
2-D	Two dimensions
3-D	Three dimensions

Chapter 1: Overall Introduction

With increased expectations for food products of high quality and safety, the need for accurate, fast and objective determination of these characteristics in food products continues to grow. Machine vision as a non-destructive technology, provides an automated and economic way to accomplish these requirements. This research explored two applications of using machine vision techniques for food quality and safety inspections. One is using a combined X-ray and laser range imaging system to detect physical contaminants inside poultry meat, and the other is using a backlight imaging system to separate black walnut shells and meat. The problems and rationales of these two projects will be introduced as follows, respectively.

1.1 *A Combined X-ray and Laser Range Imaging System to Detect Physical Contaminants in Poultry Meat*

The United States produces over 43 billion pounds of poultry meat annually, over 41% of which is boneless poultry meat (USDA, 2008). Physical contamination such as plastic, metal, glass and bone fragments, etc. is a major problem in food safety which could lead to serious injury if accidentally swallowed. Thus, the detection of bone fragments and other physical contaminations in deboned poultry meat has become increasingly important to ensure food quality and safety (Tao, et al., 2000, 2001). In many poultry processing plants, the inspection of boneless meat is still performed manually (Figure 1.1). The accuracy of manual inspection depends on the sensitivity of workers' fingers. To prevent microbial proliferation, meat must be chilled to a little above the frozen temperature. Such coldness causes the workers'

fingers to quickly become numb to the feel of a possible bone fragment. As a result, cotton gloves are provided for the workers which consequently reduces the sensitivity of the fingers and thus decreases accuracy in bone detections. Additionally, cross-contamination and high labor costs are also problems of manual inspection.

Figure 1.1 Poultry processing line. Note the physical contaminants inside meat are inspected manually by the workers.

X-ray imaging techniques, as noninvasive inspection methods, have been used for years to detect physical contamination in food products. However, the traditional X-ray imaging detection technologies have significant difficulties in detecting contaminations because of variations in the poultry thickness. In order to address this problem, Tao and Ibarra (2000A) suggested adding a laser range imaging subsystem and combining the depth information from the laser range image with the X-ray image to achieve accurate physical contaminate detection. The X-ray line-scan

camera captures high resolution X-ray images in real-time and the three-dimension (3-D) laser imager provides an accurate thickness profile for each piece of poultry. In the combined system, the 3-D thickness information is used to cancel the thickness variation in the X-ray image, thus the process of physical contamination detection is significantly simplified (Chen, 2003; Jing, 2003).

Figure 1.2 is a schematic of the described combined system which contains four CCD (charge-coupled device) cameras and one X-ray line-scan camera. When objects on the conveyor belt pass through the X-ray and laser beams, X-ray images are captured by the X-ray line-scan camera below the conveyor belt, while the laser light stripe images are recorded by the CCD cameras to get the 3-D profile of the object (based on a structured lighting method) (Chen, et al., 2000). Proper calibration of the five cameras is, of course, required to achieve high accuracy of foreign material detection. A unique three-step geometric calibration method was developed in our research to address this problem.

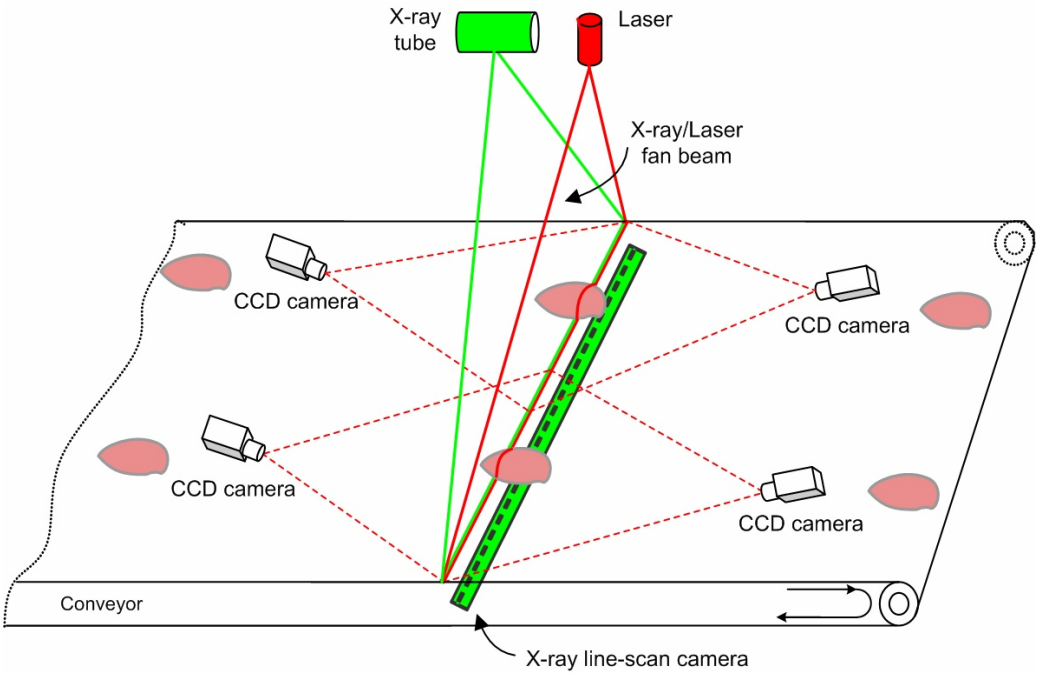


Figure 1.2 A schematic plot of the combined X-ray and laser range imaging system. Four CCD cameras are placed on the both sides of laser beam plane to reduce the influence of possible occlusion caused by object curvature and ensure enough field cover range. Note that the X-ray and laser beam are aligned carefully to be coplanar, though they are not coplanar here for illustration purposes.

1.2 *A Backlight Imaging System for Walnut Shell and Meat Separation*

The second machine vision application for food safety inspection in our research is using a backlight imaging system to separate walnut shell and meat. Black walnuts, frequently added to food from appetizers to desserts, are nutritious and flavorful with a long shelf life when refrigerated or frozen (Figure 1.3). Eastern black walnuts grow throughout the central and eastern United States. The annual yield of raw black walnuts in the U.S. is approximately 4 billion pounds (Jones, et al., 1998; Hatcher et al., 1998), however only about 20 million pounds are commercially processed (Hammons, 1998). Walnut farmers are not motivated to harvest their crop

because the raw nuts are sold at low prices to nut processors. Furthermore, nut processing capacity within the United States is limited.

(a) (b)

Figure 1.3 (a) Photograph of raw black walnuts; (b) Cracked black walnuts showing (from top to bottom row): light meat, dark meat, inner shell surface and outer shell surface.

In walnut processing plants today, nuts are cracked and the majority of the shells are removed by air lathe. Any remaining shell fragments are removed by visual inspection to meet the required marketable quality. Manual inspection is very time consuming and labor intensive because the shell and meat fragments are very similar in size and color (Krishnan and Berlage, 1984). Therefore, an accurate automated inspection method is desirable to reduce labor and time while ensuring product quality. A backlight imaging system was developed in this research to differentiate black walnut shells from walnut meat since the backlit images show distinctive texture patterns. This study further explored efficient texture analysis and pattern recognition methods to achieve high accuracy for walnut shell and meat separation.

This dissertation is organized into the following chapters. Chapter 2 details the overall objective of this dissertation research and Chapter 3 provides the literature review. Chapter 4 describes the detail calibration method of the combined X-ray and laser range imaging system. Chapter 5 describes the design of a backlight imaging system and algorithms to achieve high accuracy in separating walnut shells and meat. Chapter 6 follows with overall conclusions and Chapter 7 describes suggestions for future studies.

Chapter 2: Overall Objective

The major goal of this research was to explore the two machine vision applications in the food safety inspection area. For the first project, a combined X-ray and laser range imaging system is used to detect physical contaminants in poultry meat which includes several imaging devices, four CCD cameras and one X-ray line-scan camera. Our goal was to develop an efficient system calibration method. For the second project, our goal was to develop a machine vision system to separate walnut shells from meat.

Specifically, the aims of this research were to:

(1) Develop a calibration method to calibrate CCD cameras and X-ray line-scan camera used in the combined X-ray and laser range imaging system. Develop hardware and software to achieve automated or semi-automated system calibration so that the whole calibration procedures can be accomplished in a timely fashion.

(2) Evaluate the precision of the calibration method.

(3) Develop a backlight imaging system for walnut shell and meat separation. Explore texture analysis and pattern classification methods for the acquired backlit walnut images.

(4) Evaluate the performance of the backlight imaging system.

Chapter 3: Review of Literature

3.1 *X-ray Imaging*

X-rays are a type of electromagnetic radiation with wavelengths ranging from 10^{-3}nm to 10 nm, with photon energy ranging from 120 eV to 1.2 MeV (Selman, 1994). The basic production of X-rays is by accelerating electrons with a high voltage and allowing them to collide with a metal target in an X-ray tube (Figure 3.1).

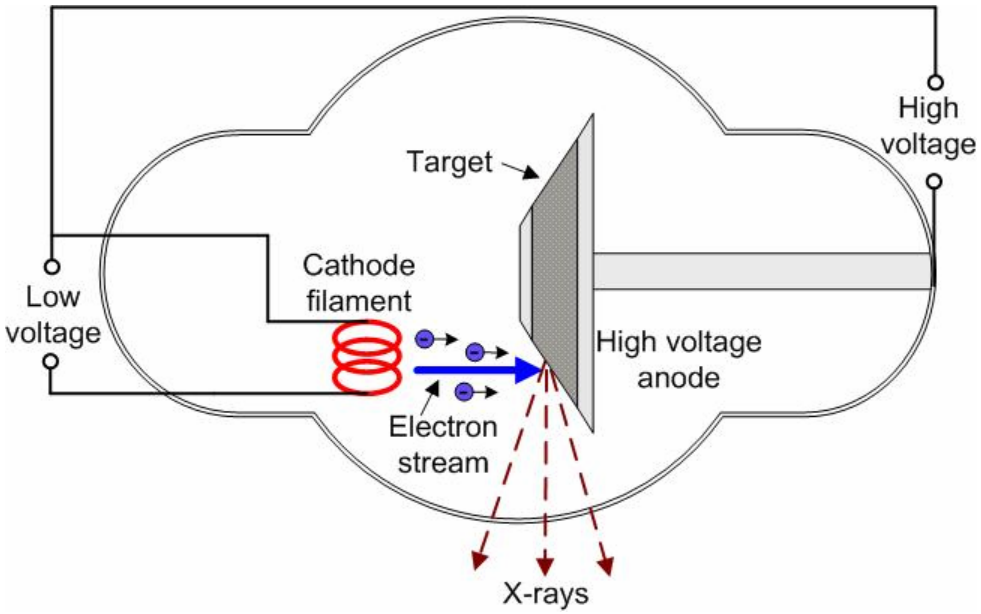


Figure 3.1 Model of an typical X-ray tube which is composed of: a tube from which the air has been evacuated as completely as possible, a hot filament which liberates electrons (the filament is driven by a low voltage power supply and it also serves as the cathode when high voltage is correctly applied), a target which serves as the anode, and a high voltage applied across the cathode and anode, causing the electrons to rush at an extremely high speed through the tube from cathode to anode.

X-ray radiation is generally classified into two types: bremsstrahlung radiation and characteristic radiation. Bremsstrahlung radiation, also called braking radiation, is emitted when the electrons suddenly decelerate upon colliding with the metal target. The kinetic energy lost by each electron is radiated as an X-ray of equivalent energy.

Braking radiation is poly-energetic (nonuniform in energy) because the amount of braking or deceleration varies among electrons. If the accelerated electron contains enough energy, it is able to knock out an electron from the inner shell of the target metal atom. As a result, electrons from higher energy levels drop down to fill up the vacancy and X-ray photons are emitted whose energies are determined by the electron energy levels. These X-rays are called characteristic X-rays (Dowsett, et al., 2006). A typical X-ray spectrum contains both braking radiation and the characteristic radiation as shown in Figure 3.2 (Hamamatsu, 2008). Notice that the spectrum of braking radiation is continuous, while characteristic radiation forms line spectra.

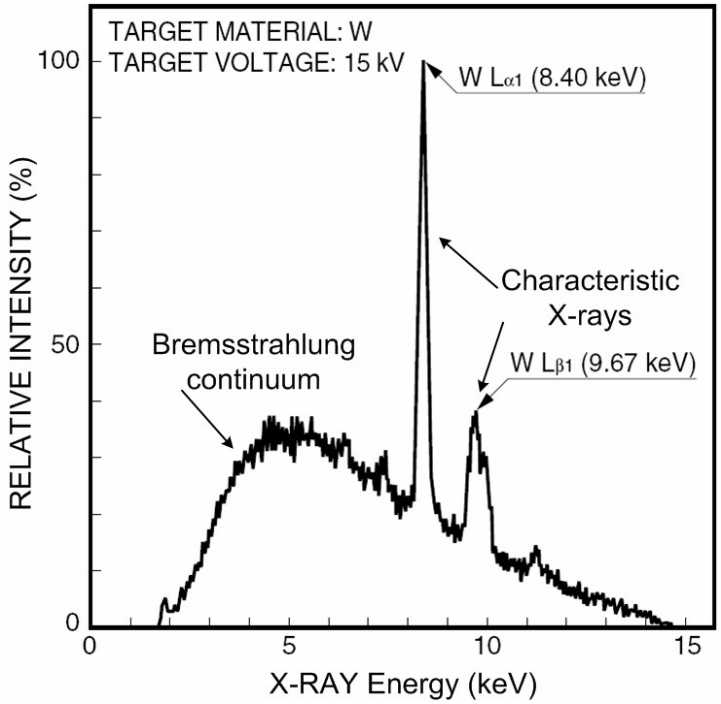


Figure 3.2 A typical X-ray spectrum (Hamamatsu, 2008). Note the characteristic X-ray spectrum depends on the target material used.

One important property of X-rays is that they are invisible rays with high penetration ability. The penetrating nature of X-rays makes them ideal tools to study the internal structure of objects based on the interaction between the X-rays and the

matter. A typical X-ray imaging system is depicted in Figure 3.3, where X-rays are generated from the X-ray tube, attenuated by the target object, and then intercepted by an X-ray imager. The imager can be an X-ray film (Selman, 1994), a video camera coupled with an image intensifier (Molloi, et al., 1995), a linear digital X-ray detector array (Iacobaeus, et al., 2007), or a two dimensional (2-D) digital X-ray detector arrays (X-ray panel) (Zentai, et al., 2006; Szeles, et al., 2007).

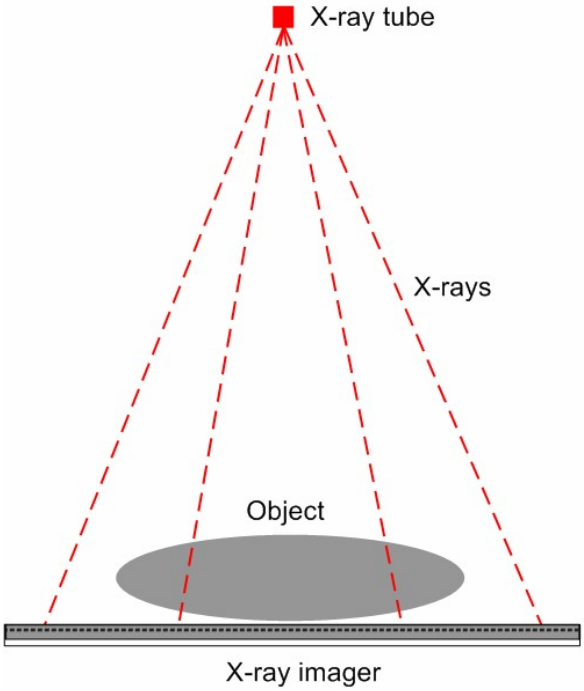


Figure 3.3 Typical setup of an X-ray imaging system

X-ray imaging is a well-established technology and has been widely used in many areas, including medical diagnosis, homeland security, material research, and industrial inspections (AS&E, 2008; Fahrig, et al., 2008; Roh, et al., 2003). In the field of food inspection, X-ray technology also provides an invaluable means to ensure food safety and to assess food quality. Shahin, et al. (2002a) developed an X-ray imaging to detect apple surface bruises. Spatial and transform features comprise

the basis for apple bruise detection in the study, and a back propagated artificial neural network (ANN) classifier was found to be efficient for apple classification. The same imaging system was also revealed to be successful in detecting internal defects of sweet onions (Shahin, et al., 2002b). It has been shown that X-ray imaging can also be applied to detect infested pistachio nuts (Casasent, et al., 2001, 2003). The internal product detail that X-ray images provide allows the presence of worm damage and other defects to be determined non-destructively.

Significant research efforts have also been made on meat inspection using X-ray. The difficulty of X-ray imaging in the meat inspection application is that the thickness of the poultry meat is not uniform, which makes it hard to differentiate physical contaminants and thicker meat portions. Various approaches have been used to address this thickness issue. One method is to immerse the meat in water so the thickness variation can be largely smoothed out. Although successful in thickness cancellation, the method is unlikely to be acceptable for on-line inspection applications due to the sanitation problem and the difficulties in implementing an efficient meat handling and rejection apparatus. Another thickness cancellation method adopted by the industry is to press the meat mechanically. A pump is used to press the meat into a pipe, compressing it into a rectangular block of uniform thickness before it is scanned by an X-ray imager. It was reported that the method worked particularly well for ground meat inspection (Hartman, 2001). However, for products such as poultry fillets, where the preservation of the natural shape is desired, this compression method seems unattractive due to inevitable meat damage. Furthermore, once the X-ray imaging system reports a contamination in the pipe, it is

often difficult for human inspectors to determine the exact location of the defect in the chunk of rejected meat (Graves and Batchelor, 2003).

Note that the above methods for thickness cancellation are based on X-ray imaging with a single energy X-ray. The dual energy X-ray imaging method can also help address the thickness variation problem (Graves and Batchelor, 2003). It is known that the attenuation coefficient of any material changes with X-ray energy, and the change is also different for different materials which make it possible to differentiate two materials based on the images acquired under two different X-ray energy settings (Taibi, et al., 2003; Lemacks, et al., 2002). Commercial inspection machines using dual X-ray energy methods are available for bone detection (Jamieson, 2002). The dual energy system can detect ossified bone fragments in chicken fillets, thighs and nuggets. However, soft bones are still not detectable by the X-rays (Graves and Batchelor, 2003). Limitations of this dual energy method come from the underlying physics. One is due to the difficulty to make a monochromatic X-ray source. The common approach to achieve dual energies is by using X-ray filters. Although the two filtered X-ray are peaked at two different energy levels, both of them are still multi-chromatic which breaks the assumption of the dual-energy X-ray imaging model. In addition, in order to have high quality X-ray images, the high energy level has to be limited due to the fact that poultry meat usually isn't very thick. Otherwise, the X-ray sensors are prone to be saturated by overexposure to the high energy X-ray photons. This implies that the difference between the high and low energies is small, which leads to insufficient contrast between the images taken at the two energies.

In order to address the thickness variation problem, a more plausible method is to acquire the thickness information directly by using optical range imaging technologies. The resultant profile of the object can then be used to compensate for the thickness effects in the X-ray images. A brief review of the optical way of 3-D measurement is given in the next section.

3.2 *Optical Range Imaging Technologies*

Optical range imaging refers to the technology of determining three-dimensional information of an object surface via optical sensing. Range imaging systems collect large amount of 3-D coordinate data from visible object surface in a scene. These systems can be used in a wide variety of automation applications, such as object shape acquisition, inspection, robotic assembly, and medical diagnosis (Davies, 2005; Peters, 2006).

Optical range imaging methods can be categorized into passive and active approaches (Klette, et al., 1998). Passive methods require no interaction with the 3-D scene while active methods recover the object surface by intentionally projecting energy into the scene (e.g. project a structured lighting pattern, or produce variations of the illumination).

Typical passive approaches include static stereo analysis, dynamic stereo analysis, shape from shading, and shape from texture (Klette, et al., 1998).

Static stereo analysis infers the 3-D information of a scene from two or more images taken from different viewpoints (Klette, et al. 1998). The so-called standard stereo geometry is shown in Figure 3.4. A feature point (P) in the scene is viewed by

the two cameras at different positions, thus showing at different locations in the corresponding image plane ($p1$ and $p2$). If both cameras are calibrated beforehand, the 3-D information of scene point P can be recovered by analyzing the disparity between image points $p1$ and $p2$. The main problem of static stereo analysis is locating matching points. One way of simplifying this task is to precede the matching algorithm by an edge detection algorithm, to reduce the number of points involved. This method is especially useful when there are only simple geometric objects in a scene. Another way of reducing the required computation is to use a hierarchical approach that first matches images at a very coarse resolution and then processes to matches at progressively finer resolution scales.

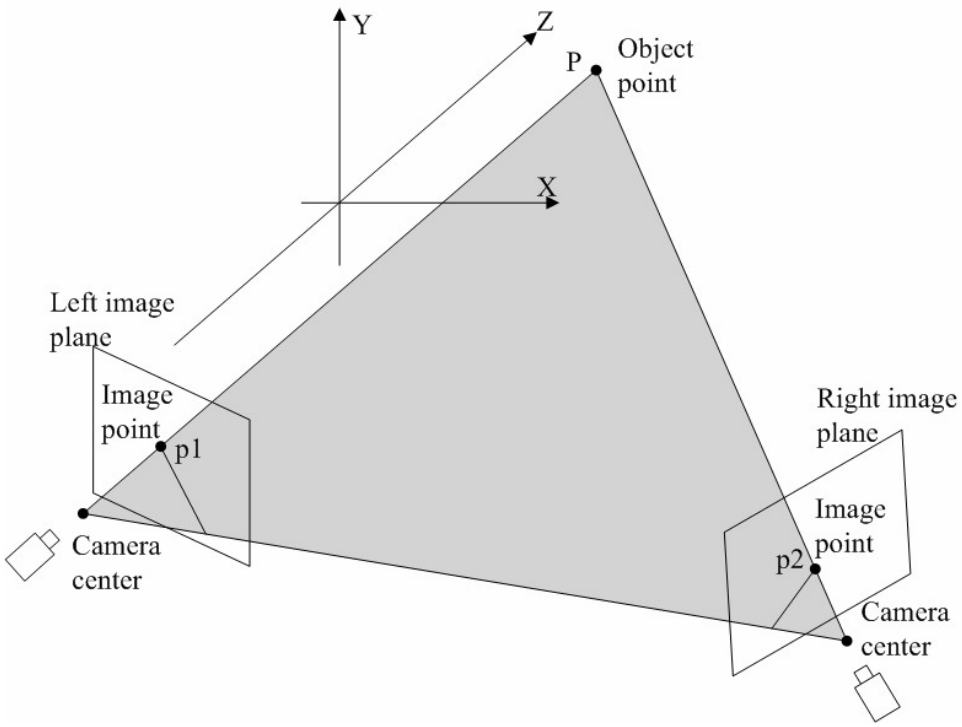


Figure 3.4 Stereo vision using two cameras.

Unlike static stereo analysis where no object and camera movement occur during the time interval of image acquisition of two (or more) stereo images, dynamic stereo analysis assumes that images are taken from a continuously moving scene (or a moving camera) (Soatto, 2007). The aim is to recover the 3-D geometry of the scene as well as its motion relative to the camera from a sequence of images. The movement of the scene object in three-dimensional space corresponds to a 2-D motion in a captured image sequence. Two steps are usually adopted in numerous dynamic stereo analysis methods: (1) compute observables in images (various observables can be used, such as points, lines, occluding contours, and optical flow), and (2) relate these observables to recover the 3-D information.

Shape from shading (SFS) methods exploit the changes in the image intensity (shading) to recover surface shape information. This is done by calculating the orientation of the scene surface corresponding to each point in the image. Because the intensity of a particular pixel in an image depends on the light source(s), surface reflectance properties of the object, and local surface orientation (expressed by a surface normal **n**), SFS methods try to extract information about surface normals in view solely on the basis of the image intensity. SFS is one of the classic ill-posed problems of scene analysis (Klette, et al., 1998) as we can imagine there is not sufficient information contained in an arbitrary irradiance image to reconstruct the object surface unambiguously. Simplifying assumptions need to be made about illumination, surface reflectance properties, and surface smoothness, to make the SFS task solvable (Sonka, et al., 1999) and SFS can only be expected to work well by itself in highly controlled environments.

Shape from texture methods (SFT) try to infer 3-D information of the scene using texture as a cue. Image plane variations in the texture properties such as density, size, and orientation are usually explored by SFT algorithms. For example, if an object surface is covered with circle patterns, quantifying the changes in the shape of texture elements (circles appearing as ellipses) enable us to determine surface orientation. To recover the surface orientation, we must have accurate methods to delineate each primitive in the image plane. For simple binary primitives such as circles and triangles, fairly accurate segmentation of individual image plane primitives for measurement is possible. However, for more complex gray-level textures corrupted by noise, it is usually difficult to accurately estimate the image plane features, which limits practical use of SFT methods (Davies, 2005).

All the above mentioned passive techniques do not rely on special scene illumination. Although passive approaches usually require little special-purpose hardware, they do not yield the highly dense and accurate range images that many applications require. Most of the shape recovery techniques in use today are active systems which illuminate the scene in an advantageous manner. The active methods can be divided into three broad categories: time of flight methods, interferometric methods, and structured lighting methods.

The time of flight (TOF) based 3-D measurement makes use of the time it takes for light to travel the distance from a light source, to an object surface, and reflect back onto the light receiver (Seta and O'ishi, 1990). There are three basic ways to accomplish TOF range imaging. The most straight forward type is the pulse-modulated TOF measurement which simply turns on a light source and at the same

time start a stop watch. This stop watch is then read as soon as the detector senses an increase in intensity of the incoming light. The second type, continuous-wave modulated TOF, measures the phase difference between the continuous modulating signal and the modulation of the received light. The third type is pseudo-noise coded TOF. By modulating a light source with pseudo-noise, autocorrelation of the modulated signal achieves pulse compression for more accurate phase detection (Möller, 2005).

Interferometric methods rely on the fact that waves with a certain fixed wavelength can be used as a measuring stick based on wave interference. The interference light waves are usually generated by splitting one light beam into two components or using two different light sources with different wave length. One beam travels a known reference path while the other beam is sent to the object, reflected back, and also allowed to interfere with the reference beam. The two beams are then compared and the relative shift between these beams is measured in terms of wavelength. In general, interferometric measurement can achieve extremely high accuracy. However, if the light beams share the same wavelength, the measurements are only unique over a limited range of $\lambda/2$ (λ is the wavelength) and are periodic with period λ . Multiple-wavelength techniques can be used to stretch the unambiguity range of interferometers (Möller, 2005).

The structured lighting method is also widely used for 3-D reconstruction (Jalkio, et al., 1985; Dorsch, et al., 1994; Chen, et al., 2000). In a typical structured light system, a specially designed lighting pattern is projected on the object, and the deformed fringe pattern on the surface is recorded by a camera (Figure 3.5). If the

spatial relation of the light source and camera is calibrated beforehand, the 3-D information of the points in the fringe pattern can be reconstructed using triangulation.

The structured lighting approach is preferred in many applications over other ranging methods. They have the advantage of simplicity and compactness over the time-of-flight methods. Compared with interferometric techniques, they are better in the size and depth of the field of view. Generally speaking, structured lighting approaches usually have the following advantages: low cost, high 3-D resolution, convenient customization, and robustness against environment or object variation (Chen, et al., 2000). In this research, a structured lighting system containing a laser source and four high-speed cameras has been developed.

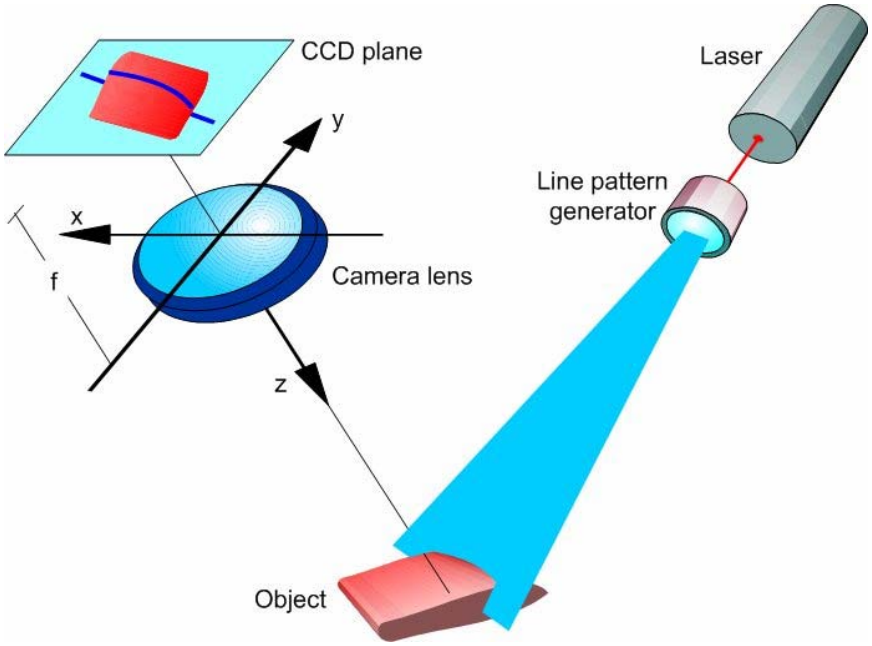


Figure 3.5 Structured lighting based shape recovery by projecting single light stripe (Tao and Ibarra, 2000B)

3.3 *Camera Calibration*

A machine vision system usually contains more than one optics and electrical imaging devices to pursue better performance. Calibrating each imaging device is a crucial problem for further metric scene measurement. In our currently developed multi-modality imaging system (X-ray and laser range imaging), the major imaging components include four high speed CCD cameras and one X-ray line-scan camera. Calibration of these cameras is the first step towards quantitative image analysis for food quality inspection.

The goal of camera calibration is to determine the relationship between a 2-D image perceived by a camera (or images perceived by multiple cameras) and the 3-D information of the real object. Camera calibration can be divided into two phases: camera modeling and model parameter estimation. First, camera modeling deals with mathematical approximation of the physical and optical behavior of the camera using a set of parameters. The second phase of camera calibration deals with the use of direct or indirect methods to estimate the values of the parameters. Camera models usually consist of two sets of parameters: intrinsic and extrinsic (Figure 3.6). Intrinsic parameters, including camera focal length, principal point, skew factor, and lens distortions, models the internal geometry of a camera. They determine how to derive an image point position given the spatial position of the point with respect to the camera. Extrinsic parameters model the position and orientation of the camera with respect to the scene, or between different cameras.

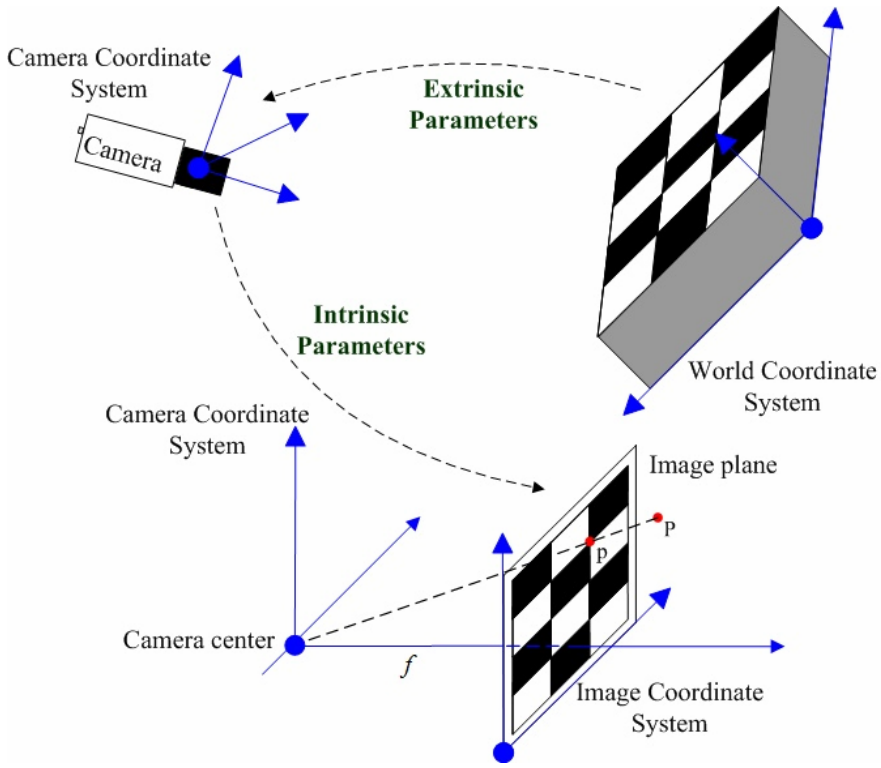


Figure 3.6 Camera calibration. The purpose of camera calibration is to determine the relationship between a 2-D image perceived by a camera and the 3-D information of the real object. Extrinsic parameters model the position and orientation of the camera with respect to the scene, or between different cameras. Intrinsic parameters determine how to derive an image point position given the spatial position of the point with respect to the camera.

Camera calibration can be classified into three types according to the calibration method used to estimate the parameters of the camera model:

1. *Direct Linear Transform techniques.* The direct linear transform techniques use least square methods to obtain a transform matrix which relates 3-D points in the scene to their 2-D projections in the image plane (Abdel-Aziz and Karara, 1971; Hall, et al., 1982; Faugeras and Toscani, 1986). It has the advantage of model simplicity and rapid calculation since only linear equations have to be solved. However, due to the linear property of the

equation, lens distortion is not incorporated into the model which causes a rough accuracy of the system. Moreover, it is sometimes difficult to extract the parameters from the matrix since cameras are calibrated without explicitly computing their physical parameters (Salvi, et al., 2002).

2. *Non-linear optimization techniques*. To account for any kind of lens imperfection into the camera model, the calibrating equations becomes non-linear. In this case, the camera parameters are usually obtained through iteration with the constraint of minimizing some predefined function (McGlone, et al., 2004). A typical minimizing function can be the summation of the distance between the imaged points and the modeled projections obtained through iteration. Although almost any camera model can be calibrated using this iterative way, these methods require a good initial guess in order to guarantee convergence.

3. *Two-step techniques*. To make use of the advantages of previously described methods, a two-step procedure, a closed-form solution followed by a nonlinear refinement, have been widely used for better calibration results (Tsai, 1987; Zhang, 2000). These techniques use a direct linear transform method to compute some of the parameters at first step. The calculated parameter values are then used as the initial values for the following iterative procedure. Two step techniques permit a rapid calibration by considerably reducing the number of iterations. Moreover, the convergence is nearly guaranteed due to the linear guess obtained using the direct linear transform method (Salvi, et al., 2002).

There are two general ways to evaluate the accuracy of camera calibration. One is to analyze the discrepancy between the real position of the 3-D object point with respect to the 3-D position estimated from its 2-D projection based on the calculated model parameters. The other way is to compare the real position of a 2-D image point with the calculated projection of 3-D object point on the image plane.

A comparative study of the most commonly used camera calibration methods of the last few decades has been presented by Salvi, et al. (2002). The major differences among the techniques covered are in the step concerning lens modeling. Camera models without considering lens distortion (Faugeras and Toscani, 1986), with only radial distortion (Tsai, 1987), and with radial distortion, decentering distortion and thin prism distortion (Weng, et al., 1992) are compared. Results show that the modeling of radial distortion is quite sufficient when high accuracy is required. The use of more complicated models does not improve the accuracy significantly because the complexity of the model ill effects the calibration. The results also show that the accuracy of non-linear calibration methods on the image plane is much better than linear methods. Thus, a two step nonlinear method is applied in our research to calibrate CCD cameras and only radial distortion is considered.

X-ray camera is different from normal optical imaging devices (e.g., CCD camera). The image intensity reflects the total absorption of X-ray in its path through the measured object. Common X-ray imagers used in the product inspection industry interpose the object to be viewed between a point X-ray source and a linear array of detectors. As the object moves perpendicularly to the fan beam of the X-ray, a 2-D

image consisting of several 1D projections is collected. The scanning geometry of the sensor can be defined by a combination of orthographic (in the motion direction) and perspective (in the orthogonal direction) transformations (Nobel et al. 1998). The linear pushbroom model can be used to analyze this imaging setup (Nobel, et al., 1994; Gupta and Hartley, 1997).

In sections 3.1 to 3.3, we've reviewed the technologies used in the combined X-ray and laser 3D imaging system, which included the X-ray imaging, optical range imaging, and camera calibration methods. Next, we'll review the techniques used in the backlight imaging system for automatic walnut shell and meat separation. Sections 3.4 to 3.6 will cover texture analysis, feature selection, and pattern recognition methods, respectively.

3.4 *Texture Analysis Methods*

Our second machine vision application in food safety inspection is to develop a backlight imaging system for automatic walnut shell and meat separation. Our experiments showed that due to the difference in light transmittance of walnut shell and meat, the texture patterns of the acquired backlit images are quite different. Appropriate texture feature analysis methods are required for the differentiation between walnuts' shells and their meat. Approaches of texture analysis are usually categorized into four types: structural, statistical, model-based, and signal processing methods.

Structural approaches (Haralick, 1979; Fu, 1982) assume that textures are composed of texture primitives and the texture is produced by the placement of these

primitives according to certain organization rules. The structure and the spatial organization of texture primitives are the key components of the resulting models. Texture primitives are building blocks of a texture, representing local spatial organization of spatially varying image signals. Examples of texture primitives include edges, shapes, and Voronio polygons (Tuceryan and Jain, 1990). The spatial organization of primitives can be random or dependent on each other within a defined neighborhood. To describe the texture, one must define the primitives and the organization rules. The advantage of the structural approach is that it provides a good symbolic description of the image. However, this feature is more useful for texture synthesis than texture analysis tasks. The symbolic descriptions can be ill defined for natural textures because of the variability of both primitives and organization rules.

In contrast to structural approaches, statistical methods do not attempt to model explicitly the hierarchical structure of the texture. Instead, they represent the texture indirectly by the statistical properties that govern the spatial distribution and relationships between the gray levels of an image. Many of these methods are based on Julesz's (1975) finding that the texture discrimination power of the human visual system has a close relationship with the spatial statistics of the image gray levels. Commonly applied statistics can be categorized into first-order, second-order, and high-order according to the number of gray levels that are taken into account simultaneously during texture analysis. First-order statistics depend only on individual pixel values and not on the interaction or co-occurrence of neighboring pixel values. They can usually be computed from the histogram of the pixel gray values in the image. Typical examples include the mean, standard deviation, and

moments of the image histogram. Second-order statistics account for the spatial inter-dependency or co-occurrence of two pixels at specific relative positions. The most popular second-order statistical features for texture analysis are derived form the so-called gray-level co-occurrence matrix (Haralick, et al., 1973). Although the approach has been proposed for several decades, it remains among the most popular and most discriminative types of texture features and has been successfully used in various applications, such as iris detection (He, et al., 2007), satellite image analysis (Ünsanlan, 2007), and content based image retrieval (Tao, et al., 2007). Other second-order statistical methods, such as autocorrelation function (Tuceryan and Jain, 1990), and local binary pattern operator (Ojala, et al., 2000, 2002), have also been widely used in texture analysis. Currently, statistics up to the second order are applied in most applications. Although, higher than second-order statistical features have also been investigated (Tsatsanis and Giannakis, 1992), the computational complexity increases exponentially with the order of statistics which limits their practical application.

Model based methods are based on the construction of an image model that can be used to both describe and synthesize texture. The model parameters estimated from an image capture the essential perceived qualities of texture. For example, the Autoregressive (AR) model is one of the stochastic models frequently used for the representation of texture characteristics. Let $\{I(i, j) \mid i \in \{1, 2, \cdots, M\},\ j \in \{1,2, \cdots, N\}\}$ be the gray value of a given texture image with a size of $M \times N$. The 2-D AR model can be represented as

$$I(i, j) = \sum_{(k,l)\in S} a_{k,l} I(i+k, j+l) + b\varepsilon(i, j) \tag{3.1}$$

where S denotes the associated neighborhood, $\{a_{k,l}\}$ is a set of parameters of the AR model which characterizes the dependence of a pixel to its neighbors, and $\varepsilon(i, j)$ is an error term. The assumption of this 2-D AR model is that in a texture image, the pixel gray level $I(i, j)$ at location (i, j) is related to the linear combination of gray values of its neighbors through a set of parameters $\{a_{k,l}\}$. The values of these parameters can be estimated from the texture image and further be used as a feature vector to distinguish different texture types (Lee, et al., 2003; Joshi, et al., 2006). Several other stochastic models, such as Gaussian Markov random fields (Deng and Clausi, 2004; Rellier, et al., 2004), moving average (MA) models (Chanyagorn and Eom, 2000), autoregressive moving average (ARMA) models (Zhong and Sclaroff, 2003) and fractal models (Petrou and Sevilla, 2006) have also been proposed for texture modeling, synthesis, segmentation, and classification. In practice, the computational complexity arising in the estimation of model parameters is the primary problem.

Signal processing based methods try to compute texture features from filtered images by using signal processing techniques. Signal processing in the spatial domain is the most direct way to capture image texture properties. For example, simple edge detection masks, such as Robert, Sobel, and Laplacian operators (Gonzalez and Woods, 2002), can be applied to the image to perform image convolution, thus highlight edges. The edge density computed from the filtered image can be used to distinguish fine and coarse texture patterns since fine textures tend to have a higher density of edges per unit area than coarse textures. Texture analysis can also be

analyzed in the image frequency domain by performing Fourier transform (Chan and Pang, 2000). The Fourier transform is an analysis of the global frequency content in the signal, however the resolution in spatial domain is totally lost. Many applications require the analysis to be also localized in the spatial domain. This is usually handled by making a compromise in the resolutions between the spatial and frequency domains, which results in the widely used Gabor transform (Lee, 1996; Kumar and Pang, 2002) and the wavelet transform (Daubechies, 1992). The Gabor representation is optimal in minimizing the joint uncertainty in space and frequency. Each Gabor filter is characterized by an orientation or spatial frequency, which generates the Gabor convolution energy measure of an image. It is particularly efficient in detecting the frequency channels and orientations of texture pattern with high precision. Gabor filters have been successfully used in various applications, such as fabric defects detection (Kumar and Pang, 2002), face recognition (Zhou and Wei, 2006), and on-road vehicle detection (Sun, et al., 2005). The 2-D wavelet transform performs a spatial-frequency analysis on an image by repeatedly decomposing the image in the lower frequency sub-bands (Starck, et al., 1998). The rationale behind this spatial-frequency joint representation is to cut the image of interest into several parts using sets of scaleable modulated windows and then analyze the parts separately. Analyzing an image in this way gives us detailed information about the image under different scales. Typical applications of wavelet transform based texture analysis methods include content-based image retrieval (Kokare, et al., 2005), medical image classification (Khademi and Krishnan, 2007), and texture image segmentation (Charalampidi and Kasparis, 2002).

All texture analysis methods introduced above have their own application areas. In a typical texture classification application, texture features are first extracted by using texture analysis methods to characterize each data pattern. A machine learning algorithm uses these feature set to find a function (or classifier) that is generalized from the data so that the classifier is able to predict the classes for new instances. Usually, there is more than one texture analysis method that can be used in a specific application, thus the resulting feature vector will increase significantly depending on the number and the type of methods used. In theory, the more features extracted the greater possibility that each pattern can be correctly classified in the infinite data set limit. However, the data size is always limited in reality, and the classifier must be estimated on the basis of the limited amount of available data. In these cases, too many features can make the estimated classifier unreliable or even lose any practical use, which is the so-called *curse of dimensionality* (Bishop, 2006). Feature selection is one important means to attack this problem and a brief review of current feature selection methods is given in the following section.

3.5 *Feature Selection*

Feature selection is a process that chooses an optimal subset of features according to a certain criterion (Liu and Motoda, 1998). The criterion specifies the details of measuring feature subsets, and the choice of which can be influenced by the purpose of feature selection. For example, the criteria can be measured in terms of whether selected features help improve the classifier's accuracy. Feature selection has many benefits, such as reducing the measurement and storage requirements in the next round of data collection by removing irrelevant and redundant features,

improving data quality, making machine learning algorithms work faster, defying the curse of dimensionality to ensure more reliable model generation, and facilitating data visualization and understanding (Guyon and Elisseeff, 2003).

Three general approaches of feature selection exist: filters, wrappers and embedded methods (Guyon, et al., 2006).

Filter methods (Almuallim and Dietterich, 1994; Kira and Rendell, 1992) filter out features on the basis of their relevance or discriminate power with regard to the targeted classes. The filtering process is independent of any machine learning algorithm and only the resulting best feature subset is used to train the machine learning algorithm [Figure 3.7(a)].

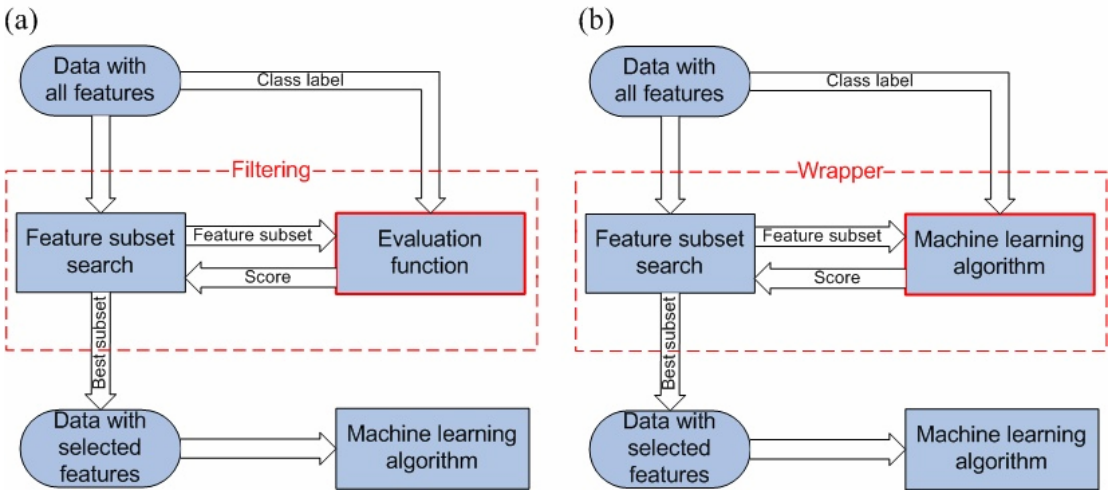


Figure 3.7 Filter and wrapper feature selection procedures. (a) Filter approach; (b) Wrapper approach. Filter and wrapper methods differ mostly by the evaluation criterion. Filters use criterions do not involve any learning algorithm, while wrappers use the performance of a learning machine trained with a given feature subset.

Wrapper methods (John et al. 1994; Kohavi and John, 1997) wrap feature selection around a specific classifier and use the classifier's accuracy to directly judge

the usefulness of a feature. In similarity to the filter model, a classifier is learned on the training data with the selected features [Figure 3.7(b)].

As we can see, filters and wrappers differ mostly by the evaluation criterion (highlighted red boxes in Figure 3.7). Filters use criteria that do not involve any learning algorithm, while wrappers use the performance of a learning machine trained with a given feature subset. Comparing with filters, wrapper methods may achieve better classification performance since their feature selection procedure already take the classifier's performance into account. However, the filter approach is computationally less expensive since measuring relevance (or other criterion such as distance and information gain) is usually cheaper than measuring accuracy of a classifier (Liu and Motoda, 1998).

Embedded methods (Weston, et al., 2001, 2003; Guyon, et al., 2002; Bradley and Mangasarian, 1998; Rakotomamonjy 2003) differ from the two former feature selection methods in the way feature selection and learning interact. As mentioned before, filter methods do not incorporate learning, while wrapper methods use a learning machine to score feature subsets according to their predictive performance, without incorporating knowledge about the specific structure of the learning algorithm. Embedded methods, however, do not separate the learning from the feature selection and perform feature subset generation and evaluation in the process of training an algorithm (Figure 3.8). By treating the learning machine as a "black box", wrappers can generally be combined with any kind of learning machines, while embedded methods are usually specific to given learning machines.

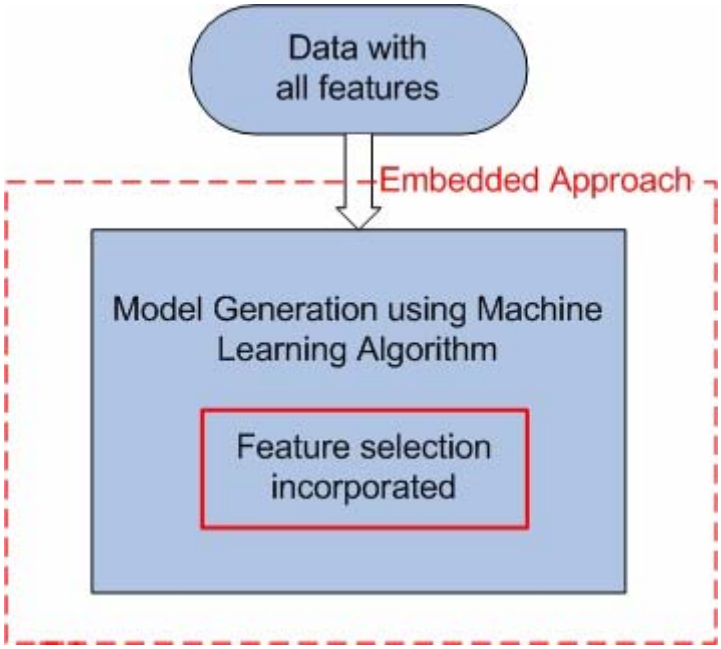


Figure 3.8 Embedded feature selection procedures.

Every family of feature selection methods (filter, wrapper and embedded) has its own advantages and drawbacks. In general, filters are less computational expensive than wrappers and embedded methods. Wrappers and embedded methods, however, have higher capacity and are more powerful. Compared to wrappers, embedded methods are less universal and more complicated in structure, while they usually make better use of the available data and can reach a solution faster (Guyon, et al., 2002). There are also theoretical arguments showing that filers are less prone to overfitting than wrappers and embedded methods. If only a small amount of data are available, filters usually perform better. Embedded methods and wrappers will eventually outperform filters as the number of training points increase (Guyon, et al., 2006).

Removing redundant or irrelevant features results in an improved classification performance and the extent of which depends on the feature selection

methods and the pattern classification or pattern recognition algorithms used. In the next section, we will briefly review pattern recognition methods.

3.6 *Pattern Recognition*

Pattern recognition is the study of how machines can observe the environment, learn to distinguish patterns of interest from their background, and make sound and reasonable decisions about the categories of the patters (Jain, et al., 2000). Automatic machine recognition, description and classification of patterns are important problems in a variety of engineering applications. The best known approaches for pattern recognition can be divided into four types: template matching, syntactic or structural matching, statistical classification, and neural networks (Jain, et al., 2000).

Template matching is one of the simplest approaches for pattern recognition. The matching is performed by comparing an unknown pattern to templates using a similarity measure, taking into account all allowable scale, pose (translation and rotation), and illumination changes, and possible object deformations. Templates or prototypes can be user specified or learned from a training set and the type of which depends on the specific applications. It can be points, curves, a 2-D shape or even a 3-D object. The basic way of template matching is to loop through all the pixels in the search image and compare them to the template. While the method is simple to implement, it has high computational demands. One approach to alleviate the problem is to filter the image in the frequency domain since the correlations between the template and the scene image can be computed through an implementation of the fast Fourier transform (FFT) algorithm (Bandeira, et al., 2007). Another method is to perform searching in a hierarchical way, where low-resolution versions of the

template are compared against low-resolution versions of the image, to find the location of a good match. This coarse-to-fine template matching usually requires much less computation compared to the full-resolution search. However, the typical coarse-to-fine template matching does not always find the global best match, thus the reduction in computation comes at the price of lower precision (Gharavi-Alkhansari, 2001).

The syntactic approach takes the view that a pattern is composed of simpler subpatterns, which are themselves built from even simpler subpatterns (Basu, et al., 2005). The simplest subpatterns are known as *primitives* and a complex pattern is represented in terms of the interrelationship between these primitives. A formal analogy between the structures of patterns and the syntax of a language is used to establish the foundation for syntactic pattern recognition. The patterns are viewed as sentences in a language, while the primitives are viewed as the alphabet of the language. The sentences are generated based on the grammar of a language. The primitive set is extracted from the training data and the grammar rules can also be inferred from the sample patterns. The idea that a large collection of complex patterns can be described by a small number of primitives and grammatical rules makes this approach intuitively appealing. Syntactic pattern recognition can be used in situations where the patterns have a definite structure which can be captured in terms of a set of rules. One example of this is the electrocardiogram (ECG) pattern recognition. ECG waveforms can be approximated with primitives such as peaks, straight line segments and parabolic segments. If normal and unhealthy ECG waveforms can be described as formal grammars, measured ECGs can be classified as healthy or unhealthy by first

describing them in term of the primitives and then trying to parse the descriptions according to the grammars (Trahanias and Skordalakis, 1990; Tümer, et al., 2003).

Statistical methods have been most intensively studied and used in practice. In statistical pattern recognition, a pattern is represented in terms of k features and is viewed as a point in a k-dimensional space. Given a set of training patterns from different classes, the objective is to establish decision boundaries in the feature space which separate patterns belonging to each class. The decision boundaries can be obtained either directly (geometric approach) or indirectly (probabilistic approach). The probabilistic approach requires firstly estimating class-conditional probability density functions and then constructing the discriminant functions which specify the decision boundaries (Devroye, et al., 1996). The class-conditional density can be estimated either parametrically or non-parametrically depending on whether the form of the class-conditional density function is known beforehand. If the form of the class-conditional density is known, but some of the parameters of the density are unknown, the density function can be estimated in a parametric way. Otherwise, nonparametric procedures can be used with arbitrary distributions and without the assumption on the forms of the underlying density function (Duda, et al., 2000). After the class-conditional density functions have been estimated, a number of well-known decision rules are available to define the decision boundary, such as the Bayes decision rule, the maximum likelihood rule and the Neyman-Pearson rule (Duda, et al., 2000). On the other hand, the geometric approach directly specifies a parametric form of the decision boundary (e.g., linear or quadratic) and then constructs the decision boundary by optimizing certain error criterion. A classical example of this

type is the Fisher linear discriminant, which projects high-dimensional data onto a line and performs classification in this one-dimensional space yielding the maximum ratio of between class scatter to within-class scatter (Duda, et al., 2000). Another example is the decision tree (Quinlan, 1993) which is trained by iterative selection of individual features that are most salient at each tree node. The decision boundaries generated are parallel to the feature axes since only one feature is used at each node. Support vector classifier (Vapnik, 1998; Cristianini and Shawe-Taylor, 2000), one of the most popular classifier, can also be grouped into this type. The method consists of mapping the input vector into a higher dimensional space and constructing an optimal hyper-plane to achieve maximum separation between different classes in this space.

Artificial neural networks are elaborate computer models of neurons constructed by computational neurobiologists in order to simulate the functions of human brain. They can be viewed as massively parallel processing systems consisting of large quantities of simple processors (neurons) with many interconnections (weights between neurons). Neural networks have emerged as an important tool for classification, due to their ability to learn complex nonlinear input-output relationships, and the ability to adjust themselves to the data through training. The most commonly used family of neural networks for pattern classification include multilayer perceptron (MLP), radial basis function (RBF) network, and self-organizing map (SOM) (Bishop, 1995; Kohonen, 2001). Both MLP and RBF are feed-forward networks which are organized into layers and have unidirectional connections between the layers. SOM, on the contrary, contains only one layer of neurons (except the input nodes). During the network training procedure, weights

between neurons or even the network structure are updated so that a trained network can efficiently perform a specific classification task. Neural networks have gained increasing attention since they usually do not require much domain-specific knowledge and can be treated as a black box, enabling "automatic" mapping between inputs and outputs.

The above four pattern recognition approaches are not necessarily independent. Attempts have also been made to design hybrid systems involving multiple methods (Kuncheva, 2004). A number of experimental studies have shown that the pattern recognition or classification accuracy can be improved by combining a set of classifiers. However, to train a set of classifiers may require more training data and more computation time. Also, the classifier combination results are not always better than those of the best individual classifier (Jain, et al., 2000).

Chapter 4: Geometric Calibration of a Combined X-ray and Laser Range Imaging System

As mentioned in chapter 1, the detection of bone fragments and other physical hazards in deboned poultry meat has become increasingly important in food quality and safety. The poultry physical contamination detection problem solution depends largely on how well the thickness variation is neutralized. In this research, a thickness estimation method based on laser range imaging was adopted to measure the poultry profile in real-time, which leads to the possibility of thickness compensation.

4.1 *Principle of Thickness Compensation for Physical Contamination Detection*

Figure 4.1 illustrates the conceptual procedure of thickness compensation for easy physical contamination detection. Figure 4.1(a) shows a piece of poultry meat with bone fragments inside. The associated X-ray image intensity profile is shown in Figure 4.1(b). Generally, the intensity of objects in an X-ray image reflects the X-ray absorption, which is determined by the physical characteristics of the material and its thickness. The intensity of the imagery I_X can be described using an exponential law (Dowsett, et al., 2006) as follows:

$$I_X = I_0 \exp(-\mu d), \tag{4.1}$$

where I_0 is the incident X-ray intensity, μ is the X-ray attenuation coefficient of the material (assuming the material is homogenous), and d is the length of the X-ray pathway through the material.

As we can see from equation (4.1), the thickness variation leads to grayscale variations in the X-ray image. In other words, for a homogenous object, if its X-ray image is darker in some location, it also means the object is thicker at that place. However, since the X-ray attenuation coefficients are different for meat tissue and bone fragments (or other physical contaminants), the X-ray image intensity of a piece of meat with uniform thickness but also embedded with bones will not be constant within the image. So due to the thickness variation of poultry meat, it is very difficult to detect the physical contaminations inside accurately by checking the X-ray image alone. To address this problem, Tao and Ibarra (2000A) suggested acquiring the thickness information of the inspected objects, and then compensating for the thickness effects in the X-ray images.

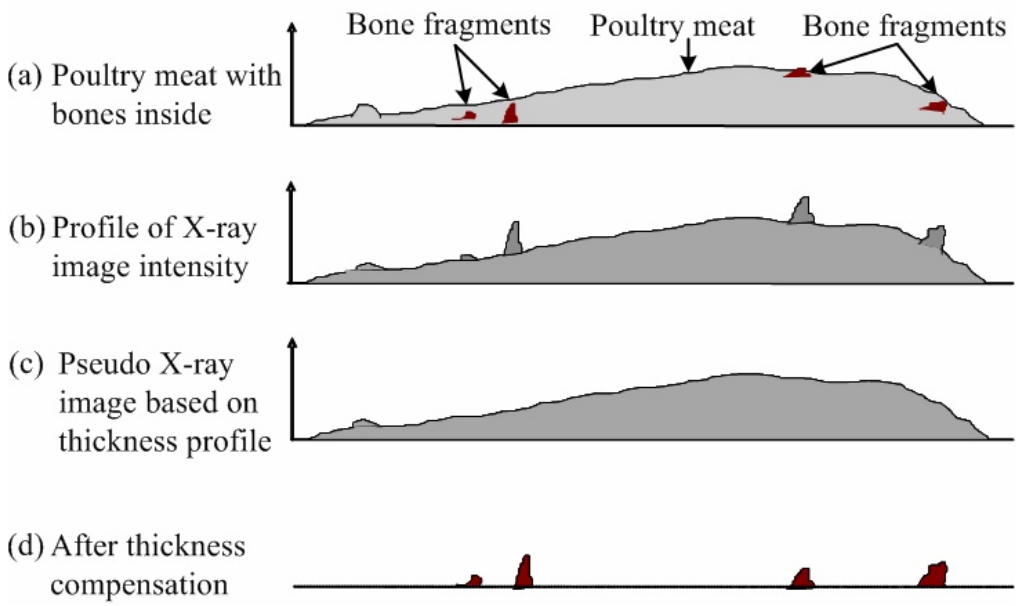


Figure 4.1 Principle of thickness compensation for physical contamination detection. (a) a piece of poultry meat with bones inside; (b) profile of the X-ray image intensity; (c) pseudo X-ray imagine generated based on the thickness profile; (d) thickness compensated image (subtract the pseudo X-ray image from the X-ray image). Notice bone fragments can be easily detected after the thickness compensation

The surface topography $D(x,y)$ is obtained using laser range imaging, and further mapped to a pseudo X-ray image $I_{pX}(x,y)$ [Figure 4.1(c)] under the assumption that the meat is homogenous. Because both X-ray and laser images are functions of the thickness, a mapping relationship exists between them. The transformation function f_{XL} can be determined by constructing a model between the X-ray image and laser image (Jing, 2003; Xin, 2003):

$$I_{pX}(x,y) = f_{XL}(D(x,y)). \tag{4.2}$$

Thus, the grayscale variation in the X-ray image due to uneven meat thickness can be compensated by subtracting the pseudo X-ray image from the acquired X-ray image:

$$I_{tc}(x,y) = I_X(x,y) - I_{pX}(x,y), \tag{4.3}$$

where $I_{tc}(x,y)$ is the resultant thickness compensated image [Figure 4.1(d)]. By combining the mapped pseudo X-ray image with the original X-ray image, the physical contaminants detection inside meat becomes much easier.

The overall system performance depends on the accuracy of the information provided by the individual subsystems. Due to the unique synergism of X-ray imaging and laser range imaging in this research, special calibration methods were developed to ensure the combined systems work seamlessly. Two types of cameras, four CCD cameras and one X-ray line-scan camera were utilized in our combined system. The calibration of each kind of camera and the method to synthesize the calibration results will be introduced in the following sections.

4.2 *CCD Camera Calibration*

The key idea for CCD camera calibration is to link the known coordinates of a set of 3-D points and their projections on the CCD camera and then to estimate camera parameters (camera focal length, lens distortion, etc.) based on that information. To perform a camera calibration, a camera model needs first to be established to serve as the foundation for the following calibration procedures.

4.2.1 *CCD Camera Model*

The geometry of the widely used pinhole camera model is shown in Figure 4.2. In this camera model, an object point $M = (X_w, Y_w, Z_w)$ is transformed from the world coordinate system to the camera coordinate system, (X_c, Y_c, Z_c), in order to perform perspective projection. This transformation consists of a translation and a rotation, which can be formulated as

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + T, \tag{4.4}$$

where $R = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix}$ is the 3-D rotation matrix, and $T = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}$ is the 3-D translation vector.

Through perspective projection, the ideal (undistorted) image coordinate of the object point (x_u, y_u) is expressed as

$$\begin{bmatrix} x_u \\ y_u \end{bmatrix} = \frac{f}{Z_c} \begin{bmatrix} X_c \\ Y_c \end{bmatrix}, \tag{4.5}$$

where f is the effective focal length.

The above projection is a mapping from the Euclidean 3-D space to Euclidean 2-D space. The center of the projection is called the *camera center*, also known as the *optical center*. The line from the camera center perpendicular to the image plane is called *principal axis*, and the point where the principal axis meets the image plane is called the *principal point*.

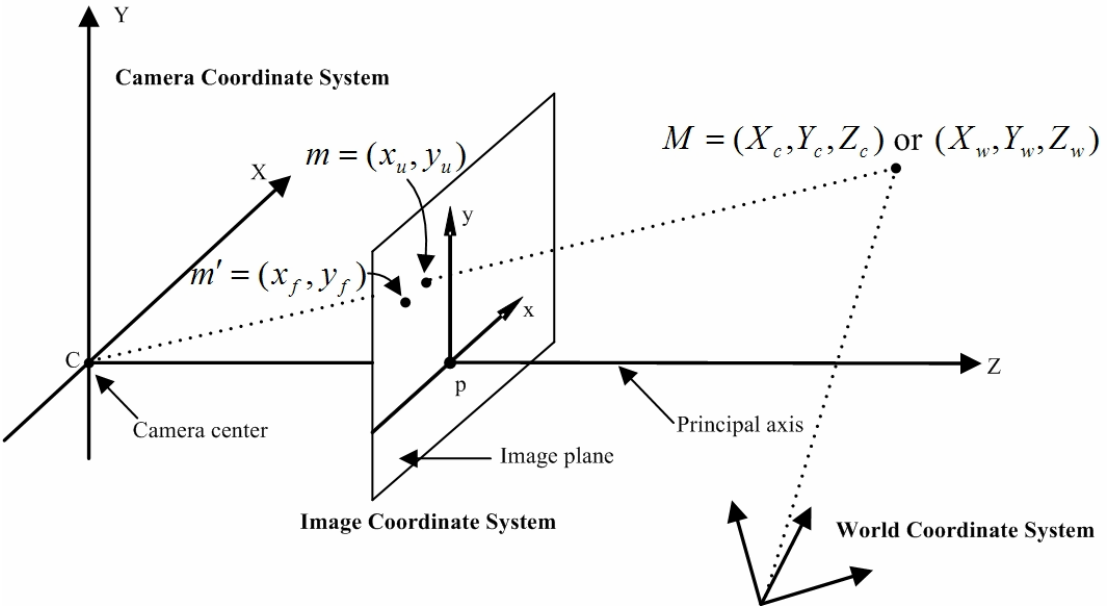


Figure 4.2 Pinhole camera geometry, where C is the camera center and also the origin of the camera coordinate system, p is the principle point, M is a point in the space that has different coordinates in the world and camera coordinate systems, and m is the image of M in the image plane that is placed in front of the camera center.

The camera model (4.5) is based on an ideal optical system. However, common cameras often display somewhat distorted projection behavior. Two kinds of lens distortion, as shown in Figure 4.3, are usually modeled. One is radial distortion, which causes the actual image points to be displaced radially in the image plane. This type of distortion is mainly caused by flawed radial curvature of the lens elements.

The other is tangential distortion, which causes the displacement of image points perpendicular to a radius from the center of the field. The tangential distortion is caused by the lens elements not being perfectly aligned in the lens itself (Heikkila, 2000). Research shows that for most industrial machine vision applications, only radial distortion needs to be considered, and only one term is needed (Tsai, 1987). In this case, the relationship between the distorted image coordinates (x_d, y_d), and the ideal image coordinates (x_u, y_u), can be represented as

$$\begin{bmatrix} x_u \\ y_u \end{bmatrix} = (1 + k_1 r^2) \begin{bmatrix} x_d \\ y_d \end{bmatrix}, \tag{4.6}$$

where $r = \sqrt{x_d^2 + y_d^2}$, k_1 is the first order lens radial distortion coefficient.

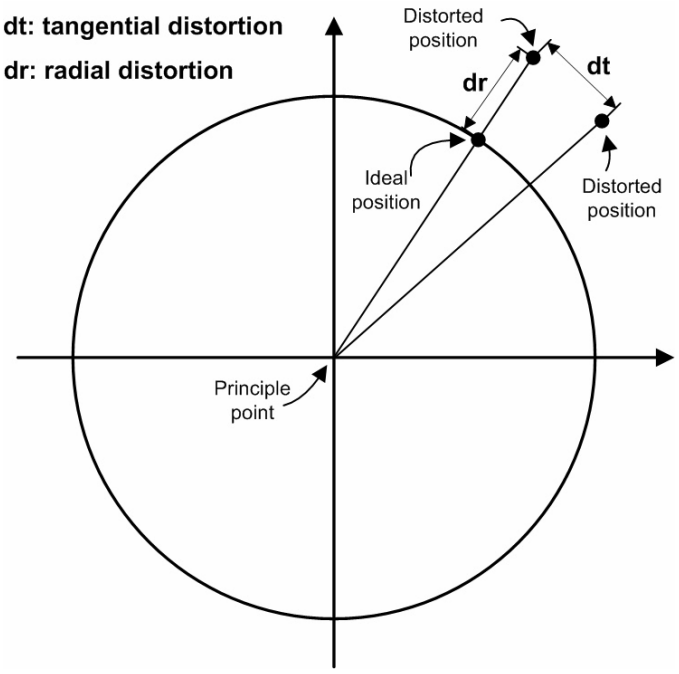


Figure 4.3 Two types of camera lens distortion: radial distortion and tangential distortion. dt in the figure is caused by tangential distortion while dr is caused by radial distortion.

In practice, the image in the image plane will be further sampled by a frame grabber card and stored in an image memory buffer. Usually the origin of the image buffer coordinate system is located in the lower left corner (see Figure 4.4). Since the unit of the image buffer coordinates is the pixel, coefficients d_x and d_y are needed to change the distance units into pixels. Here, d_x and d_y are the center-to-center distances between adjacent CCD sensor elements in the row and column directions, respectively. The 2-D image buffer coordinates (x_f, y_f), can be obtained by

$$\begin{bmatrix} x_f \\ y_f \end{bmatrix} = \begin{bmatrix} d_x^{-1} s_x x_d \\ d_y^{-1} y_d \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}, \tag{4.7}$$

where s_x is the horizontal scaling factor, and (x_0, y_0) are the coordinates of the principal point in the image buffer coordinate system. In our application, the horizontal scaling factor s_x is set to one and the initial value of the principal point, (x_0, y_0), is assumed to be at the image buffer center.

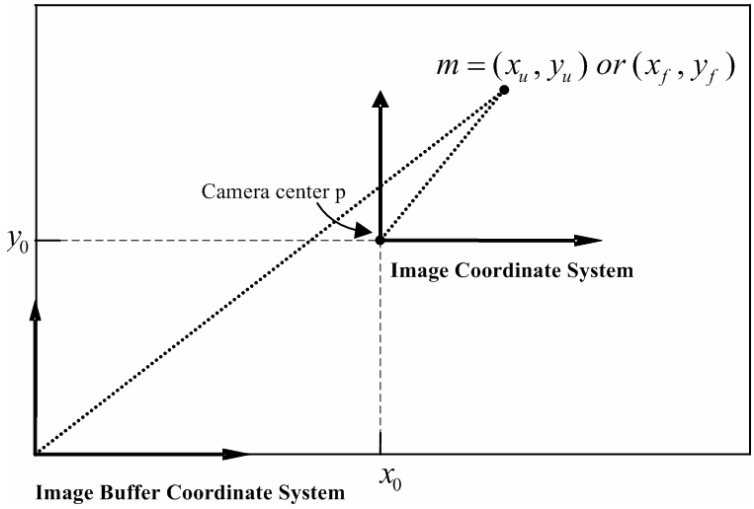


Figure 4.4 The relationship between the image coordinate system and the image buffer coordinate system, where (x_u, y_u) are the image coordinates of

point m, (x_f, y_f) are the corresponding image buffer coordinates, and (x_0, y_0) are the coordinates of the principle point in the image buffer coordinate system.

From equations (4.4) to (4.7), the relationship between a point (X_w, Y_w, Z_w) in the world coordinate system and its image (x_f, y_f) in the image buffer can be represented as

$$d_x(x_f - x_0)(1 + k_1 r^2) = f\,\frac{r_1 X_w + r_2 Y_w + r_3 Z_w + T_x}{r_7 X_w + r_8 Y_w + r_9 Z_w + T_z}, \tag{4.8a}$$

$$d_y(y_f - y_0)(1 + k_1 r^2) = f\,\frac{r_4 X_w + r_5 Y_w + r_6 Z_w + T_y}{r_7 X_w + r_8 Y_w + r_9 Z_w + T_z}, \tag{4.8b}$$

where $r = \sqrt{(x_f - x_0)^2 d_x^{\,2} + (y_f - y_0)^2 d_y^{\,2}}$.

4.2.2 *CCD Camera Calibration Using a Monoview Coplanar Set of Points*

Physical camera parameters can be categorized as *extrinsic* parameters, which represent the 3-D position and orientation of the camera, and *intrinsic* parameters, which model the camera's optical setting and characteristics. In the CCD camera model employed in this study, the intrinsic parameters include focal length f, principal point (x_0, y_0), and radial distortion coefficient k_1. The CCD pixel sizes, d_x and d_y, are available from the manufacturer. The extrinsic parameters include the rotation matrix R and the translation vector T. The objective of the camera calibration procedure is to determine optimal values for these parameters based on image observations of a known 3-D target.

As mentioned in the introduction section, the laser range imaging subsystem works to capture the profile of scanned objects. Laser light in conjunction with a cylindrical lens is used to generate the laser light plane, which has been arranged to overlap with the X-ray beam plane when the machine is set up. The intersection of the laser beam with the object surface is visible as a light stripe in the image grabbed by the CCD camera. The thickness variation of the scanned object is registered by the shift in the stripe lines (Figure 4.5). Since all the stripes grabbed by the CCD camera are within the laser light plane, the CCD camera can be calibrated using monoview coplanar set of points according to Tsai's method (1987). It is worth to mention that the monoview of the calibration points should be taken when these points are all sitting in the plane of the laser fan beam. This is achieved by synchronizing the X-ray and CCD camera image acquisition during system calibration and also by using a specially designed calibration model (Figure 4.7a), which will be introduced in detail in a later section.

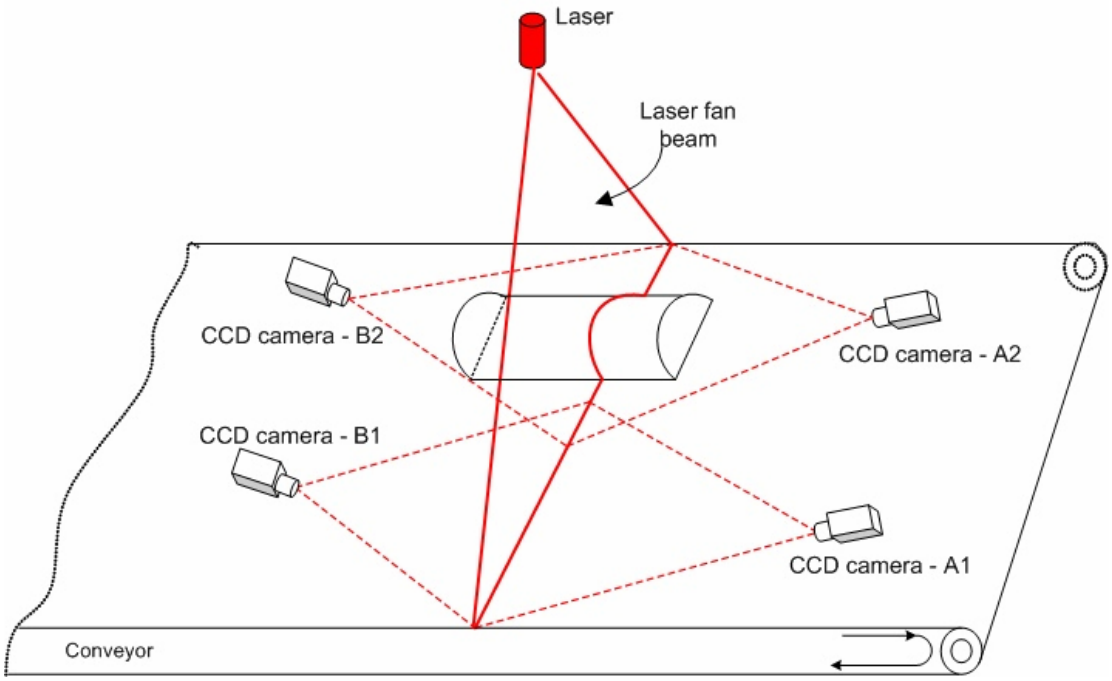


Figure 4.5 Laser range imaging system setup. The thickness variation of the scanned object is registered by the shift of the stripe lines in the CCD camera image plane.

Because the calibration points are on a common plane, the (X_w, Y_w, Z_w) coordinate system is defined such that $Z_w = 0$. The CCD camera calibration method requires at least five accurately detected calibration points, whose corresponding geometric configurations are known. The seven processing steps of this method are explained below.

(1) *Compute the distorted image coordinates* (x_d, y_d) *from the image buffer coordinates* (x_f, y_f).

First, an image of the calibration object is taken. The feature point coordinates (x_f, y_f) are extracted from the image. The horizontal and vertical distances between two sensor elements, d_x and d_y, are obtained from the

manufacturer data sheet. For the CCD camera used in this research (Pulnix TM-6740CL, California, USA), d_x and d_y are both 7.4 μm. The corresponding image coordinates, (x_d, y_d), are calculated for all extracted calibration feature points (x_f, y_f) by

$$\begin{aligned} x_d &= d_x(x_f - x_0), \\ y_d &= d_y(y_f - y_0). \end{aligned} \tag{4.9}$$

(2) *Calculate the five parameters* $\left(T_y^{-1}r_1,\ T_y^{-1}r_2,\ T_y^{-1}T_x,\ T_y^{-1}r_4,\ T_y^{-1}r_5\right)$ *for the transformation of image coordinates into world coordinates.*

The projection of calibration points (X_w, Y_w, Z_w) in the world coordinate system onto corresponding image points (x_d, y_d) in image coordinates is characterized by the following linear equation [Appendix I of Tsai's paper (1987)]:

$$\begin{bmatrix} y_d X_w & y_d Y_w & y_d & -x_d X_w & -x_d Y_w \end{bmatrix} \cdot L = x_d \ , \tag{4.10}$$

where $L = \begin{bmatrix} T_y^{-1}r_1 & T_y^{-1}r_2 & T_y^{-1}T_x & T_y^{-1}r_4 & T_y^{-1}r_5 \end{bmatrix}^T$ for $T_y \neq 0$. Here T_x and T_y are the components of the translation vector T in x and y axis directions, and r_1, r_2, r_4, and r_5 are elements of the rotation matrix R. With more than five calibration points, an over determined system of linear equations can be established and solved for L using the pseudo-inverse technique (Hartley and Zisserman, 2000).

(3) *Calculate* T_y *from L.*

From step (2), we already have all the elements of the vector L. We now define a new matrix C as

$$C \equiv \begin{bmatrix} r_1' & r_2' \\ r_4' & r_5' \end{bmatrix} \equiv \begin{bmatrix} r_1 / T_y & r_2 / T_y \\ r_4 / T_y & r_5 / T_y \end{bmatrix} . \tag{4.11}$$

If there is no whole row or column of C that vanishes (the value of matrix elements equal zero), the magnitude of T_y is calculated by

$$| T_y | = \sqrt{\frac{S_r - [S_r^2 - 4(r_1'r_5' - r_2'r_4')^2]^{1/2}}{2(r_1'r_5' - r_2'r_4')^2}} \tag{4.12}$$

with $S_r = r_1'^2 + r_2'^2 + r_4'^2 + r_5'^2$. Otherwise, $| T_y | = (r_i'^2 + r_j'^2)^{-1/2}$, where r_i' , r_j' are the elements in the row or column of C that do not vanish (Tsai, 1987). The sign of T_y can then be determined by first projecting a point in the world coordinate system (the image coordinate of this point should be far away from the principal point) into the image plane and then comparing the projected with the original position.

(4) *Compute the 3-D rotation matrix R and the x component of the translation vector T_x.*

Once T_y is known, the rotation matrix elements r_i ($i = 1, 2, 4, 5$) and the x component of the translation vector, T_x, become easy to solve for:

$$r_i = r_i' T_y \ (i = 1, \ 2, \ 4, \ 5), \tag{4.13}$$

$$T_x = \left(\frac{T_x}{T_y} \right) T_y \ . \tag{4.14}$$

The still missing components r_3, r_6, r_7, r_8, and r_9 can be calculated based on the orthonormal property of the rotation matrix R:

$$R^T R = I = RR^T, \tag{4.15}$$

$$|R| = 1. \tag{4.16}$$

(5) *Calculate focal length f and the z component of the translation vector, T_z, ignoring the lens distortion.*

By setting k_1 to zero in Equation 4.8(b), for each calibration point we can establish the following linear equation with f and T_z as unknowns:

$$[Y_c \quad -d_y(y_f - y_0)]\begin{bmatrix} f \\ T_z \end{bmatrix} = wd_y(y_f - y_0), \tag{4.17}$$

where $Y_c = r_4 X_w + r_5 Y_w + r_6 \cdot 0 + T_y$ and $w = r_7 X_w + r_8 Y_w + r_9 \cdot 0$. Again, if more than two calibration points are used, this over determined set of equations can be solved using pseudo-inverse technique:

$$\begin{bmatrix} f \\ T_z \end{bmatrix} = (W^T W)^{-1} W^T Q, \tag{4.18}$$

where W and Q are calculated using N calibration points with

$$W = \begin{bmatrix} Y_{c1} & -d_y(y_{f1} - y_0) \\ Y_{c2} & -d_y(y_{f2} - y_0) \\ \vdots & \vdots \\ Y_{cN} & -d_y(y_{fN} - y_0) \end{bmatrix} \text{ and } Q = \begin{bmatrix} wd_y(y_{f1} - y_0) \\ wd_y(y_{f2} - y_0) \\ \vdots \\ wd_y(y_{fN} - y_0) \end{bmatrix}.$$

(6) *Estimate the radial distortion k_1, and refine the estimation for f and T_z.*

The utilization of a standard optimization technique allows more accurate calculations of the camera constant f, the depth T_z, and the distortion coefficient k_1. These three parameters can be calculated by minimizing the following function:

$$\sum_{i=1}^{N} \| m_{ui}(f, T_z) - \tilde{m}_{ui}(k_1) \|^2 , \tag{4.19}$$

where $m_{ui}(f, T_z)$ is the undistorted projection of the point (X_{wi}, Y_{wi}, Z_{wi}) in camera coordinate calculated from equations (4.4) and (4.5), while $\tilde{m}_{ui}(k_1)$ is the undistorted camera coordinate for the image point (x_i, y_i), recovered by using equations (4.7) and (4.6). Basically we are projecting both the points in world coordinates and the points in image frame coordinates to image coordinates, to check for coordinate discrepancies. Equation (4.19) is a nonlinear minimization problem, which is solved with the Levenberg-Marquardt algorithm (Gill and Murray, 1978; Ranganathan, 2004). It requires an initial guess of f and T_z, which can be obtained from step (5). Zero is assumed as the initial value for the radial distortion coefficient k_1.

(7) *Calculate optimal solutions for* f, R, T, k_1, x_0, y_0.

In this last step, all the parameters are further refined by minimizing the function

$$\sum_{i=1}^{N} \| m_{ui}(f, R, T) - \tilde{m}_{ui}(k_1, x_0, y_0) \|^2 , \tag{4.20}$$

where all the unknown parameters (f, R, T, k_1, x_0, y_0) are as defined in the previous sections. Again Levenberg-Marquardt algorithm is used to solve this nonlinear minimization problem. Initial guesses of these parameters are obtained from the results of steps (1) to (6). Better results can be achieved after this step, since all the camera parameters are taken into account for the optimization.

4.3 *X-ray Line-Scan Camera Calibration*

For the X-ray imaging subsystem, X-ray images were collected by a high-resolution line-scan digital X-ray camera (Hamamatsu, C8750-10FC, Japan) featuring 1280 pixels per line at 0.4 mm resolution.

4.3.1 *X-ray Line-Scan Camera Model*

Geometrically, the X-ray line-scan camera can also be approximated by the pinhole camera model as shown in Figure 4.6. The distance between the X-ray source and the detector is considered the effective focal length, f', of the X-ray camera; the image plane is where the detector is located; c' is the X-ray focal spot, which is also the camera center and the origin of the camera coordinate system; p' is the principal point, which is the intersection of the principal axis and the image plane. The gray rectangle represents an object interposed between the X-ray source and the X-ray detector.

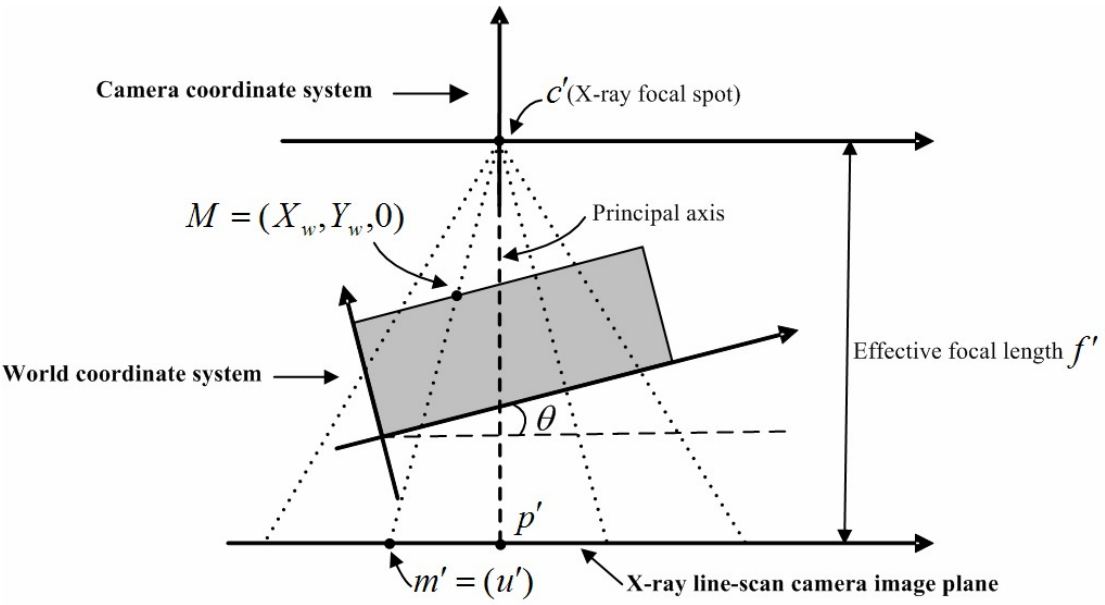


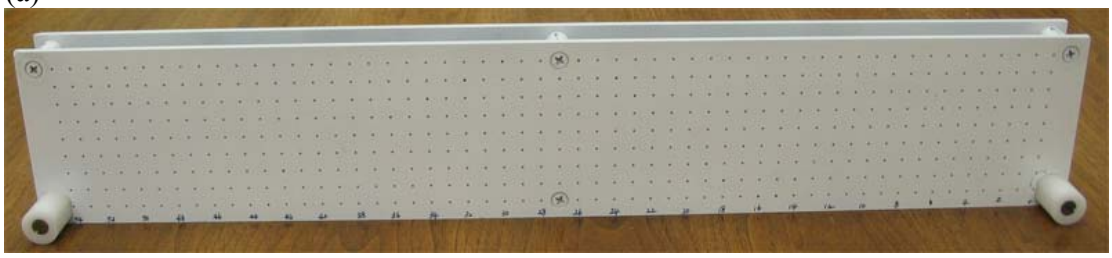
Figure 4.6 Pinhole camera geometry for an X-ray camera. Here c' is the X-ray focal spot which also is the camera center and the origin of the camera coordinate system; p' is the principal point; f' is the effective focal length; M is a point on the object surface, and m' is its image in the image plane (detector); θ is the angle between the conveyor belt and detector image plane.

To calibrate the X-ray camera, the 2-D pinhole model is used. As before in CCD camera calibration, four coordinate systems — world, camera, image, and image buffer — are needed to model the transformation between a feature point in the object and its image in the image buffer. We use homogenous coordinates for easy representation of geometrical transformations. Homogeneous coordinates form a basis for the projective geometry used extensively to project a 3-D scene onto a 2-D image plane. They also unify the treatment of common graphical transformations and operations. For a 3-D point $P = (X, Y, Z)$, its corresponding homogeneous coordinate are the components of a vector augmented by adding 1 as the last element: $\hat{P} = (X, Y, Z, 1)$. Similar changes were made to represent a point in 1-D or 2-D space. By representing points in homogenous coordinates, inversions or combinations of

linear transformations are simplified to inversion or multiplication of the corresponding matrices.

A special calibration object was made and shown in Figure 4.7. As shown in Figure 4.7(b), a set of fine steel pins are inserted at known geometrical positions. The X-ray beam will cut across these steel pins and cast shadows in the image plane. The correspondence between steel pins and shadows is used to calibrate the X-ray camera. The details of this method will be discussed below.

(a)



(b)

Figure 4.7 Photos of the calibration object: (a) side view; (b) top view with the steel pins inserted at known geometrical positions.

Without loss of generality, we assume the X-ray beam plane is on $Z = 0$ of the world coordinate system; thus the feature point $M = (X_w, Y_w, 0)$ is simplified as $M = (X_w, Y_w)$ in 2-D space (X-ray beam plane). By using homogenous coordinate representation, M is further represented as $M = (X_w, Y_w, 1)$. The relationship between the coordinates of point M and its image projection m' can be expressed as

$$s\begin{bmatrix} u' \\ 1 \end{bmatrix} = K[R \mid t]\begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} \tag{4.21}$$

with s is an arbitrary scale factor. Here $[R \mid t] = \begin{bmatrix} \cos\theta & \sin\theta & t_x \\ -\sin\theta & \cos\theta & t_y \end{bmatrix}$, called the *extrinsic matrix*, represents the rotation and translation that relate the world coordinate system to the camera coordinate system; $K = \begin{bmatrix} -f' & u_0' \\ 0 & 1 \end{bmatrix}$ is the camera intrinsic matrix; f' is the effective focal length; and u_0' is the principal point.

From equation (4.21), a linear mapping between the homogenous coordinates of a model point M and its image m' can be achieved by a homography H :

$$s\begin{bmatrix} u' \\ 1 \end{bmatrix} = H\begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} \tag{4.22}$$

with $H = K[R \mid t] = \begin{bmatrix} -f'\cos\theta - u_0'\sin\theta & -f'\sin\theta + u_0'\cos\theta & -ft_x + u_0't_y \\ -\sin\theta & \cos\theta & t_y \end{bmatrix}$.

4.3.2 *Estimate Homography between the Calibration Model Plane and its Image*

Let $M_i = (X_{wi}, Y_{wi}, 1)$ and $m_i' = (u_i', 1)$, $i = 1, 2, \ldots, n$, be sets of model and image calibration points, respectively. Ideally, each point pair should satisfy equation (4.22). In practice, however, due to the noise in the extracted image points, equation (4.22) cannot be satisfied. Assume m_i' is corrupted by Gaussian noise with mean zero and covariance matrix $\sum_{m_i'}$, then the maximum likelihood estimation of H can be obtained by minimizing the following function:

$$\sum_{i=1}^{n}(m_i' - \tilde{m}_i')\sum_{m_i'}^{-1}(m_i' - \tilde{m}_i')^T , \tag{4.23}$$

where $\tilde{m}_i'$ is the estimated image coordinates computed from equation (4.22), $\sum_{m_i'}$ is the covariance matrix of m_i', and n is the number of feature point pairs.

If feature points are extracted independently with the same procedure, we can simply assume a covariance matrix $\sum_{m_i'} = \sigma^2 I$ for all i. In this case, minimizing equation (4.23) is simplified to be the following equation:

$$\min_H \sum_{i=1}^{n} \| m_i' - \tilde{m}_i' \|^2 . \tag{4.24}$$

The minimization of equation (4.24) is conducted as follows.

Let us denote the elements of the 2×3 matrix H by h_1, h_2, …, h_6. Then equation (4.22) can be rewritten as

$$\begin{bmatrix} M & -u'M \end{bmatrix}\xi = 0 , \tag{4.25}$$

where $\xi = \begin{bmatrix} h_1 & h_2 & h_3 & h_4 & h_5 & h_6 \end{bmatrix}^T$.

If more than six point pairs are given, we have an over-determined matrix equation

$$V\xi = 0 \tag{4.26}$$

where $V = \begin{bmatrix} X_{w1} & Y_{w1} & 1 & -u_1'X_{w1} & -u_1'Y_{w1} & -u_1' \\ X_{w2} & Y_{w2} & 1 & -u_2'X_{w2} & -u_2'Y_{w2} & -u_2' \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{wn} & Y_{wn} & 1 & -u_n'X_{wn} & -u_N'Y_{wn} & -u_n' \end{bmatrix}$.

The solution is well known to be the right singular vector of V associated with the smallest singular value.

4.3.3 *Extraction of Calibration Parameters from the Homography Matrix H*

By using the homogeneous coordinates, the homography matrix H can be estimated from equation (4.26) only up to a scale factor, i.e., we can get the following equation:

$$H = \begin{bmatrix} -f'\cos\theta - u_0'\sin\theta & -f'\sin\theta + u_0'\cos\theta & -ft_x + u_0't_y \\ -\sin\theta & \cos\theta & t_y \end{bmatrix} = \lambda \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \end{bmatrix} \quad (4.27)$$

where λ is an arbitrary scale.

All the intrinsic and extrinsic parameters of the X-ray camera can be calculated from the preceding equation. The results are listed as follows:

$$\theta = \arctan\left(\frac{-h_4}{h_5}\right) \quad (4.28)$$

$$t_y = \frac{h_6}{\sqrt{h_4^{\,2} + h_5^{\,2}}} \quad (4.29)$$

$$u_0' = \frac{h_1h_4 + h_2h_5}{h_4^{\,2} + h_5^{\,2}} \quad (4.30)$$

$$f' = \frac{-(h_1 - u_0'h_4)}{h_5} \quad (4.31)$$

$$t_x = \frac{-(h_3 - u_0'h_6)}{f\sqrt{h_4^{\,2} + h_5^{\,2}}} \quad (4.32)$$

4.3.4 *Compensation of the X-ray Fan Beam Effect*

As previously mentioned, the laser subsystem is used to capture the 3-D profile of the scanned objects. In practice, the 3-D thickness image is mapped to a pseudo X-ray image that will further be subtracted from the real grabbed X-ray image to remove the thickness influence in the X-ray image, thus highlighting the suspicious areas (Jing, 2003; Xin, 2003).

To get this pseudo X-ray image from the laser 3-D profile, for each point in the profile, we need to project it to the X-ray camera image plane and calculate its corresponding position and intensity in that image plane. For example in Fig. 4.8, for any point M at the object surface, we want to know its projection position, m', in the X-ray camera image plane, as well as the pixel intensity that corresponds to the X-ray attenuation. The position on the detector of the image of any point can be obtained from equation (4.22), so only its intensity needs to be determined. Assuming the incident X-ray intensity is uniformly distributed and the scanned object is homogenous, the X-ray attenuation is affected by the length of the X-ray path through the object (D in Figure 4.8).

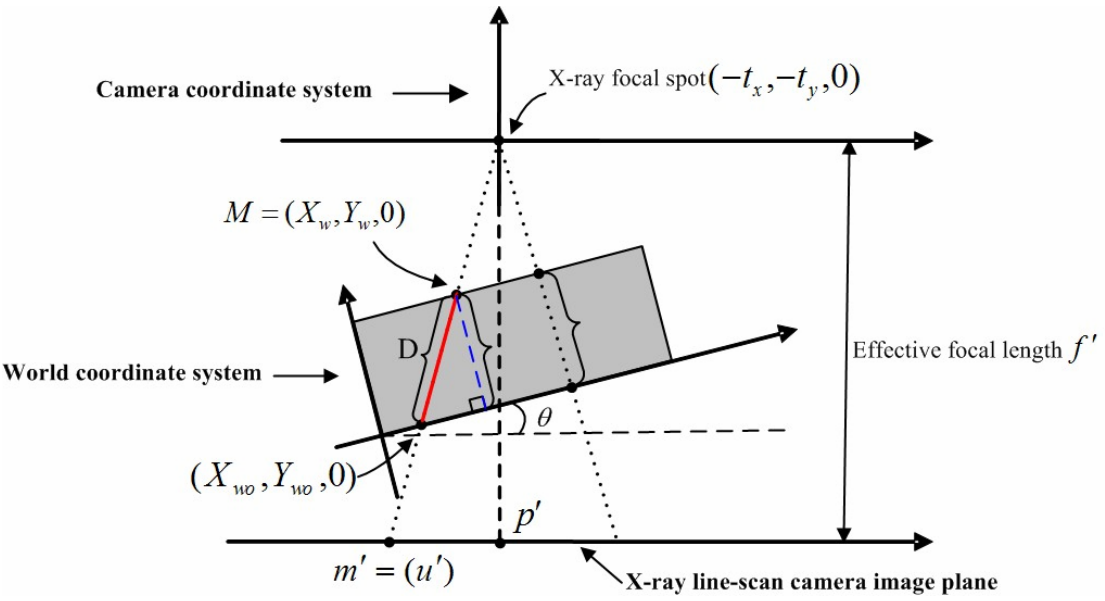


Figure 4.8 X-ray fan beam effect. The X-ray absorption is determined by the distance it passes through the object. Notice that the coordinates of the X-ray focal spot are $(-t_x,-t_y,0)$ in the world coordinate system.

As shown in Figure 4.8, the coordinates of the X-ray focal spot are $(-t_x,-t_y,0)$ in the world coordinate system. For any surface point $(X_w,Y_w,0)$ on the object, the line function of the X-ray passing through it can be determined by the coordinates of these two points (neglecting the *z* axis components):

$$y-Y_w=\frac{Y_w+t_y}{X_w+t_x}(x-X_w) \tag{4.33}$$

The coordinates for the point where the X-ray exits the object, $(X_{wo},Y_{wo},0)$, are determined by the intersection of the incident X-ray (equation 4.33) with the line $y=0$. So the X-ray path corresponding to point $(X_w,Y_w,0)$ on the object surface can be calculated as

$$D=\sqrt{(X_w-X_{wo})^2+(Y_w-Y_{wo})^2} \tag{4.34}$$

4.4 *Direct Mapping from CCD Camera to X-ray Line-Scan Camera*

The CCD camera and X-ray camera calibration methods have been introduced in detail in previous sections. After CCD camera calibration, every intrinsic and extrinsic parameter for the CCD camera is known. So for each pixel in the image frame of the CCD camera, it is easy to get the corresponding coordinates in the world coordinate system by solving equations (4.8a) and (4.8b). Also, since the homography H is known after X-ray camera calibration, it becomes straightforward to calculate the image coordinates (in the X-ray line-scan camera) for each point in the world coordinate system from equation (4.22). It is intuitive to combine the results of the two calibrations (Figure 4.9). So for each pixel in the CCD camera, we can directly calculate its coordinates in the X-ray line-scan camera image domain. From equation (4.34), the corresponding X-ray pathway can also be determined. It is worthy to mention that in a typical multi-modality imaging system, images acquired from different imaging modalities normally need to be registered together before any further image analysis. However, image registration is usually computationally expensive. For our dual-modality imaging system, the acquired laser range image can be directly mapped to the X-ray image domain after system calibration, without any complicated calculation. In addition, the X-ray fan beam effect can be compensated at the same time. This great time/computation saving is valuable especially for a real-time application.

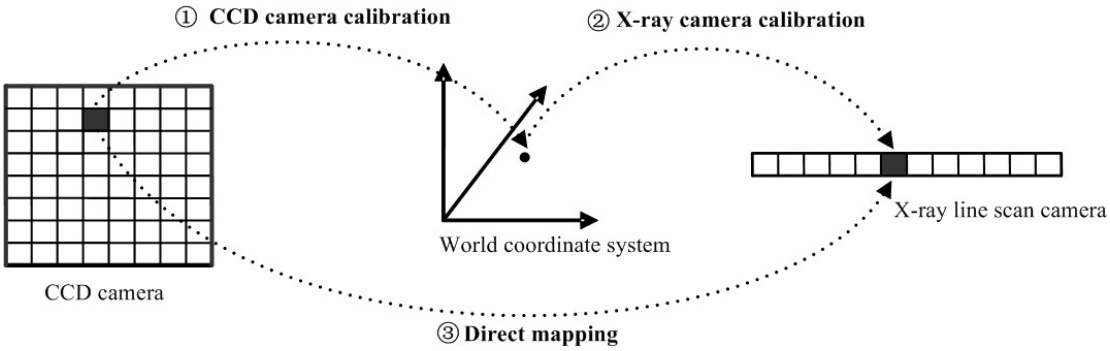


Figure 4.9 Direct mapping from CCD camera to X-ray line-scan camera. The mapping from CCD camera to world coordinates can be realized after CCD camera calibration. The mapping from world coordinates to X-ray line-scan camera can be realized after X-ray camera calibration. The direction mapping from CCD camera to X-ray line-scan camera can be achieved by combining those two calibration results.

4.5 *Engineering Implementations for the Semi-automatic System Calibration*

We have introduced the calibration method of the combined imaging system from a mathematical point of view in the previous sections. This section covers its engineering implementation.

4.5.1 *System Calibration Setup*

The system calibration setup is shown in Figure 4.10. Before we start the calibration, the calibration model needs to be placed on the conveyor with its surface parallel to the laser/X-ray beam plane. An encoder is mounted on the conveyor rotor shaft and connected to the imaging system. The encoder records the movement of the conveyor belt and sends out continuous electronic pulses to synchronize image acquisitions of the CCD cameras and the X-ray line-scan camera. Each pulse will trigger both the X-ray line-scan camera and the CCD cameras to acquire one image.

The image acquisition frequency of each camera is proportional to the speed of the conveyor.

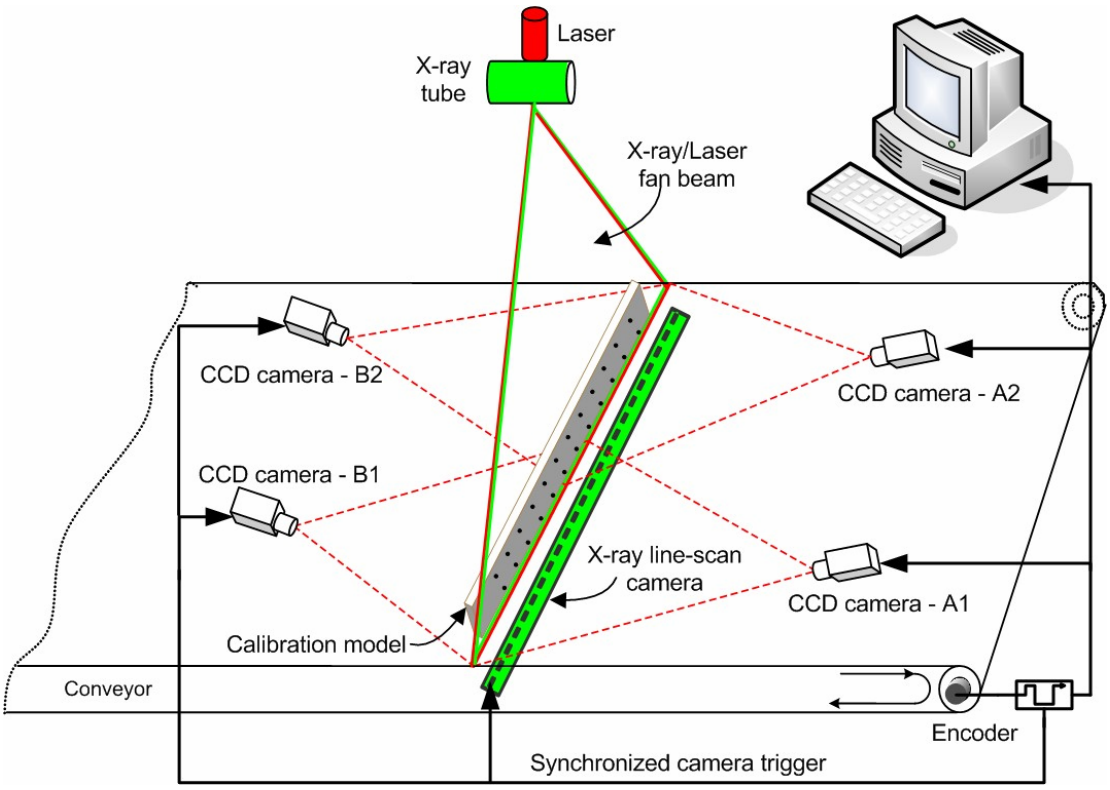


Figure 4.10 System calibration setup. The calibration model is placed on the conveyor with its surface parallel to the laser/X-ray beam plane.

4.5.2 *Calibration Image Selection*

The first step of the system calibration is to acquire the calibration images. As shown in Figure 4.10, the conveyor is setup to move from left to right. When the calibration model moves slowly with the conveyer belt, its front surface will be coplanar with the laser plane, and eventually so will its back surface. During this procedure, the X-ray line-scan camera captures an X-ray image of the calibration model, which can be used to calibrate the X-ray camera itself. At the same time, each CCD camera acquires a sequence of images, and the one captured when the model

surface is coplanar with the laser beam plane will be used to calibrate that camera (Figure 4.11). For example, only the image captured by the camera A1 (A2) when the calibration model front plane is coplanar with the laser beam plane can be used to calibrate camera A1 (A2). Similarly, only the image captured by the camera B1 (B2) when the calibration model back plane is coplanar with the laser plane can be used to calibrate camera B1 (B2). But how can we know when the calibration model surface will be coplanar with the laser light plane? This is achieved by referring to the X-ray image, since the image acquisition of the X-ray line-scan camera and the CCD cameras are synchronized. As shown in Figure 4.11, the locations of the front and back planes of the calibration model are first detected by using image processing algorithm and this location information is further used to select correct CCD camera calibration images. With this method, we can acquire the calibration images automatically by making the calibration model run through the laser beam only once, with minimum human interaction.

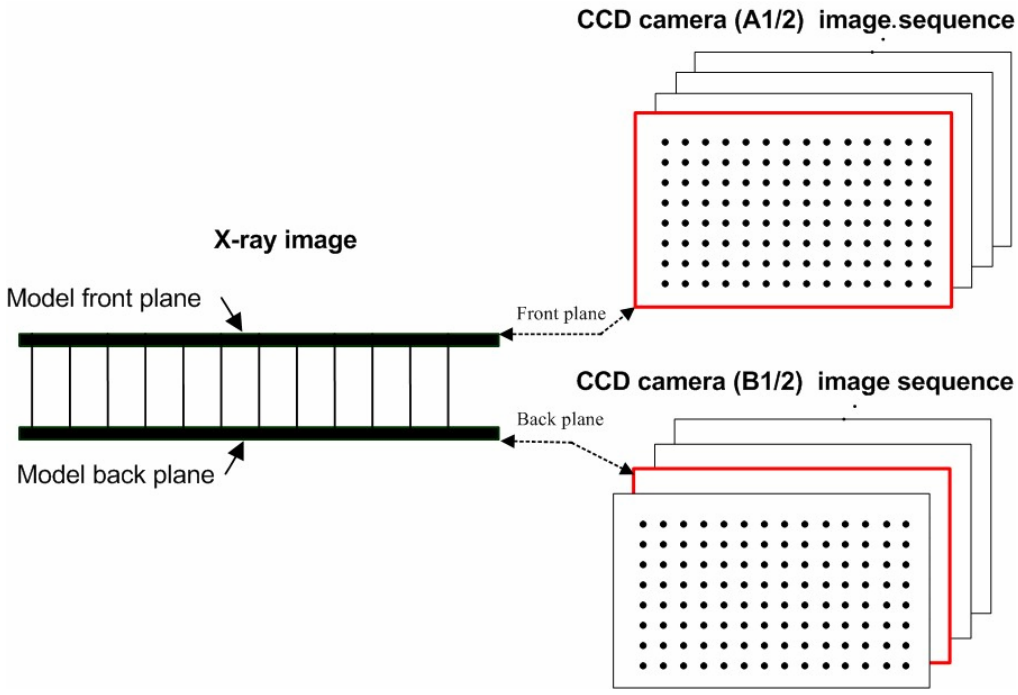


Figure 4.11 Automatic calibration image selection. The X-ray line-scan camera acquired X-ray image is used to calibrate itself. The CCD camera calibration image is selected by analyzing the X-ray image, since the image acquisitions for all the cameras are synchronized.

4.5.3 *Calibration Procedure Summary*

The complete calibration procedure is shown in Figure 4.12. In step 1, the CCD camera calibration image is captured when the calibration model plane is coplanar with the laser beam. In step 2, the X-ray line-scan camera calibration image is recorded when the X-ray beam intersects the steel pins. Calibration feature points are then detected and used to calibrate the CCD and X-ray line-scan camera respectively. In step 3, the X-ray fan beam effect is compensated with the calibration results of step 1 and 2. Lookup tables are then generated to achieve direct mapping from CCD pixels to X-ray detector sensors and to improve the performance of the whole system.

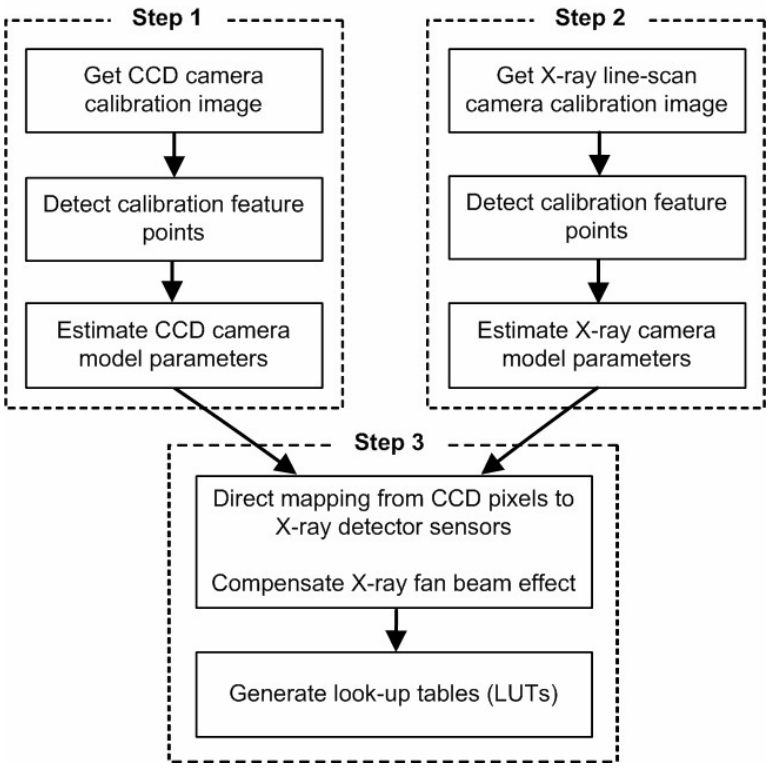


Figure 4.12 Flow chart of the calibration procedure. Step 1 is CCD camera calibration, step 2 is X-ray line-scan camera calibration, and step 3 is the direct mapping from CCD pixels to X-ray detector sensors.

Figure 4.13 shows the aforementioned two look-up tables (LUTs), one for image position mapping and the other for X-ray path-length (fan beam effect) compensation. For any pixel (i, j) in the camera CCD, its associated image position in the X-ray line-scan camera will be stored at (i, j) in the LUT_POSITION, and the distance of the X-ray path through the object is available at (i, j) in the LUT_PATH. The speed gain using the lookup table is significant, since retrieving the value from memory is much faster than executing an expensive computation.

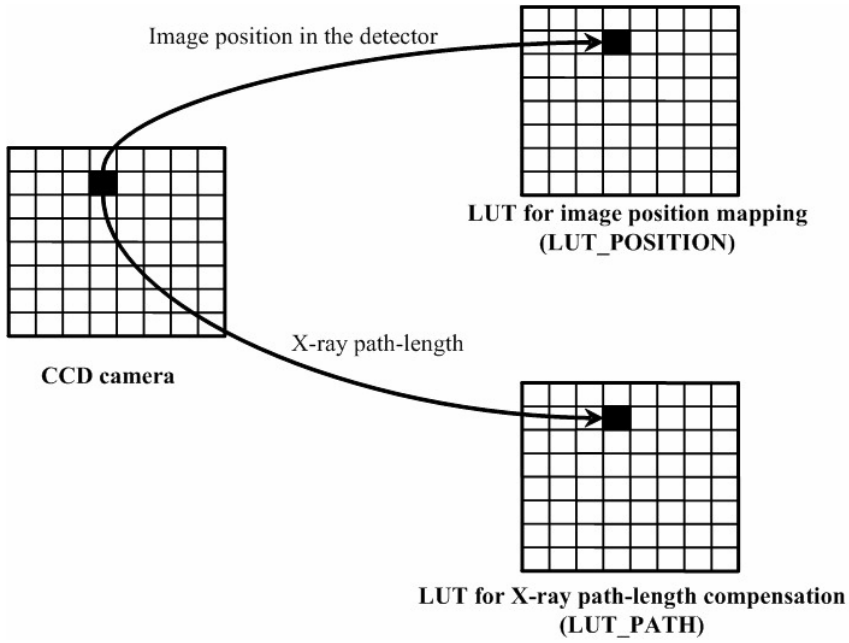


Figure 4.13 Look-up tables generated after the CCD and X-ray line-scan camera calibration. For any pixel (i, j) in the camera CCD, its associated image position in the X-ray line-scan camera will be stored at (i, j) in the LUT_POSITION, and the distance of the X-ray path trough the object is available at (i, j) in the LUT_PATH.

To further illustrate the function of the two LUTs, the procedures of how to use them in real-time application are listed bellow:

(1) Take a laser image with a CCD camera when an object passes through the laser beam.

(2) Detect the laser stripe in the image from (1).

(3) For each pixel on a laser line from (2), use the two lookup tables to get its corresponding position on the X-ray detector plane, and its associated X-ray path-length. Thus, a laser line on the CCD image plane will be mapped to a line on the detector plane.

(4) Repeat steps (1-3) to generate a laser range image. Lines received from

each loop are collected sequentially to generate a laser 3D image with X-ray fan beam effect compensated. The laser range image can be further combined with the X-ray image to compensate the thickness variation of the scanned object.

4.6 *Results and Discussion*

To evaluate the performance of the calibration procedure, experiments were conducted to determine system precision.

4.6.1 *CCD Camera Calibration Results*

Coplanar feature points [see Figure 4.7(a)] are used to calibrate the CCD camera. The centroid method is used to get sub-pixel accuracy in feature point extraction (Valkenburg, et al., 1994). From the camera calibration, the coordinates of the principal point is known; then the distance from each feature point to the principal point (the radius) can be easily calculated. Using equation (4.8), the feature points can be projected from the world coordinate system back to image buffer coordinates, yielding what are called *estimated feature points* here. The coordinate difference between the extracted and estimated feature points as a function of radius is plotted in Figure 4.14. It is observed that the error is almost uniformly distributed across different radii, which means the lens distortion is well calibrated.

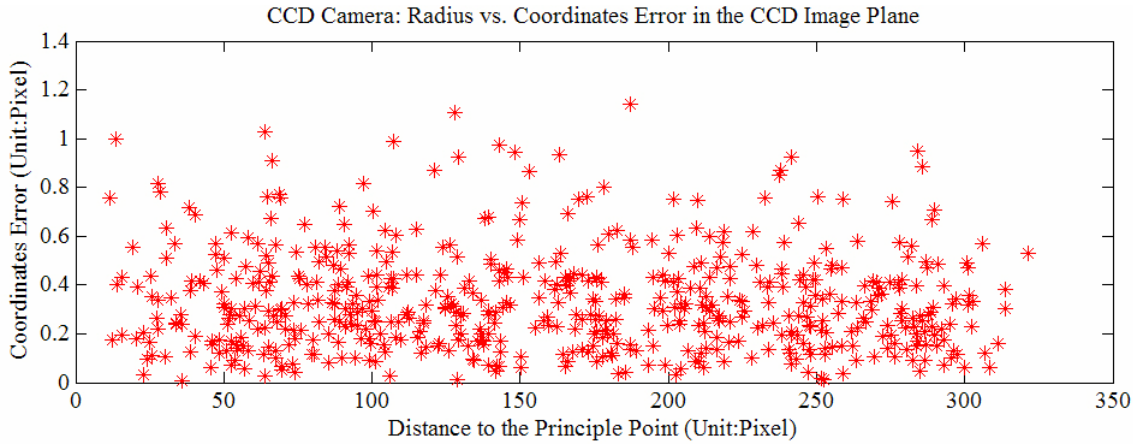


Figure 4.14 Error analysis of CCD camera calibration. The horizontal axis represents the distance to the camera principal point, and the vertical axis represents the distance between extracted and estimated feature points in the CCD camera image plane.

Table 4.1 shows the detailed error statistics. The root mean square (RMS) of the error is approximately 0.39 pixels. Since the CCD camera filed of view is about 160 mm × 120 mm and the active pixels of the CCD camera is 640 × 480, the spatial resolution of our CCD camera is 0.25 mm/pixel. Therefore, RMS error is only 0.10 mm, and its standard deviation (STD) is 0.05 mm – showing that high accuracy has been achieved after camera calibration. Even the maximum (MAX) error of 0.29 mm is small compared to the size of the foreign materials within the food.

Table 4.1 Error statistics for CCD camera calibration.

Unit	MEAN	RMS	STD	MAX
Pixel	0.34	0.39	0.21	1.14
mm	0.09	0.10	0.05	0.29

Figure 4.15 shows one image of the calibration model and its calibrated image based on the resultant calibration parameters. Notice that all the calibration points in

Figure 4.15(b) are aligned horizontally and vertically after calibration, which constitutes visual verification of the CCD camera calibration results.

(a) (b)

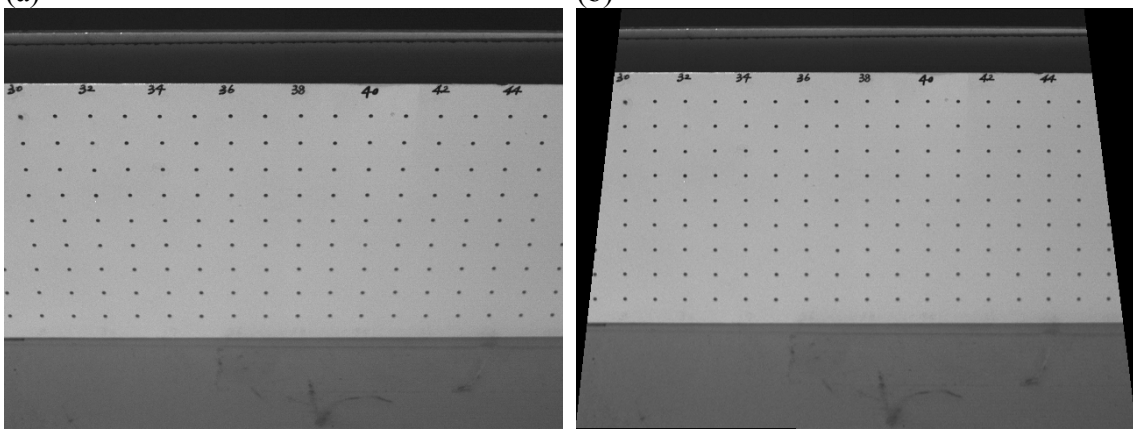


Figure 4.15 The original versus the calibrated calibration model image: (a) calibration model image obtained by CCD camera; (b) calibrated image of (a). Notice the alignment of calibration points within the two images.

4.6.2 *X-ray Line-Scan Camera Calibration Results*

The X-ray camera calibration analysis is implemented on the image plane of the X-ray camera. Figure 4.16 shows the extracted and estimated feature point coordinates in the detector plane. All the asterisks are almost on the diagonal line, which means small discrepancy between the extracted and estimated feature points. This can be further verified with Figure 4.17, which shows that the feature coordinate errors are almost uniformly distributed across the full range of the X-ray line-scan camera, mostly within one pixel.

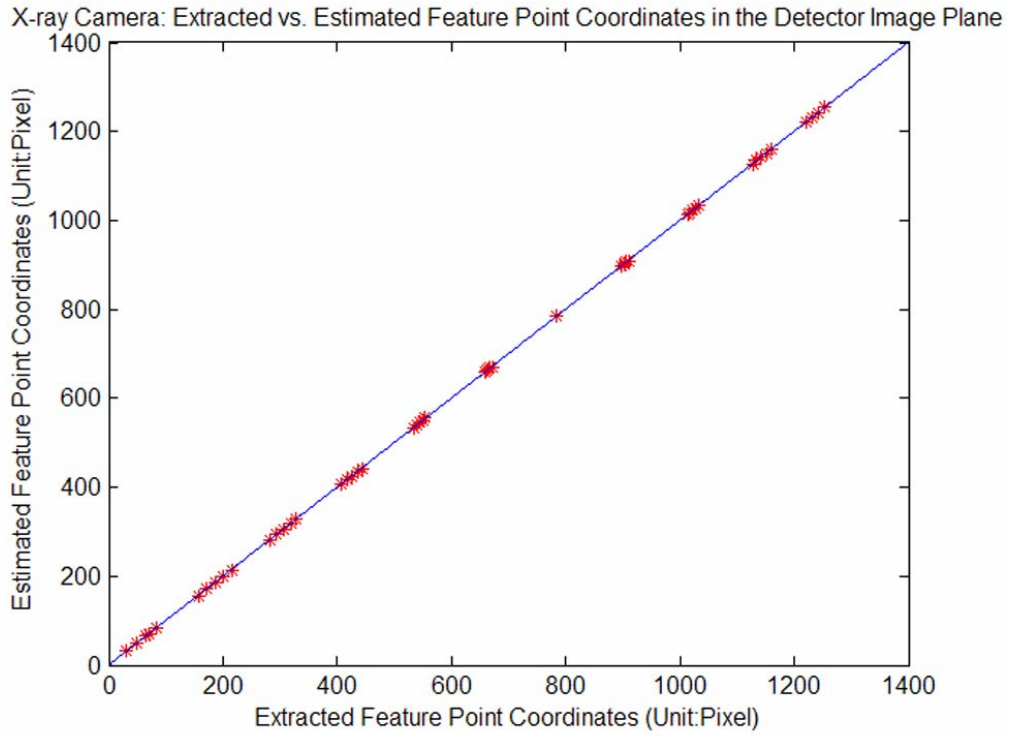


Figure 4.16 Extracted versus estimated feature point coordinates in the X-ray camera calibration.

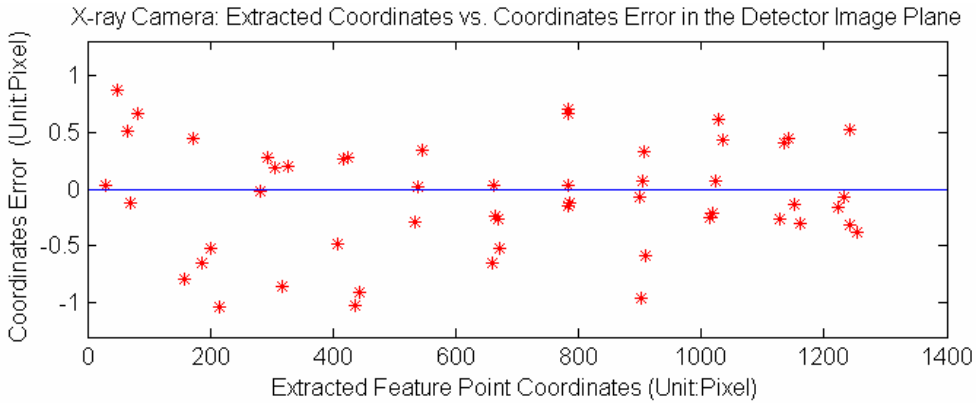


Figure 4.17 Error analysis of X-ray camera calibration. The horizontal axis represents the extracted feature point coordinates, while the vertical axis represents the distance between extracted and estimated feature points in the X-ray camera image plane.

Table 4.2 shows the error statistics for the X-ray calibration results. The RMS error is 0.48 pixels. The spatial resolution of the X-ray detector sensor is 0.4 mm/pixel, yielding an RMS for the X-ray camera of 0.19 mm, which is acceptable for food safety inspection.

Table 4.2 Error statistics for the X-ray line-scan camera calibration.

Unit	MEAN	RMS	STD	MAX
Pixel	0.39	0.48	0.48	1.03
mm	0.16	0.19	0.19	0.41

4.6.3 *Combined CCD Camera and X-ray Line-Scan Camera Calibration Results*

For the combined CCD camera and X-ray camera calibration, each pixel of the CCD camera is mapped directly to an X-ray line-scan sensor based on the calibration results. By using equations (4.8a), (4.8b) and (4.22), for each feature point in the CCD image plane, we can estimate its corresponding position in the X-ray camera image plane, which is called the estimated feature point coordinates in Figure 4.18. The coordinates of the estimate feature points are then compared with those of the extracted feature points from the X-ray image. If calibration procedure is well implemented, the difference between each coordinate pair should be very small. This is obviously shown in Figure 4.18 and 4.19.

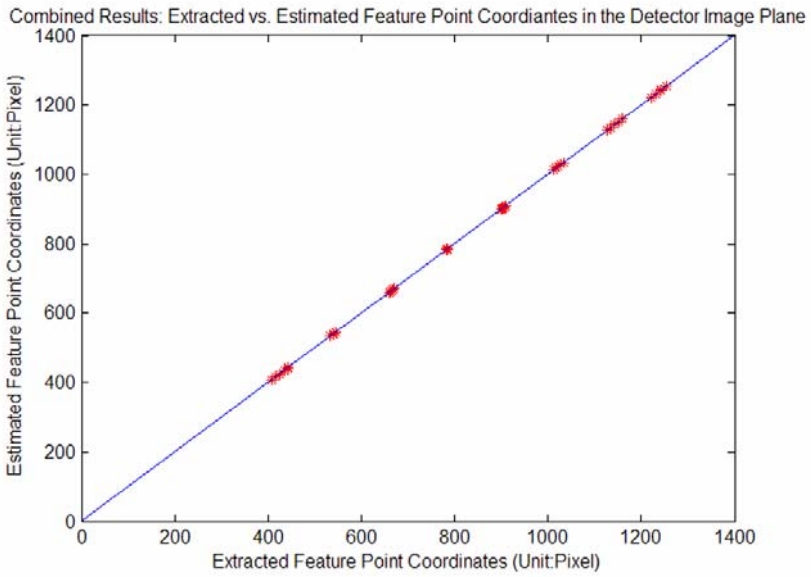


Figure 4.18 Extracted versus estimated feature point coordinates in the detector image plane.

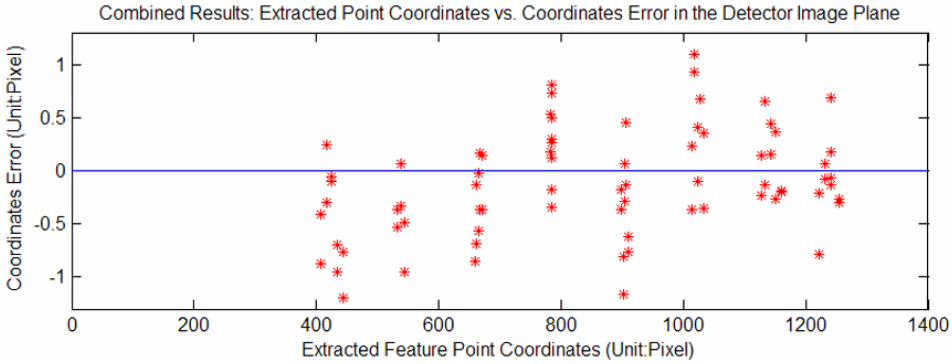


Figure 4.19 Error analysis of combined CCD camera and X-ray camera calibration. The horizontal axis represents the extracted feature point coordinates, while the vertical axis represents the distance between extracted and estimated feature points in the X-ray camera image plane.

Table 4.3 shows the error statistics for the combined calibration results. The final calibration is very precise. It has RMS error as low as 0.2 mm and a standard deviation of 0.2 mm. Though the RMS error for this combined calibration is a little

higher than that of the X-ray camera calibration result, it still gives a very good calibration performance. Together with the 0.16-mm MEAN, 0.2-mm STD, and 0.48-mm MAX, this calibration approach makes the whole combined X-ray and laser range imaging system accurate and robust.

Table 4.3 Error statistics for combined CCD camera and X-ray line-scan camera calibration.

Unit	MEAN	RMS	STD	MAX
Pixel	0.41	0.51	0.50	1.19
mm	0.16	0.20	0.20	0.48

To better illustrate the calibration result and its application, Figure 20 shows images of a chicken (*Gallus Gallus*) sample. Figure 4.20(a) is the photo of a chicken breast. Figure 4.20(b) is the corresponding 3D laser image from the result of section 4.5.3 steps 1-4. Figure 4.20(c) is the X-ray image. Notice that the 3D laser image and X-ray image are already perfectly aligned after the system calibration. This will make the further image combination and processing much easier, since no image registration is needed. These two images from different modalities will be further processed to find possible contaminations within the chicken meat (Jing, 2003; Xin, 2003).

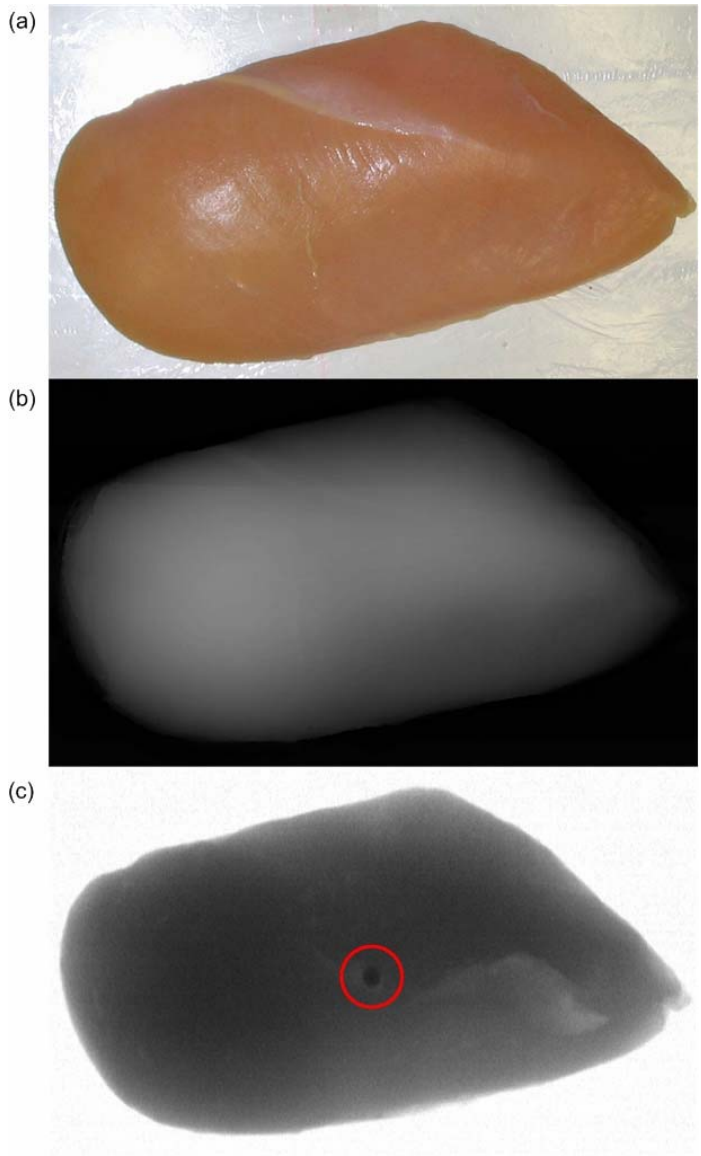


Figure 4.20 Photo of a chicken breast sample; (b) 3D laser image of the sample; (c) X-ray image of the sample chicken breast with the foreign material highlighted.

4.7 *Conclusions*

We have demonstrated a three-step method to calibrate a combined X-ray and laser range imaging system based on pinhole camera models. The CCD camera was calibrated using a coplanar set of points, giving a RMS error of 0.09 mm, a standard deviation of 0.05 mm, and a maximum error of 0.29 mm. The X-ray line-scan camera calibration was performed using a steel pin pattern (RMS error 0.19mm, standard deviation 0.19 mm, maximum error 0.41 mm). The distortion caused by the X-ray fan beam effect was compensated after X-ray camera calibration. Direct mapping from the CCD camera to the X-ray line-scan camera was achieved. Two look-up tables, one for mapping image positions from CCD to X-ray camera and another for finding X-ray path lengths, are created to replace the run-time computation. Experiments show that the combined laser 3-D and X-ray imaging system can be well calibrated by our three-step calibration method (RMS error 0.20 mm, standard deviation 0.20 mm, and maximum error 0.48 mm).

Chapter 5: Texture Analysis based Backlight Imaging System for Walnut Meat and Shell Separation

In the previous chapter, we introduced the combined X-ray and laser range imaging system for physical contamination detection, mainly focusing on the method of calibrating the whole system. In this chapter, we will introduce another machine vision food safety inspection system – a backlight imaging system for automatic walnut meat and shell separation.

5.1 *Introduction*

The black walnut, *Juglans nigra*, grows throughout the central and eastern U.S. Unisexual flowers emerge on black walnut from mid-April to mid-June, appearing with the leaves on a separate inflorescence of the same tree. A globular fruit is produced which contains a corrugated nut in its yellowish-green husk. The nut is usually 40 to 60 mm in diameter and spherical shaped, containing an oil-rich, sweet, and edible seed. The seed (walnut meat) has a rich and distinctive flavor and is often used as an additive in value-added foods, such as ice cream, bakery items, and gift bags. The weight of a single nut varies from 15 to 24 gram and the meat weight is about 30%. There are over 15.4 million acres of black walnut trees with approximately 700 million black walnut trees in the United States (Jones, et al., 1998). An acre of mature black walnut trees (from 20 to 70 years old) produces about 1000 to 1700 pounds of raw nuts annually (Hatcher, et al., 1998). However, of the 4 billion pounds of walnuts produced annually, only about 20 millions pounds of the raw nuts are commercially processed (Hammons, 1998). With nut processors often paying only $0.12 per pound of hulled nuts (Hammons, 2007), growers were not motivated

to harvest the nuts, and the U.S. has insufficient nut processing capacity available. However, there is a great potential economic benefit if an efficient walnut processing method is available. Compared with the low price of the hulled nuts ($0.12/lb), the current market price of the black walnut meat is $12 and up per pound, depending on the quality and physical size of the meat .

In current walnut processing plants, black walnuts are cracked and most of the shells are removed by air lathe. The remaining shell fragments are then removed by visual inspection to meet the required marketable quality. This manual inspection is very time consuming and labor intensive because shell and meat fragments are similar in size and color (Krishnan and Berlage, 1984). Therefore, an accurate automated inspection method is desirable to reduce labor and time, while ensuring product quality.

In the past decades, machine vision and non-vision based technologies have been applied to separate the walnut meat and shell. Kishnan and Berlage (1984) investigated the feasibility of using iron and magnetic fluid to remove shells from walnut meat. In their work, the walnuts were coated with either iron powder or magnetic fluid for comparison purposes. The coated nuts were cracked in a commercial nut cracker and then conditioned over a permanent magnetic drum separator. Although both of their methods are successful in removing shell fragments from nut meat, they could not be adopted by the industry immediately because both the iron powder and magnetic fluid are considered food additives and therefore require the approval of the Food and Drug Administration.

A notable study was presented by Pearson and Young (2002), who tried to automatically sort almonds with embedded shells using laser transmittance imaging. Their research work was based on the fact that laser light could penetrate the almond kernel, but could not go through the almond shell. A shell fragment blocks nearly all the light, forming a dark spot in the image which can be captured by a line-scan camera. The imaging setup is effective, while the laser transmittance image of almond kernels does not contain many unique texture patterns which makes the sorting relatively difficult.

In this study, we propose to use a backlight imaging method to differentiate black walnut shells and meat based on the observation that their backlit images showed quite distinctive texture patterns. The objectives include: (1) to build a backlight imaging system for image acquisition of walnut meat and shell; (2) to explore texture analysis and pattern classification methods based on the acquired backlit walnut images; and (3) to evaluate the performance of the backlight imaging system. So, this chapter is structured as follows. Section 5.2 introduces the imaging system setup. Section 5.3 -5.4 describes the texture analysis and pattern classification methods. Finally, Section 5.5 demonstrates the performance of the system.

5.2 Materials and Backlight Imaging System Setup

Intact current-year and one-year-old harvested black walnuts were provided by the USDA Agricultural Marketing Service. The walnuts were cracked and the walnut meat and shells were separated manually. Photos of various sample types are shown in Figure 5.1(a). These include light meat (light colored walnut meat), dark meat (dark colored walnut meat), and shells with either inner or outer surfaces facing

the camera. Due to the shape, color and size variations of walnut shells and meat, it is difficult to separate them based on the traditional color imaging method. For example, a dark meat surface may look quite similar to a shell's outer surface and some shells' inner surfaces have similar patterns to meat. However, the difference between meat and shell surfaces is quite obvious with the backlight imaging system [shown in Figure 5.1(b)]. Texture analysis methods thus allow for easy separation of shell and meat with high classification accuracy.

(a) (b)

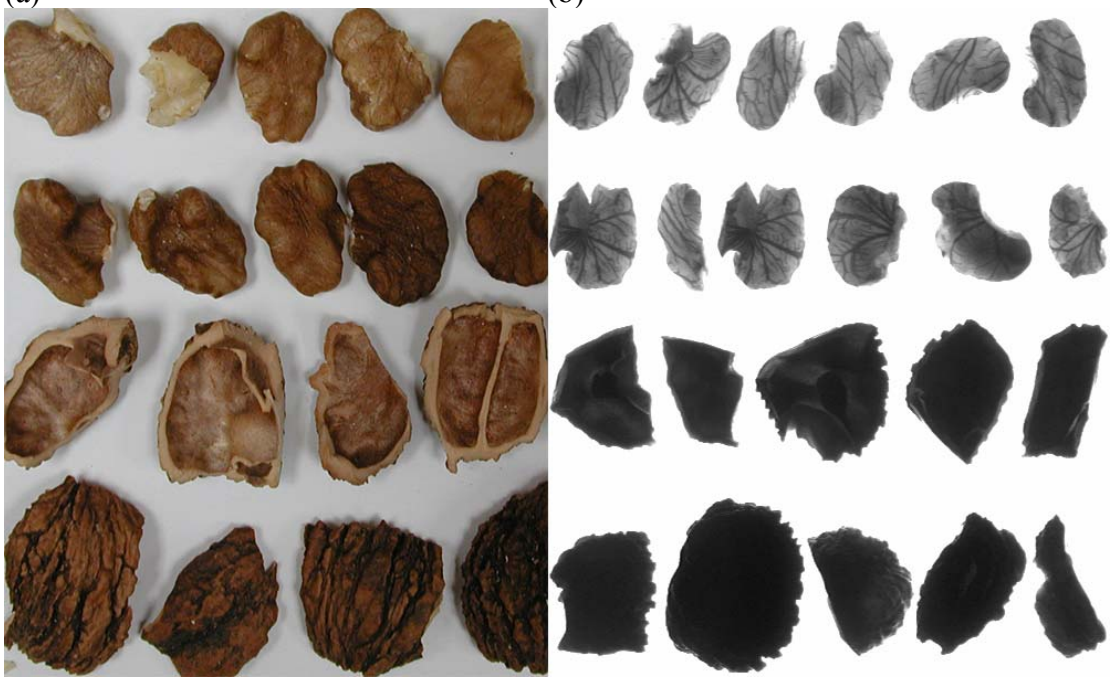


Figure 5.1 (a) Walnut samples showing (from top to bottom row): light meat, dark meat, inner shell surface and outer shell surface; (b) Typical backlit images of walnut samples showing (from top to bottom row): light meat, dark meat, inner shell surface and outer shell surface.

The experimental backlight imaging system is shown in Figure 5.2(a) and its schematic diagram is shown in Figure 5.2(b). Walnut samples were placed on transparent glass with a diffusion film underneath. The film diffused the light making the background of the backlit image bright. Below the film was a box that held a

fiber-optic annular ring laminator to illuminate the sample. A monochrome CCD camera was positioned above the transparent glass and pointed down toward the sample. The camera was connected to an imaging board with its field of view (FOV) adjusted to cover the entire glass window. A PC with imaging board installed inside controlled image acquisition and handled image analysis.

As shown in Figure 5.2(b), among the light rays emitted from the annular ring laminator, those rays that traveled directly to the camera were blocked by the light shielding plate. Only the rays between lines O-A and O-B in the figure penetrated the glass stage and escaped out of the lighting box. When a nut sample was placed on the stage, the light rays struck it and scattered. Part of the scattered light was then captured by the camera. By this means, the CCD camera did not capture direct light from the laminator, but captured only the scattered light from both the sample and the diffusion film. This backlight imaging system is the initial proof-of-principle setup for data collection. The design of the real-time similar principle imaging system will be proposed in section 5.5.4.

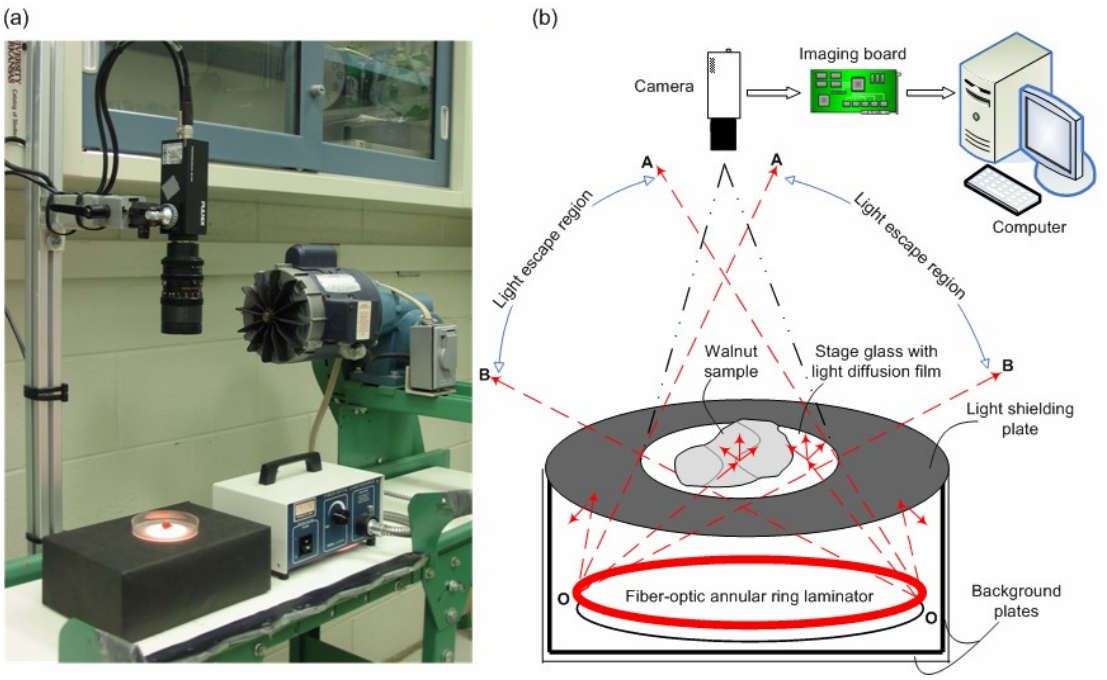


Figure 5.2 (a) Experimental backlight imaging system; (b) Schematic plot of the backlight imaging system.

The acquired backlit images of walnut fragments are shown in Figure 5.1(b). Due to the difference in light transmission, the backlit images showed distinct texture patterns. Walnut meat images were bright with striped texture, while shell images were much darker with less texture. In some cases, when a shell's inner or broken surface faced the camera, the light color of the surface made it prone to the reflected light, which gave the image some intensity variation and texture patterns [see first image in the third row of Figure 5.1(b)]. Fortunately, these texture patterns were different from those of the walnut meat. Therefore, they could still be correctly separated by using proper texture operators and classification algorithms.

The walnut classification procedure is shown in Figure 5.3. Walnut samples were randomly selected and the backlit image of each sample was captured by the imaging system (Figure 5.2). The foreground area of each image was segmented from

the background and different texture features were extracted (section 5.3). These texture features were then concatenated into one vector for each sample. All the features are further normalized (data pre-processing) and stratified random sampling (Witten and Frank, 2005) was performed to split the samples into a training set and a testing set (section 5.5). The support vector machine based recursive feature elimination (SVM-RFE) method was then used to select good feature subsets and obtain a trained SVM classifier. The performance of the classifier was further estimated using the testing data set.

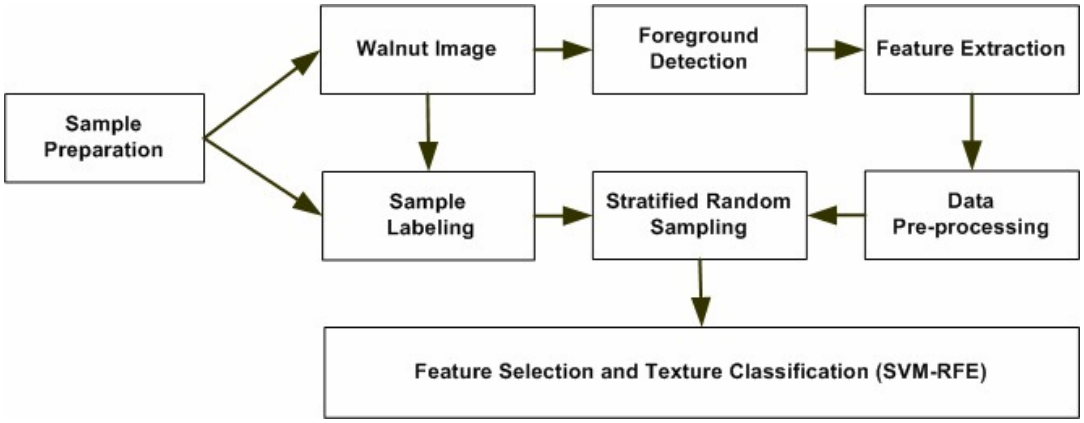


Figure 5.3 Flowchart of the walnut classification.

5.3 *Texture Analysis of Backlit Walnut Images*

Texture plays an important role in the composition of natural images. Texture analysis and classification are essential in many areas, such as applications in biomedical image analysis, remote sensing, industrial surface inspection, etc. (Kumar and Pang, 2002; Zhong and Sclaroff, 2003; Sun, et al., 2005; Kokare, et al., 2005; Khademi and Krishnan, 2007). Many existing texture analysis methods make the assumption, explicit or implicit, that texture images are acquired from the same viewpoint (i.e. the same spatial scale and orientation). However, in many real-world

applications, it is very difficult to ensure that the captured images have the same orientation and scaling. Therefore, texture analysis methods should be invariant to viewpoint changes. In our research, all of the images were captured under the same experimental setup to assure the same scale. However, the orientation of walnut samples was not controlled, which requires our texture descriptors to be rotationally invariant. In this study, we evaluated five texture analysis methods: local binary pattern operator, circular Gabor filters, Wavelet transform, circular symmetrical gray-scale co-occurrence matrix and image histogram based statistics. They were all modified, if necessary, to achieve a rotationally invariant property on the foreground area of the images.

5.3.1 *Local Binary Pattern Operator* (*LBP*)

The LBP operator is an efficient gray scale and rotation invariant texture descriptor (Ojala, et al., 1996, 2000, 2002). In a 3x3 neighborhood (Figure 5.4a), the eight neighbors of a pixel were thresholded at the value of the center pixel (Figure 5.4b). The eight binary numbers associated with the eight neighbors were then read sequentially in the counter clockwise direction to form an 8-bit number (Figure 5.4c). Although the resultant 8-bit number was invariant against any monotonic gray scale transformation by definition, it could vary due to image rotation. To achieve rotation invariance, an arbitrary number of circular bit-wise shifts were made until the resultant 8-bit number achieved its minimum (Figure 5.4c). There could be 36 different values, corresponding to 36 unique rotation invariant local binary patterns. However, the occurrence frequencies of the 36 different patterns varied greatly. Some of them were encountered rarely, making them statistically unstable. To quantify the

varying performance of individual patterns, the number of spatial transitions (bitwise $0 \to 1$ or $1 \to 0$ changes) in the patterns was counted and defined as the uniformity measure U (Ojala, et al., 2002). For example, patterns "00000000" and "11111111" had a U value of 0, while the pattern "00001111" in Figure 5.4c had a U value of 2 as there were exactly two transitions ($0 \to 1$ or $1 \to 0$) in the pattern. Patterns that had U values less than or equal to 2 were designated uniform and each was assigned a unique LBP index. All other patterns were labeled non-uniform and collapsed into one value, 9 (Figure 5.4d), which made the number of total different patterns 10. By compressing the non-uniform patterns, the total number of LBP values decreased from 36 to 10. This mapping is based on the observation that non-uniform patterns are not dominative in deterministic textures (Ojala, et al., 2002).

For each image, a histogram of LBP values was calculated from the foreground pixels and further normalized. The value of each bin was collected to form the feature vector that described the local spatial patterns of the image:

$$FeaVec_{LBP} = \{hist_1, hist_2, \cdots, hist_{10}\}, \tag{5.1}$$

where $hist_i$ ($i = 1,2,\cdots,10$) is the ith bin of the normalized histogram.

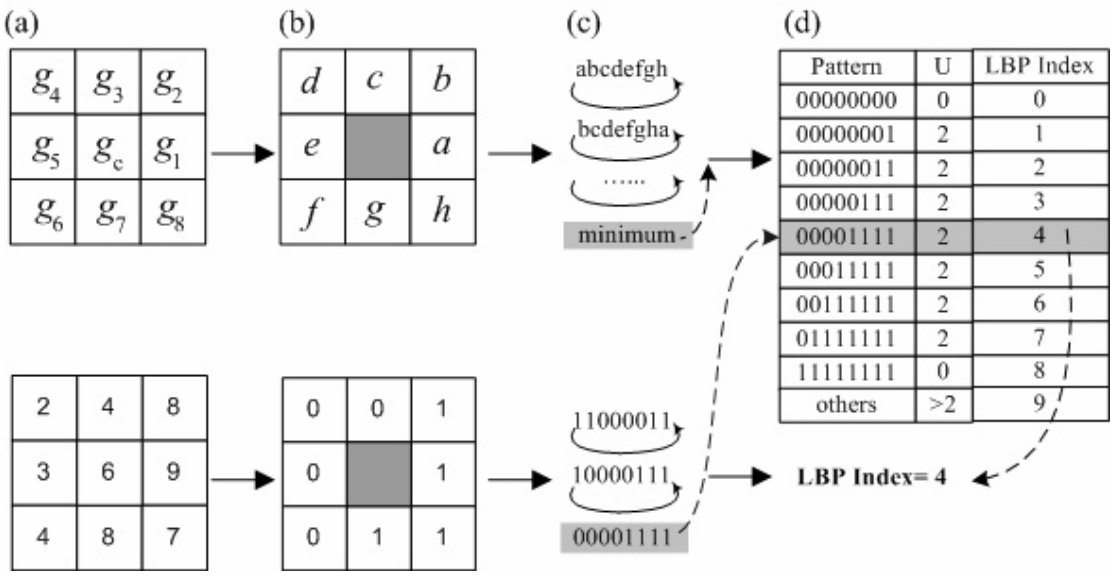


Figure 5.4 Computation of LBP. (a) A 3x3 neighborhood of a pixel; (b) Neighborhood is thresholded at the value of the center pixel; (c) A 8-bit binary number is formed by reading sequentially the eight neighbors and further being circularly shifted until reaching its minimum; (d) The 8-bit binary patterns are further compressed to 10 LBP values based on the uniformity measure.

The rotation invariant LBP operator provides robust information about local spatial patterns, but by definition it discards detailed contrast (Ojala, et al., 2000). This immediately suggested that the contrast of local image textures should be incorporated as well. Under stable lighting conditions, the contrast could be measured with rotation invariant local variance:

$$VAR = \frac{1}{8}\sum_{i=1}^{8}(g_i - \mu) \quad \text{, where } \mu = \frac{1}{8}\sum_{i=1}^{8} g_i \tag{5.2}$$

For each image, any foreground pixel (except pixels on the boundary) has a corresponding VAR output. All the output values were pooled together to obtain a histogram. The histogram was further normalized and mapped to a normalized cumulative histogram (NCH). m centiles [m =11 in our research (Jin, el al., 2008)]

were then calculated from the NCH by finding the cut values that divide NCH into desired $m+1$ parts. These centiles were then used to characterize local texture contrast of the images:

$$FeaVec_{LBP\_VAR} = \{centile_1, centile_2, \cdots, centile_m\} \tag{5.3}$$

5.3.2 *Circularly Symmetric Gabor Filters*

Multi-channel Gabor filters are recognized to be a very useful tool in texture analysis (Daugman, 1988; Bovik, et al., 1990; Kumar and Pang, 2002; Kamarainen, et al., 2006). The Gabor representation is optimal in minimizing the joint two-dimensional uncertainty in space and frequency (Daugman, 1988). It is mathematically expressed as

$$G(x, y) = g(x', y')\exp(-2\pi j(Ux + Vy)), \tag{5.4}$$

where (U,V) defines the position of the filter in the Fourier domain with a center frequency of $F = \sqrt{U^2 + V^2}$ and an orientation angle $\theta = \arctan(V/U)$. The term $g(x', y')$ represents a Gaussian function rotated at an angle ϕ, where

$$(x', y') = (x\cos\phi + y\sin\phi, -x\sin\phi + y\cos\phi) \tag{5.5}$$

are the rotated coordinates. The general form of the Gaussian function is

$$g(x, y) = \left(\frac{1}{2\pi\lambda\sigma^2}\right)\cdot\exp\left[-\left(\frac{(x/\lambda)^2 + y^2}{2\sigma^2}\right)\right], \tag{5.6}$$

where λ defines the aspect ratio and σ the scale factor.

The above traditional Gabor function in a 2-D image domain can be considered as orientation and scale tunable edge, and line detectors as shown in Figure 5.5(a). It is very effective at orientation dependent texture analysis. However, in our case, texture orientation was less important. This is why the circular symmetric Gabor filters (Porter and Canagarajah, 1997) are more appropriate:

$$h(x,y) = g(x,y)\exp(-2\pi Fj(x^2+y^2)) \tag{5.7}$$

where F is the center frequency and $g(x,y)$ is same as equation (5.6) with $\lambda = 1$.

The Gabor filters are not orthogonal to each other which means filtered images contain redundant information. The common strategy of filter design is to ensure that the half-peak magnitude support of the filter responses in the frequency spectrum touch each other, which requires F and σ to satisfy the following equation:

$$F\sigma = \frac{\sqrt{2\ln 2}}{2\pi}\left(\frac{2^B+1}{2^B-1}\right) \quad , \tag{5.8}$$

where B is the bandwidth of the Gabor filter (Jain and Farrokhnia, 1991). Since several experiments have shown that the frequency bandwidth of simple cells in the visual cortex is about one octave, we chose B=1, and four center frequency values, F, are used: $\sqrt{2}$ {1/4, 1/8, 1/16, 1/32}. Figure 5.5(b) shows the magnitude of these filters in the Fourier domain.

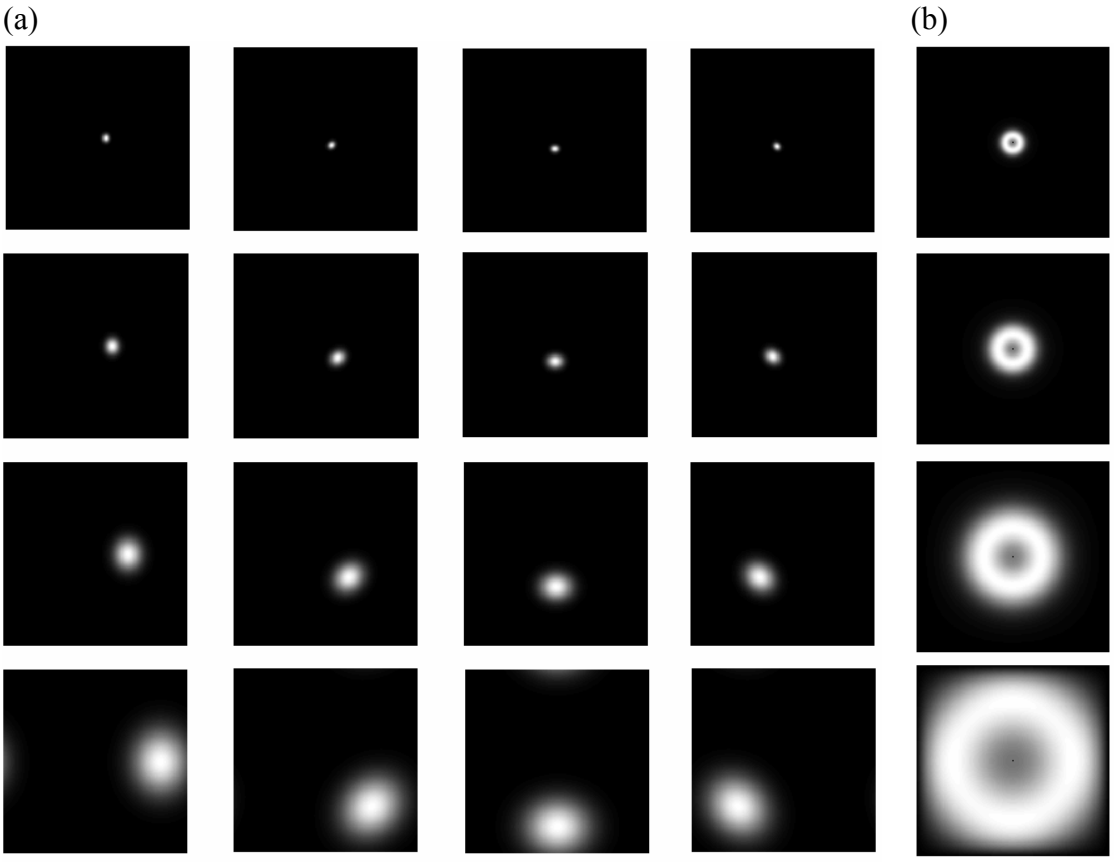


Figure 5.5 Magnitude of traditional vs. circularly symmetric Gabor filters in the Fourier domain. (a) Magnitude of traditional Gabor filters in the Fourier domain with F=$\sqrt{2}$ { 1/4, 1/8, 1/16, 1/32 } , B=1 and $\lambda = 1$, and four orientations; (b) Magnitude of circularly symmetric Gabor filters in the Fourier domain with F=$\sqrt{2}$ { 1/4, 1/8, 1/16, 1/32 } , B=1 and $\lambda = 1$.

Given an image $I(x, y)$, its circular Gabor filtered image can be defined as the convolution:

$$G(x, y) = \sum_{x_1} \sum_{y_1} I(x_1, y_1) h^*(x - x_1, y - y_1) \tag{5.9}$$

where $*$ indicates the complex conjugate. For the four center frequency values F, four filtered image $G_i(x, y)$ i=1, 2, 3 and 4 were gained. The mean μ_i and the standard deviation σ_i of the magnitude of each filtered image were used to characterize the different texture patterns (Manjunath and Ma, 1996):

$$\mu_i = \sum_x \sum_y |G_i(x,y)| \text{ and } \sigma_i = \sqrt{\sum_x \sum_y \left(|G_i(x,y)| - \mu_i\right)^2} \quad i=1,2,3,4 \qquad (5.10)$$

A feature vector based on the circularly symmetric Gabor filters was then defined as

$$FeaVec_{GABOR} = \{\mu_1, \sigma_1, \mu_2, \sigma_2, \mu_3, \sigma_3, \mu_4, \sigma_4\} \qquad (5.11)$$

5.3.3 *Wavelet Transform*

The 2D wavelet transform performs a spatial-frequency analysis on an image by repeatedly decomposing the image at the lower frequency sub-bands (Starck, et al., 1998; Kokare, et al., 2005; Khademi and Krishnan, 2007). The rationale behind this spatial-frequency joint representation is to cut the image of interest into several parts using sets of scaleable modulated windows and then analyze the parts separately. Analyzing an image in this way gives us detailed information about the image under different scales. A full wavelet image decomposition results in an array of wavelet coefficients which is of the same shape and size as the original image. Figure 5.6 shows the result of a three-level wavelet decomposition. The channels indicated in the diagram are the outputs from the three stages of the sub-band filtering process and contains information of the original image at different spatial frequencies and orientations. The HH (LL) channel represents image information of both high (low) horizontal and vertical frequency, while the HL (LH) channel represents image information of high (low) horizontal and low (high) vertical frequency. If the image contains strong texture content at the frequency and orientation represented by a channel, the energy (defined as $l2$-norm) of the reconstructed image associated with

that channel will be large. Denote $\mathrm{Im}age_i(x,y)$ as the reconstructed image from channel i, $i \in \{HL_k,\ LH_k,\ HH_k,\ LL_3 \mid k = 1,2,3\}$, the corresponding energy is:

$$E_i = \sum_x \sum_y \mathrm{Im}age_i(x,y)^2 \tag{5.12}$$

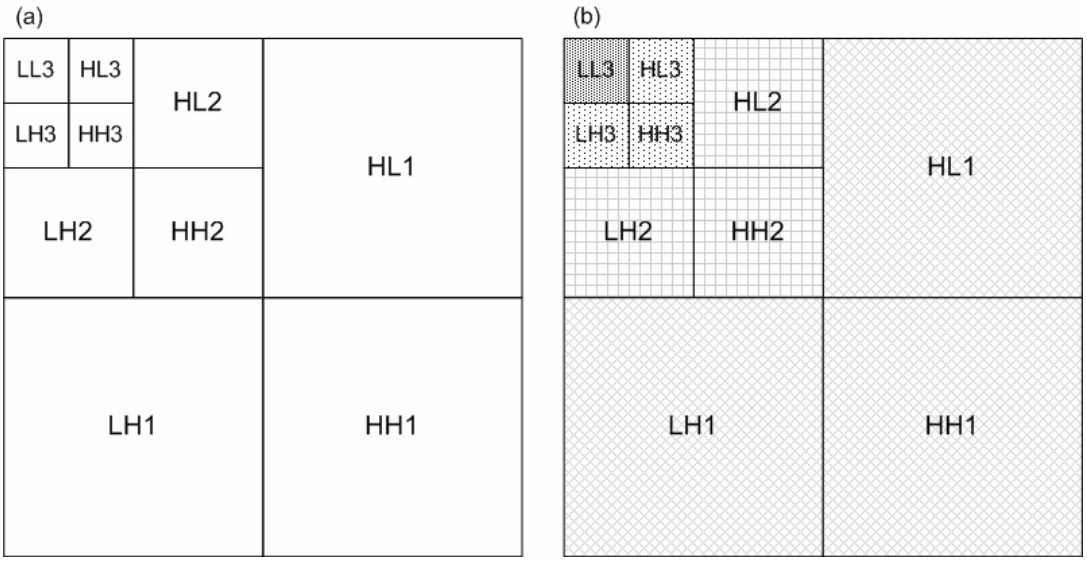


Figure 5.6 Three-level wavelet image decomposition. (a) Ten channels of three-level wavelet decomposition; (b) Four frequency bands (labeled by the shade pattern) to make the extracted texture features less sensitive to image rotation.

As shown in Figure 5.6(a), there are a total of 10 frequency/orientation channels for a three-level wavelet decomposition. The energy in these channels is not rotationally invariant. To eliminate rotation variance, HL, LH and HH channels of each decomposition level were combined (Figure 5.6b) to produce four frequency bands which are insensitive to image rotation. The normalized energy at each level was therefore used for texture classification. The wavelet energy based feature vector can be represented as

$$FeaVec_{WAVELET} = \{P_{L_3}, P_{H_1}, P_{H_2}, P_{H_3}\}, \tag{5.13}$$

where $P_{L_3} = E_{LL_3} / E_{Total}$, $P_{H_k} = (E_{LH_k} + E_{HL_k} + E_{HH_k}) / E_{Total}$ $(k = 1,\ 2,\ 3)$ with $E_{Total} = E_{LL_3} + \sum_{k=1}^{3} \left(E_{LH_k} + E_{HL_k} + E_{HH_k} \right)$.

5.3.4 *Circularly Symmetric Gray Level Co-occurrence Matrix*

The gray level co-occurrence matrix (GLCM) (Haralick, et al., 1973) is one of the most widely used texture analysis methods. It is a second-order spatial histogram that estimates the distribution of co-occurring image pixel values at a given offset. Let $\delta(r, \theta)$ denote a vector in the polar coordinate of the image, as shown in Figure 5.7. For any such vector, we can compute the joint distribution of pairs of gray levels that occur at pairs of locations separated by δ .

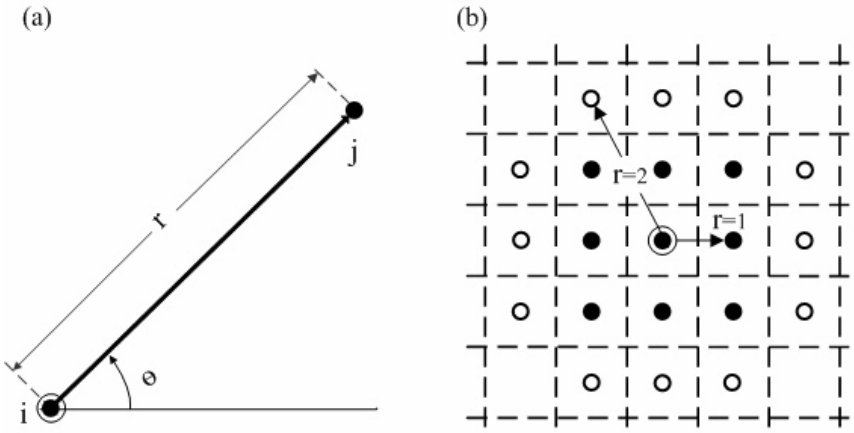


Figure 5.7 Displacement $\delta(r, \theta)$ in the polar coordinate system; (b) neighborhood pixels for r=1 and r=2.

Note that the $\delta(r, \theta)$ parameterization makes the co-occurrence matrix sensitive to image rotation. Therefore, it should be modified to achieve rotation invariance. Mathematically, a circularly symmetric gray level co-occurrence matrix (CS-GLCM) C is defined over an $n \times m$ image I , parameterized by offset r , as:

$$C_r(i,j) = \sum_{p,s=1}^{n} \sum_{q,t=1}^{m} \begin{cases} 1, & if \quad I(p,q) = i, I(s,t) = j \; and \; Dist\big((p,q),(s,t)\big) = r \\ 0, & otherwise \end{cases}, \quad (5.14)$$

where $Dist\big((p,q),(s,t)\big)$ is the distance between a point pair located at (p,q) and (s,t). The neighborhood of one pixel with d=1 and d=2 is shown in Figure 5.7(b). Basically, for each reference pixel (p,q) in image *I*, the neighboring pixels at distance *r* were checked one by one. The pixel location within the given neighborhood was ignored and only its gray level was taken into account. By doing so, the co-occurrence matrix *C* is only a function of the neighborhood distance r without being affected by image rotation. The matrices were further normalized to approximate the joint distribution probability of gray level pairs separated by *r*. In our research, to make the co-occurrence matrix less sparse and also to save the computation time, the 8-bit gray images were first rescaled into 16 gray levels. Two sets of rotationally invariant co-occurrence matrix C_1, C_2 with $r = 1, 2$ were then computed. Five statistical texture features, entropy, energy, contrast, homogeneity and correlation, were derived from each co-occurrence matrix (Petrou and Sevilla, 2006). The CS-GLCM based feature vector was the combination of all these properties which can be represented as

$$FeaVec_{CS-GLCM} = \{Ent_k, Eng_k, Con_k, Hom_k, Cor_k \mid k = 1,2\} \quad (5.15)$$

where $Ent_k, Eng_k, Con_k, Hom_k, Cor_k$ are the entropy, energy, contrast, homogeneity and correlation derived from normalized co-occurrence matrix (with $r = k$), respectively (see Appendix A for the definition).

5.3.5 *Image Histogram Based Statistics*

Apart from the above sets of features, image histogram and its associated statistical features were also used to describe the texture patterns. The purpose of an image histogram is to graphically summarize the distribution of pixel values. It is obtained by splitting the range of the gray scale into equal-sized bins and counting the number of pixels whose intensity fall into each of the bins. The count in each bin is further divided by the total number of pixels to get a normalized histogram $NH_k, (k = 1,2,...,N)$, where N is the number of bins. In this study, we employed six statistics derived from the normalized histogram, including mean, variance, entropy, energy, skewness and kurtosis, to characterize the histogram distribution (see Appendix B for the definition).

The image histogram based statistical texture feature can then be represented as

$$FeaVec_{HIS} = \{NH_k\ (k = 1,...,N), Mean_{NH}, Var_{NH}, Eng_{NH}, Ent_{NH} Ske_{NH}, Kur_{NH}\} \quad (5.16)$$

where $Mean_{NH}, Var_{NH}, Eng_{NH}, Ent_{NH} Ske_{NH}, Kur_{NH}$ are the mean, variance, energy, entropy, skewness and kurtosis of the normalized histogram respectively (Ott and Longnecker, 2001).

5.4 *Feature Selection and Texture Classification*

All the features introduced above can be combined together and fed to the learning algorithm for texture classification. However, not all these features are equally useful for the learning algorithm and a feature selection method can help in this case. Feature selection has many benefits, including facilitating data visualization

and understanding, reducing the measurement and storage requirements, decreasing training and utilization time, and defying the curse of dimensionality to improve prediction performance (Guyon and Elisseeff, 2003). As we mentioned in the chapter of literature review, three general approaches of feature selection exist: filters, wrappers and embedded methods (Guyon, et al., 2006). Filter methods (Almuallim and Dietterich, 1994; Kira and Rendell, 1992) select features on the basis of their relevance or discriminate power with regard to the targeted classes. Wrapper methods (John, et al, 1994; Kohavi and John, 1997) wrap feature selection around a specific prediction method, and the estimated accuracy of the prediction method is used to directly judge the usefulness of a feature. Embedded methods (Weston, et al., 2001, 2003; Bradley and Mangasarian, 1998; Rakotomamonjy, 2003) differ from the two former feature selection methods in the way feature selection and classifier learning interact with each other. One example of this method is SVM-RFE (support vector machine based recursive feature elimination (Guyon, et al., 2002)). It was originally proposed for gene selection, where a linear version of Support Vector Machine (SVM) is used as the learning algorithm in a recursive procedure to select a subset of genes for cancer classification. In our research, a nonlinear version of SVM-RFE (with Gaussian kernel) is used for feature subset selection and texture pattern classification. In the following sections, we briefly review the SVM classification method and SVM-RFE feature selection method.

5.4.1 *Support Vector Machine*

Support vector machines have been very popular in solving classification problems. The method consists of mapping the input vector $X \in \Re^n$ into a high

dimensional space H and constructing an optimal hyper-plane to achieve maximum separation between different classes in this space (Vapnik, 1998). The mapping $X \mapsto \Phi(X) \in H$ is performed by using a kernel function $K(\cdot,\cdot)$ which is defined as an inner product in H :

$$K(X_i, X_j) = \Phi(X_i) \cdot \Phi(X_j) \tag{5.17}$$

The kernel plays an important role in SVMs. By using kernels, all necessary computations are performed directly in input space and the mapping function Φ does not need to be known explicitly. Commonly used kernels include linear kernel, polynomial kernel and Gaussian kernel. The Gaussian kernel is used in this research which is defined as

$$K(X_i, X_j) = \exp(-\| X_i - X_j \|^2 / 2\sigma^2),$$

where σ is the spread width of the Gaussian kernel.

For a typical classification problem with the training data set $\{(X_k, y_k) \mid k = 1,...,l\}$, where $y_k \in \{+1,-1\}$ is the class label of X_k, finding the discriminant function $f(X) = W \cdot \Phi(X) + b$ can be formulated into solving the following quadratic optimization problem (Burges, 1998):

$$\min_{\alpha} J(\alpha) = \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i \alpha_j y_i y_j K(X_i, X_j) - \sum_{i=1}^{l}\alpha_i \tag{5.18}$$

subject to $0 \le \alpha_i \le C$ and $\sum_{i=1}^{l}\alpha_i y_i = 0$ $(i = 1,2,...,l)$. The hyperplane decision function can then be expressed by using α_i's (Burges, 1998):

$$f(X) = \sum_{k=1}^{l} \alpha_k y_k K(X_k, X) + b \tag{5.19}$$

Since the Gaussian kernel used in this research is a nonlinear mapping function, the linear discriminant function constructed by the SVM in the feature space corresponds to a nonlinear function in the original input space.

5.4.2 *Support Vector Machine based Recursive Feature Elimination*

SVM-RFE starts with full sets of features and nested subsets of features are selected in a backward sequential elimination manner (Guyon, et al., 2002). The features ranking criterion is the change in the objective function when one feature is removed. When the Gaussian kernel is used, the recursive feature elimination (RFE) procedure can be summarized as follows:

(a) **Start**: Initialize the subset of surviving features $S = [1,2,...,n]$ where n is the number of all the features and feature ranked list $R = [\ ]$;

(b) **Repeat** following steps until the required number of features remains.

(b.1) Train the non-linear SVM (using Gaussian kernel) with all the training data and features in S;

(b.2) For each feature v in S, compute the change in cost function caused by removing this input component. The resulting ranking coefficient is:

$$\Delta J = \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \left[\alpha_i \alpha_j y_i y_j K(X_i, X_j) - \alpha_i \alpha_j y_i y_j K_{-v}(X_i, X_j) \right], \tag{5.20}$$

where the notation (-v) means that the feature component v has been removed.

(b.3) Remove the feature component which corresponds to the smallest ΔJ

(b.4) Update surviving features subset $S = S - [v]$ and feature ranked list $R = [v, R]$;

The performance of SVM classifier is expected to be improved after the redundant or irrelevant features have been removed.

5.5 *Experimental Results and Discussion*

The USDA-supplied black walnut samples were cracked and a total of 851 samples (430 walnut meat and 421 shells) were randomly selected and labeled accordingly (Table 5.1). Among the 430 meat samples, 213 were light meat and 217 were dark meat. From the 421 shells, 212 of them were placed with their inner surface facing the camera and the remaining 209 samples were placed with their outer surfaces facing the camera. As we mentioned in section 5.2, the texture analysis will only be performed on the foreground area of each image and different texture features (equations 5.1, 5.3, 5.11, 5.13, 5.15 and 5.16) were extracted and concatenated into one vector for each sample. Since different features are in a different scale, they need to be standardized. This was done by subtracting the mean from each attribute, and then dividing over its standard deviation. As a result, each element of the feature vector has a mean of zero and a standard deviation of one. For all the samples, one-third was reserved for testing and the other two-thirds for training. Stratified random sampling (Witten and Frank, 2005) was performed to guarantee that each sample class (meat/shell) was properly represented in both training and testing sets. The SVM-RFE method was then used to select the feature subsets and obtain a trained SVM classifier based on the training samples. This classifier was further tested on the testing data set for accuracy estimation.

Table 5.1 Training and test sample sets

Samples Category	# Training Samples	# Test Samples	Total
Shell	281	140	421
Meat	287	143	430
Subtotal	568	283	851

Test error on the testing data set is usually used to assess the performance of a classifier. However, the number of testing samples is not unlimited and the test error may be biased due to a not optimized partition of training and testing sets. Thus, we partitioned the total samples into a training set or a testing set by stratified random sampling for 100 times. Each time we performed SVM-RFE on the training set and tested the performance of the trained SVM on the testing set. The mean and standard deviation of the performance measurements were computed on these 100 trials. Apart from test error, "shell accuracy" and "meat accuracy" were also calculated. "shell accuracy" is defined as the probability a true shell sample will be classified as shell while "meat accuracy" is the probability a true meat sample will be classified as meat. Using the Table 5.2, the definitions can be written as

$$Shell\ Accuracy = \frac{a}{a+c} \tag{5.21}$$

$$Meat\ Accuracy = \frac{d}{b+d} \tag{5.22}$$

where a, b, c and d are defined in Table 5.2.

Table 5.2 Classification confusion matrix

	Walnut Shell	Walnut Meat
Classified as Shell	a	b
Classified as Meat	c	d

5.5.1 *Select Optimal Number of Features*

The SVM-RFE is a feature ranking method which does not indicate the optimum number of features needed. In this research, we found the performance of the classifier was quite stable over a large range of selected feature numbers, as shown in Figure 5.8. The average classification accuracy was better than 99% with the used feature number varying from 5 to 34. This suggested that feature number was not critical for the final performance in our case. As long as the best 5 features were selected, the next 29 best features did not significantly interfere with the classifier performance. However, if more than 34 features were used, the classification accuracy started to decrease.

The mean and STD classification accuracy of the training samples were also plotted in Figure 5.8. It shows that the mean accuracy reached 100% when 6 features are selected and stay constant if more features were added. The different trend of the training and testing accuracy curves show the problem of curse of dimensionality (Bishop, 2006) that the performance of a classifier will degrade with the increase of the input feature space dimension. Although SVM classifier has an effective mechanism for alleviating the curse of dimensionality problem by margin maximization, it is still beneficial to perform feature selection first. In this research,

the top 6 ranked features were selected to achieve good balance between the classification accuracy and the input feature dimension.

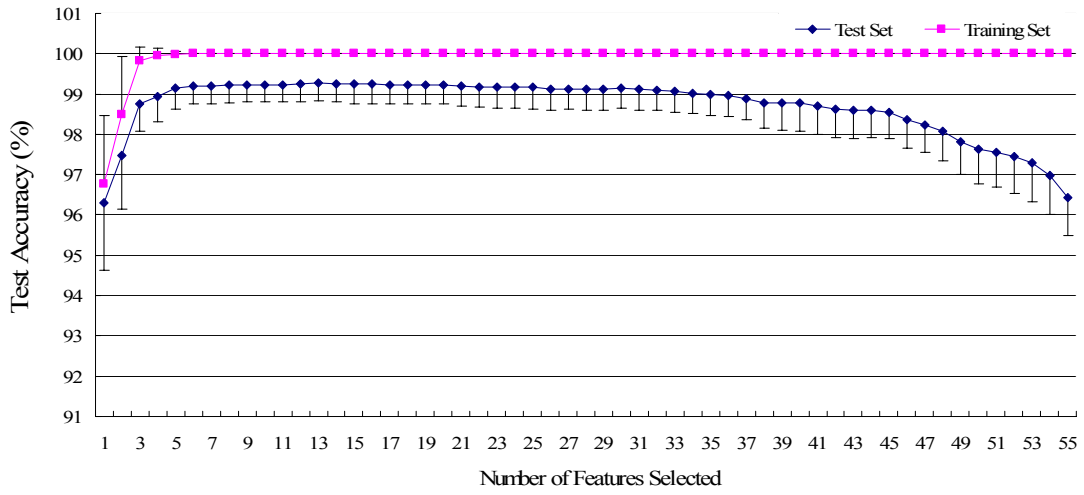


Figure 5.8 Estimated classification accuracy in training and testing data set vs. number of selected features using SVM-RFE. The mean and STD value of the 100 trails are shown in the figure. To prevent overlap, only plus error bars (+1 STD) of training set and minus error bars (-1 STD) of testing set are displayed.

5.5.2 *SVM versus SVM-RFE*

Table 5.3 compares the performances with using only the top 6 features and all of the 56 features. The average test accuracy was 96.17% using the full feature set, while it increased to 99.19% by using just 6 of the 56 features. In both cases, the mean specificities were both high (> 99%) while the sensitivity did show large differences. The average "shell accuracy" was 99.01% using the top 6 features, which dropped to 92.51% if all the features were used. This observation suggests that scientifically selecting a subset of features not only improves the efficiency of classification algorithms, but also improves the prediction accuracy.

Table 5.3 Performance (mean ± standard deviation) comparison of SVM classifier with and without RFE feature selection.

Method	# Features	Accuracy (%)	Shell Accuracy (%)	Meat Accuracy (%)
SVM-RFE	6	99.19 ± 0.45	99.01 ± 0.86	99.37 ± 0.58
SVM	56	96.17 ± 0.98	92.51 ± 1.98	99.75 ± 0.39

5.5.3 *The Performance using Each Texture Feature Category*

To further check the performance of each texture feature category (LBP+VAR, Gabor, Wavelet, CS-GLCM and image histogram based statistics), we tested the SVM classification performance by using features only from one of the categories. The results in Table 5.4 suggest that LBP_VAR, Gabor and Wavelet worked better than CS-GLCM or histogram based feature sets. Gabor was the best with an average test error as low as 1.25%, while the histogram based features gave a 4.22% error.

The ranking of the features can also be illustrated based on SVM-RFE feature sorting results. The top 6 selected features' categories were recorded every time for the 100 trials. The total occurring times of each feature category was counted and sorted in Table 5.5. Among the top 6 features, two of them were Gabor features (No. 3 and 4), two belonged to LBP_VAR category (No. 1 and 6), one was wavelet feature (No. 2) and the one was CS-GLCM feature (No. 5). Gabor and Wavelet features ranked as the top 6 features more times compared to CS-GLCM. None of the histogram based features was ranked as top 6. This result matches the observation from Table 5.4. It is not surprising that histogram based features give a poorer classification result, since no spatial information of the image is taken into account. Some walnut meats are relatively thicker than average and the corresponding backlit

images are darker. The histograms of these images are quite similar to those of walnut shells, therefore making it difficult to separate them from shells. However, the 2-D texture of these dark meat images are still different from those of shell, which can be characterized by the more sophisticated texture analysis methods and be classified correctly.

In theory, complementary features can help ameliorate classification rate. This could be shown in the following two tests. Table 5.4 shows that the classification performance of CS-GLCM features was worse than LBP even though they both similarly describe texture patterns in a local neighborhood. The reason is that only second-order (account for pixel pairs) statistical features are extracted from CS-GLCM. By simply adding the mean gray value of the image (account for single pixels) to the current CS-GLCM features, the average performance increased from 96.7% to 98.7%. Another complementary feature example is LBP and VAR. An accuracy of 98.1% and 98.4% can be achieved using only LBP or VAR respectively, but if combined, the accuracy increased to 98.5%. The two tests demonstrated that the classification performance can be improved by using complementary feature set. In reality, to find complementary features in a low dimension feature set is relatively easy. However, it is extremely difficult when large sets of features are involved. In this case, feature selection is necessary to find out the useful complementary feature subsets to reduce the feature dimensionality.

Table 5.4 Performances (mean ± standard deviation) of the SVM classifier using different feature category.

Feature Category	# Features	Test Error (%)	Shell Accuracy (%)	Meat Accuracy (%)
LBP+VAR	21	1.47 ± 0.73	97.36 ± 1.40	99.73 ± 0.38
Gabor	8	1.25 ± 0.51	98.77 ± 0.5	98.73 ± 0.97
Wavelet	4	1.67 ± 0.7	98.21 ± 1.61	98.45 ± 1.05
CS-GLCM	10	3.28 ± 0.8	96.96 ± 1.54	96.49 ± 1.32
Histogram based Statistics	13	4.22 ± 0.6	96.1 ± 1.57	95.48 ± 2.05

Table 5.5 Top 6 feature ranking result of 100 runs of SVM-RFE

Feature Rank	Feature Category	Comments
1	LBP_VAR	The No.1 ranked feature is the LBP histogram bin corresponding to the LBP pattern "00000111".
2	Wavelet	The No.2 ranked feature is the image energy percentage corresponding to the wavelet channel LL3
3	Gabor	The No.3 ranked feature is the Mean value of the 1st level (lowest frequency) Gabor filtered image magnitude.
4	Gabor	The No.4 ranked feature is the STD value of the 1st level (lowest frequency) Gabor filtered image magnitude
5	CS-GLCM	The No.5 ranked feature is the Entropy of the circular co-occurrence matrix with radius equals 1.
6	LBP_VAR	The No.5 ranked feature is the 10^{th} centile value of the normalized cumulative VAR histogram.

5.5.4 *Practical Considerations*

The backlight imaging system shown in Figure 5.2 is the initial proof-of-principle setup for data collection. To design an automatic walnut meat and shell

separation machine, several factors need to be taken into account including material handling and system speed.

Currently, samples were kept static on the stage glass when backlit images were acquired. For an automated machine, the cracked walnut could be first separated one by one using a vibratory feeder then dropped onto a semi-transparent conveyor belt which should allow backlighting. One kind of the semi-transparent belts has been successfully used for automatic quality grading of mandarin segments as shown in Figure 5.9 (Blasco, et al., 2007). Figure 5.10 shows our proposed real-time backlighting imaging system using the semi-transparent conveyor. The light source is placed under the conveyor and a line-scan camera could be used to continuously capture the backlit images, which will be transferred to the computer. Real-time image texture analysis and pattern recognition will be performed at the computer. If shell fragments are detected, the computer can control the rejection device to remove them off the conveyer.

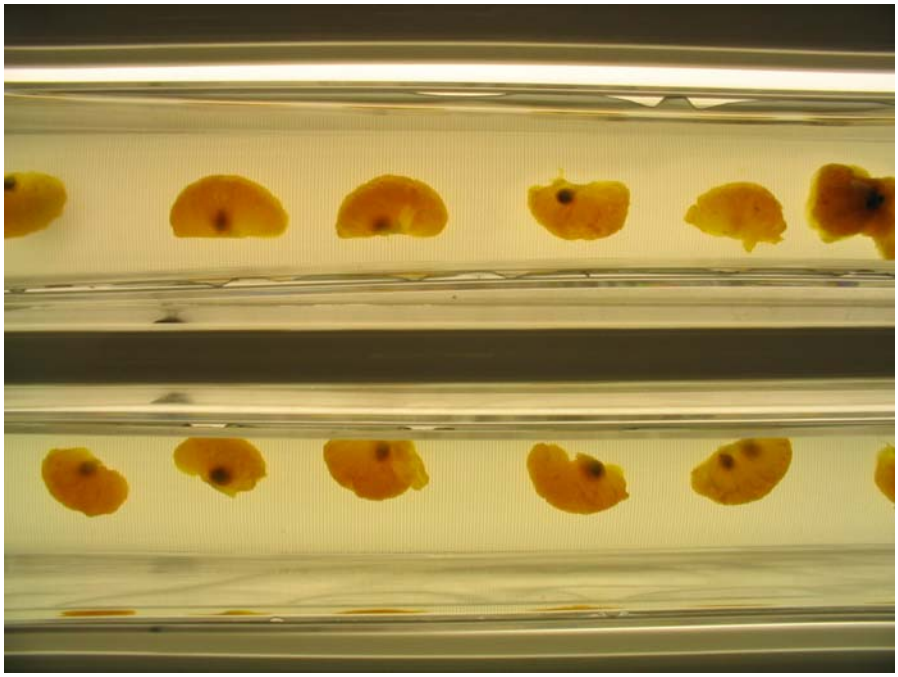


Figure 5.9 Semi-transparent conveyor belt used for automatic quality grading of mandarin segments (photo provided by Dr. José Blasco, Valencian Institute for Agricultural Research, Spain.)

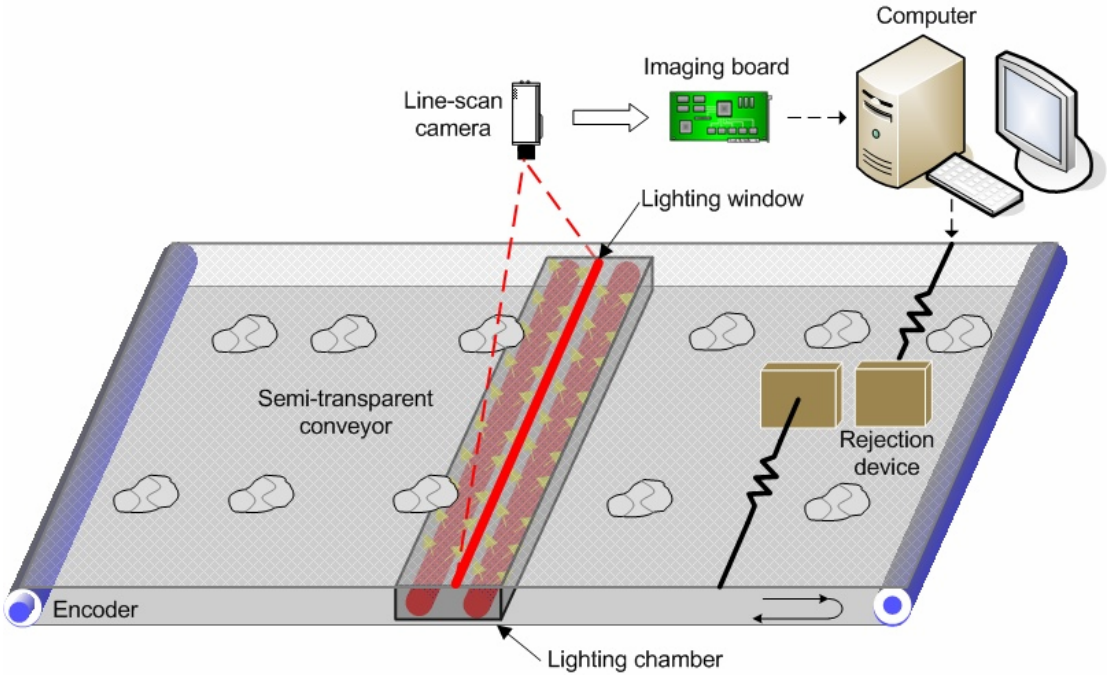


Figure 5.10 Proposed real-time imaging system for automatic walnut shell and meat separation. The light source is placed under the conveyor and a line-scan camera is used to continuously capture the backlit walnut images. Real-

time image texture analysis and pattern recognition is performed by the computer and any detected shell fragments can be further removed by the rejection device.

To ensure industrial throughput, the conveyor would need to travel at a high speed, requiring both feature extraction and pattern classification be performed at high speed. Based on our experience, an image resolution of 0.2 mm/pixel is sufficient for walnuts inspection. This would allow for an image size of 100×100 pixels for an average size walnut fragment ($20\text{mm} \times 20\text{mm}$). The computational cost is approximately 18 ms for feature calculation and 0.2 ms for the trained SVM classification (running on a PC with Pentium 4 1.8 GHz CPU and 768 MB RAM). Since the processing time for each walnut sample is less than 20 ms, the system should be able to separate more than 50 walnut fragments per second.

5.6 *Conclusion*

In this study, we proposed a backlight imaging system for black walnut meat and shell separation based on the distinct texture patterns of their backlit images. Several rotation invariant feature analysis methods, consisting of circularly symmetric gray-level co-occurrence matrix, circular Gabor filters, wavelet transforms and image histogram based statistics, were implemented to fully characterize these texture patterns. SVM-RFE method was used to select a feature subset and to improve the SVM classification accuracy. The experimental results demonstrated that the proposed approach is very effective in walnut meat and shell separation, with an overall 99.2% separation accuracy being achieved, using only the 6 top ranked features.

Some practical implementation factors, such as material handling and system speed are also briefly discussed and a real-time backlighting imaging system for automatic walnut shell and meat separation was proposed. Since the proposed backlighting imaging system requires only a little equipment, the whole system can be built at a very low price. This low instrument cost and high classification rate make the proposed imaging system a great potential in walnut processing industry.

Chapter 6: Conclusions

This research explored two machine vision applications in food inspection for higher quality and safety. The first application of a combined X-ray and laser range imaging system, we focused on developing an efficient and precise system calibration method. For the second application, we concentrated on building the effective imaging system to automatically separate walnut shells and meat. The details of system setup, texture image analysis, feature selection, and pattern classification have all been explained. The overall objectives of this research described in Chapter 2 have been achieved. The following conclusions were drawn from the research:

(1) A unique three-step system calibration method for the combined X-ray and laser range imaging system has been developed. The calibration procedures can be performed in a semi-automatic way and can be finished in a timely fashion, requiring minimum knowledge for system calibration.

(2) Results showed that high accuracy has been achieved for the whole system calibration – a root mean square error of 0.20mm, a standard deviation of 0.20mm, and a maximum error of 0.48mm.

(3) A backlight imaging system has been developed and built for walnut shells and meat separation based on the finding that backlit images of walnut shell and meat have distinct texture patterns due to their different light transmittance properties. Extensive texture analysis methods, feature selection and pattern classification algorithms have been explored to characterize and separate the unique texture patterns.

(4) Results showed that the proposed texture analysis based backlight imaging system can achieve an overall 99.2% separation accuracy. This high separation accuracy and low instrument cost make the proposed imaging system a great benefit to the walnut processing industry.

Chapter 7: Suggestions for Further Study

The system calibration method we proposed in the first study is easy to implement with high accuracy. However, we should note that our calibration is based on the assumption that the X-ray and laser beam are coplanar. This assumption is justified via system mechanical design and careful alignment during system setup. In the case of non-coplanarity of the X-ray and laser beam, it will lower the calibration precision. Future study can be conducted to identify a more comprehensive model to resolve this problem.

For the backlight imaging system for walnut shells and meat separation, future effort can be focused on developing a real-time system. During this procedure, some important factors, such as image motion blur and illumination intensity, need to be taken into account. When walnut fragments travel on the conveyor, the images captured will be blurred to an extent dependent on the conveyor speed. The illumination intensity can also affect the quality of the final acquired images. The effects of these factors can be fully analyzed in future research. We also need to mention that the size of all walnut samples used in this study is bigger than $5\text{mm} \times 5\text{mm}$. The texture analysis result may become unstable when smaller samples are involved. Future experiments can be performed on small walnut fragments to stabilize the texture analysis methods and provide an optimized uniform process for all nut sizes.

Appendices A

The definitions of the 5 statistical features used to characterize the circular co-occurrence matrixes are shown below.

$$\textbf{Entropy}: \quad Ent_k = -\sum_{i=0}^{15}\sum_{j=0}^{15} C_k(i,j)\log_2 C_k(i,j) \quad k = 1,2 \qquad \text{(A.1)}$$

where $C_k(i,j)$ is the ith row and jth column element of the normalized circular co-occurrence matrix with radius k. Note the 8-bit gray images were first recalled into 16 gray levels so the size of each circular co-occurrence matrix is 8×8. Entropy measures the randomness of the joint gray level distribution. The entropy is expected to be high if the joint gray-levels are randomly distributed.

$$\textbf{Energy}: \qquad Eng_k = \sum_{i=0}^{15}\sum_{j=0}^{15} C_k(i,j)^2 \quad k = 1,2 \qquad \text{(A.2)}$$

Energy measures the number of repeated pairs. The energy is expected to be high if the occurrence of repeated pixel pairs is high.

$$\textbf{Contrast}: \qquad Con_k = \sum_{i=0}^{15}\sum_{j=0}^{15} (i-j)^2 C_k(i,j) \quad k = 1,2 \qquad \text{(A.3)}$$

Contrast measures the local contrast of an image. The contrast is expected to be low if the gray levels of each pixel pair are similar.

$$\textbf{Homogeneity}: \qquad Hom_k = \sum_{i=0}^{15}\sum_{j=0}^{15} \frac{C_k(i,j)}{1+|i-j|} \quad k = 1,2 \qquad \text{(A.4)}$$

Homogeneity measures the local homogeneity of a pixel pair. The homogeneity is expected to be large if the gray levels of each pixel pair are similar.

$$\textbf{\textit{Correlation}}: \quad Cor_k = \sum_{i=0}^{15}\sum_{j=0}^{15} \frac{(i-\mu)(j-\mu)C_k(i,j)}{\sigma^2} \quad k = 1,2 \qquad \text{(A.5)}$$

where μ and σ are the mean and standard deviation of the rescaled gray value in the image. It provides a correlation between the two pixels in the pixel pair. The correlation is expected to be high if the gray levels of the pixel pairs are highly correlated.

Appendices B

The definitions of the 6 statistical features used to characterize image histogram patterns are shown below.

Mean: $$Mean_{NH} = \frac{1}{N}\sum_{k=1}^{N} Val_k \times NH_k \quad \text{(B.1)}$$

Variance: $$Var_{NH} = \sum_{k=1}^{N} (Val_k - Mean_{NH})^2 NH_k \quad \text{(B.2)}$$

Entropy: $$Ent_{NH} = -\sum_{k=1}^{N} [NH_k \times \log_2 NH_k] \quad \text{(B.3)}$$

Energy: $$Eng_{NH} = \sum_{k=1}^{N} NH_k^2 \quad \text{(B.4)}$$

Skewness: $$Ske_{NH} = \sum_{k=1}^{N} \frac{(Val_k - Mean_{NH})^3 \times NH_k}{Var_{NH}^{3/2}} \quad \text{(B.5)}$$

Kurtosis: $$Kur_{NH} = \sum_{k=1}^{N} \frac{(Val_k - Mean_{NH})^4 \times NH_k}{Var_{NH}^2} - 3 \quad \text{(B.6)}$$

where $NH_k, (k = 1,2,...,N)$ is the normalized image histogram and N is the number of the histogram bins.

Mean and *Varaince* measure the mean and variance of the gray level of the quantized image; *Entropy* measures the randomness of the gray value distribution; *Energy* measures the occurrence of repeated gray level pixels; *Skewness* measures of symmetry of the normalized histogram; *Kurtosis* measures the peakedness or flatness of the histogram relative to a normal distribution.

Bibliography

Abdel-Aziz, Y. I. and H. M. Karara. 1971. Direct linear transformation into object space coordinates in close-range photogrammetry. In *Proceeding of the Symposium on Close-Range Photogrammetry*, 1-18. Urbana, Illinois.

Almuallim, H., and T. Dietterich. 1994. Learning Boolean Concepts in the Presence of Many Irrelevant Features. *Artificial Intelligence* 69(1-2):279-305.

AS&E (2008). Cargo and Vehicle Inspection. [Online]. Available WWW: http://www.as-e.com/products_solutions/cargo_vehicle_inspection.asp

Bandeira, L., J. Saraiva and P. Pina. 2007. Impact Crater Recognition on Mars Based on a Probability Volume Created by Template Matching. *IEEE Transactions on Geoscience and Remote Sensing* 45(12):4008-4015.

Basu, M., H. Bunke and A. D. Bimbo. 2005. Guest Editors' Introduction to the Special Section on Syntactic and Structural Pattern Recognition. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 27(7):1009-1012.

Bishop, C. M. 1995. Neural Networks for Pattern Recognition. Oxford, U.K.: Clarendon Press.

Bishop, C. M. 2006. Pattern Recognition and Machine Learning. Singapore: Springer.

Blasco, J., S. Cubero, R. Arias, J. Gómez, F. Juste and E. Moltó. 2007. Development of a computer vision system for the automatic quality grading of mandarin segments. In *Pattern Recognition and Image Analysis.* eds. J. Martí, J. M. Benedí, A.M. Mendonça and J. Serrat, 460-465. Berlin, Germany: Springer-Verlag.

Bovik, A. C., M. Clark and W. S. Geisler. 1990. Multichannel Texture Analysis Using Localized Spatial Filters. *IEEE Transactions on Pattern Analysis and Machne Intelligence* 12(1):55-73.

Bradley, P. S. and O. L. Mangasarian. 1998. Feature selection via concave minimization and support vector machines. In *Proc. 15th International Conference on Machine Learning (ICML98)*, 82-90. San Francisco, CA.

Burges, C. 1998. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* 2(2):955-974.

Casasent, D., A. Talukder, P. Keagy and T. Schatzki. 2001. Detection and segmentation of items in X-ray imagery. *Transactions of the ASAE* 44(2):337-345.

Casasent, D. and X. Chen. 2003. New training strategies for RBF neural networks for Xray agricultural product inspection. *Pattern Recognition* 36(2):535-547.

Chan, C. and G. K. H. Pang. 2000. Fabric defect detection by Fourier analysis. *IEEE Transactions on Industry Applications* 36(5):1267-1276.

Chanyagorn, P. and K. B. Eom. 2000. Texture segmentation using moving average modeling approach. In *Proc. International Conference on Image Processing*, 2:116-119. Vancouver, Canada.

Charalampidis, D. and T. Kasparis. 2002. Wavelet-based rotational invariant roughness features for texture classification and segmentation. *IEEE Transactions on Image Processing* 11(8):825-837.

Chen, F., G. M. Brown and M. Song. 2000. Overview of three-dimensional shape measurement using optical methods. *Optical Engineering* 39(1):10-22.

Chen, X. 2003. Detection of physical hazards in boneless poultry product using combined X-ray and laser range imaging technologies. Ph.D. diss., Biological Resources Engineering Department, University of Maryland, College Park.

Cristianini, N. and J. Shawe-Taylor. 2000. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge: Cambridge University Press.

Daubechies, I. 1992. Ten Lectures on Wavelets. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Daugman, J. G. 1988. Complete Discrete 2D Gabor Transforms by Neural Networks for Image Analysis and Compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 36(7):1169-1179.

Davies, E. R. 2005. Machine Vision: Theory, Algorithms, Practicalities (3rd Ed.). San Francisco, CA: Morgan Kaufmann Publishers.

Deng, H. and D. A. Clausi. 2004. Gaussian MRF rotation-invariant features for image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(7):951-955.

Devroye, L., L. Gyorfi and G. Lugosi. 1996. A Probabilistic Theory of Pattern Recognition. Berlin: Springer-Verlag.

Dorsch, R. G., G. Hausler, and J. M. Herrmann. 1994. Laser triangulation: fundamental uncertainty in distance measurement. *Applied Optics* 33(7):1306-1314.

Dowsett, D. J., P. A. Kenny and R. E. Johnston. 2006. Physics of Diagnosis Imaging. London: Hodder Arnold.

Duda, O. R., P. E. Hart and D. G. Stork. 2000. Pattern Classification (2nd Ed.). New Youk, NY: Wiley-Interscience.

Fahrig, R., A. Ganguly, P. Lillaney, J. Bracken, J. A. Rowlands, Z. Wen, H. Yu, V. Rieke, J. M. Santos, K. B. Pauly, D. Y. Sze, J. K. Frisoli, B. L. Daniel and N. J. Pelc. 2008. Design, Performance, and Applications of a Hybrid X-Ray/MR System for Interventional Guidance. *Proceedings of the IEEE* 96(3):468-480.

Faugeras, O. D. and G. Toscani. 1986. The calibration problem for stereo. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'86)*, 15-20. Miami, FL.

Fu, K. S. 1982. Syntactic Pattern Recognition and Applications. New Jersey: Prentice-Hall.

Gharavi-Alkhansari, M. 2001. A fast globally optical algorithm for temperate matching using low-resolution pruning. *IEEE Transaction on Image Processing* 10(4):526-533.

Gill, P. E. and W. Murray. 1978. Algorithms for the solution of the nonlinear least-squares problem. *SIAM Journal of Numerical Analysis* 15(5):977-992.

Gonzalez, R. C. and R. E. Woods. 2007. Digital Image Processing (2nd Ed.). Upper Saddle River, NJ: Prentice Hall.

Graves, M. and B. Batchelor. 2003. Machine vision for the inspection of natural products. London: Springer.

Gupta, R. and R. Hartley. 1997. Linear pushbroom cameras. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 19(9):963–97.

Guyon, I., J. Weston, S. Barnhill and V. Vapnik. 2002. Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning* 46(1-3):389-422.

Guyon, I. and A. Elisseeff. 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research* 3:1157-1182.

Guyon, I., S. Gunn, M. Nikravesh and L. Zadeh (eds.). 2006. Feature Extraction, Foundations and Applications Series Studies in Fuzziness and Soft Computing. Germany: Springer Verlag.

Hall, E. L., J. B. K. Tio, C. A. McPherson and F. A. Sadjadi. 1982. Measuring curved surfaces for robot vision. *Computer* 15(12):42-54.

Hammons Products Company. 1998. World Consumption and Production Trends: International Nut Council.

Hammons Products Company (2007). It's Black Walnut Harvest Time. [Online]. Avalable WWW: http://www.hammonsproducts.com/page.asp?p_key=087DBA40341B448AB2EE569783614DFA

Hamamatsu Photonics (2008). Transmission Mode Soft X-ray Tube N7599, N7599-01. [Online]. Available WWW: http://209.73.52.252/assets/pdf/parts_N/N7599_series.pdf

Haralick, R. M., I. Dinstein, and K. Shanmugam. 1973. Textural features for image classification. *IEEE Transaction on Systems, Man, and Cybernetics* 3:610-621.

Haralick, R. M. 1979. Statistical and Structural Approaches to Texture. *Proceedings of IEEE* 67(5):786-804.

Hartley, R. and A. Zisserman. 2000. Multiple View Geometry. Cambridge, UK: Cambridge University Press.

Hatcher, R. L., L. A. Johnson, G. M. Hopper, J. W. Pease and J. E. Johnson (1998). Tree crops for Marginal Farmland Black Walnut: Virginia Cooperative Extension. [Online]. Available WWW: http://www.ext.vt.edu/pubs/forestry/446-602/446-602.html

Hartman, L. R. (2001). X-ray inspection boosts confidence through the pipeline. *Packaging digest*, 54. [Online]. Available WWW: http://findarticles.com/p/articles/mi_hb4923/is_200110/ai_n18061451

He, X., S. An, P. Shi. 2007. Statistical Texture Analysis-Based Approach for Fake Iris Detection Using Support Vector Machines, *International Conference on Biometrics*, 540-546. Seoul, Korea.

Heikkila, J. 2000. Geometric Camera Calibration Using Circular Control Points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1066-1077.

Iacobaeus, C., T. Francke, B. Lund-Jensen, J. Ostling, P. Pavlopoulos, V. Peskov, and F. Tokanai. 2007. A high position resolution X-ray detector: an "Edge on" illuminated capillary plate combined with a gas amplification structure. *Transactions on Nuclear Science* 53(2):554-561.

Jain, A. K. and F. Farrokhnia. 1991. Unsupervised Texture Segmentation Using Gabor Filters. *Pattern Recognition* 24(12):1167-1186.

Jain, A. K., R. P. W. Duin and J. Mao. 2000. Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(1):4-37.

Jalkio, J. A., R. C. Kim and S. K. Case. 1985. Three dimensional inspection using multistripe structured light. *Optical Engineering* 24(6):966-974.

Jamieson, Valerie (2002). Physics raises food standards. [Online]. Available WWW: http://physicsweb.org/article/world/15/1/3

Joshi, M.V., L. Bruzzone and S. Chaudhuri. 2006. A Model-Based Approach to Multiresolution Fusion in Remotely Sensed Images. *IEEE Transactions on Geoscience and Remote Sensing* 44(9):2549-2562

Jin, F., L. Qin, L. Jiang, B. Zhu and Y. Tao. 2008. Novel Separation Method of Black Walnut Meat from Shell using Invariant Features and a Supervised Self-Organizing Map. *Journal of Food Engineering* 88(1):75-85.

Jing, H. 2003. Laser range imaging for on-line mapping of 3-D images to pseudo-X-ray images for poultry bone fragment detection. Ph.D. diss., Biological Resources Engineering, University of Maryland, College Park.

John, G. H., R. Kohavi and K. Pfleger. 1994. Irrelevant Features and the Subset Selection Problem. In *Proc. 11th International Conference on Machine Learning ICML94*, 121-129. San Francisco, CA.

Jones, J. E., R. Mueller and J. W. V. Sambeek. 1998. Nut Production Handbook for Eastern Black Walnut. Republic, Missouri: Southwest Missouri Resources, Conservation & Development (RC&D), Inc.

Julesz, B. 1975. Experiments in the Visual Perception of Texture. *Scientific American* 232(4):34-43.

Kamarainen, J. K., V. Kyrki and H. Kalviainen. 2006. Invariance properties of Gabor filter-based features-overview and applications. *IEEE Transactions on Image Processing* 15(5):1088-1099.

Khademi, A. and S. Krishnan. 2007. Shift-invariant discrete wavelet transform analysis for retinal image classification. *Medical and Biological Engineering and Computing* 45(12):1211-1222.

Kira, K. and L. A. Rendell. 1992. The Feature Selection Problem: Traditional Methods and a New Algorithm. In *Proc. 10th National Conference of Artificial Intelligence*, 129-134. San Jose, CA.

Klette, R., K. Schlüns and A. Koschan. 1998. Computer Vision: Three-dimensional Data from Images. Singapore: Springer-Verlag Singapore Pte. Ltd.

Kohavi, R. and G. H. John. 1997. Wrappers for feature subset selection. *Artificial*

Intelligence 97(1-2):273-324.

Kohonen, T. 2001. Self-Organizing Maps, New York, NY: Springer-Verlag.

Kokare, M., P. K. Biswas and B. N. Chatterji. 2005. Texture image retrieval using new rotated complex wavelet filters. *IEEE Transactions on Systems, Man, and Cybernetics* 35(6):1168-1178.

Krishnan, P. and A. Berlage. 1984. Separation of Shells from Walnut Meats Using Magnetic Methods. *Transaction of the ASAE* 27(6):1990-1992.

Kumar, A., and G. K. H. Pang. 2002. Defect detection in textured materials using Gabor filters. *IEEE Transactions on Industry Applications*, 38(2):425-440.

Kuncheva, L. I. 2004. *Combining Pattern Classifiers*, Hoboken, New Jersey: John Wiley & Sons, Inc.

Lee, S., T. Stathaki and F. Harris. 2003. Texture characterization using a novel optimization formulation for two-dimensional autoregressive modeling and K-means algorithm. *37th Asilomar Conference on Signals, Systems and Computers*, 2, 1605-1609. Pacific Grove, CA.

Lemacks, M. R., S. Cheenu Kappadath, C. C. Shaw, X. Liu and G. J. Whitman. 2002. A dual-energy subtraction technique for microcalcification imaging in digital mammography - A signal-to-noise analysis. *Medical Physics* 29(8):1739-1751.

Liu, H. and H. Motoda. 1998. Feature Selection for Knowledge Discovery and Data Mining. Nowwell, MA: Kluwer Academic Publishers.

Möller, B. 2005. Full frame 3-D snapshot - Possibilities and limitations of 3-D image acquisition without scanning. Ph.D. diss. Department of Electrical Engineering, Linköping University, Linköping, Sweden.

Molloi, S., A. Ersahin and Y. Qian. 1995. CCD camera for dual-energy digital subtraction angiography. *IEEE Transactions on Medical Imaging* 14(4):747-752.

Nobel, A., R. Hartley, J. Mundy and J. Farley. 1994. X-ray metrology for quality assurance. *IEEE International Conference on Robotics and Automation* 2:1113-1119.

Nobel, J. A., R. Gupta, J. Mundy, A. Schmitz and R. I. Hartley. 1998. High Precision X-ray Stereo for Automated 3-D CAD-Based Inspection. *IEEE Transactions on Robotics and Automation* 14(2):292-303.

Ojala, T., Pietikäinen, M. and Harwood D. 1996. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition* 29(1):51-59.

Ojala, T., M. Pietikäinen and T. Mäenpää. 2000. Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. In *Proc. 6th European Conference on Computer Vision, Part I*, 404-420. Dublin, Ireland.

Ojala, T., M. Pietikäinen and T. Mäenpää. 2002. Multiresolution Gray-Scale and Rotation Invariant Texture Classification. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 24(7):971-987.

Pearson, T. and R. Young. 2002. Automated Sorting of Almonds with Embedded Shell by Laser Transmittance Imaging. *Applied Engineering Agriculture* 18(5):637-641.

Petrou, M. and P. G. Sevilla. 2006. Image Processing: Dealing with Texture. West Sussex, England: John Wiley & Sons Ltd.

Peters, T. M. 2006. Image-guidance for surgical procedures. *Physics in Medicine and Biology* 51(14):505-540.

Roh, Y. J., W. S. Park and H. Cho. 2003. Correcting image distortion in the X-ray digital tomosynthesis system for PCB solder joint inspection. *Image and Vision Computing* 21(12):1063-1075.

Porter, R. and N. Canagarajah. 1997. Robust rotation-invariant texture classification: wavelet, Gabor filter and GMRF based schemes. *IEE Proceedings - Vision, Image & Signal Processing* 144(3):180-188.

Manjunath, B. S. and W. Y. Ma. 1996. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18(8):837-842.

McGlone, C., E. Mikhail and J. Bethel. 2004. Manual of Photogrammetry (5th Ed.). Falls Church, Virginia: ASPRS publisher.

Ott, R. L. and M. Longnecker. 2001. An Introduction to Statistical Methods and Data Analysis (5th Ed.). Pacific Grove, CA: Duxbury.

Quinlan, J. R. 1993. C4.5: Programs for Machine Learning. San Mateo, CA : Morgan Kaufmann Publisher, Inc.

Rakotomamonjy, A. 2003. Variable selection using SVM-based criteria. *Journal of Machine Learning Research* 3:1357-1370.

Ranganathan, Ananth (2004). The Levenberg-Marquardt Algorithm. [Online]. Available WWW: http://www-tatic.cc.gatech.edu/people/home/ananth/docs/lmtut.pdf

Rellier G., X. Descombes, F. Falzon, and J. Zerubia. 2004. Texture feature analysis using a gauss-Markov model in hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 42 (7):1543-1551.

Salvi, J., X. Armagué, J. Batlle. 2002. A comparative review of camera calibrating methods with accuracy evaluation. Pattern Recognition, 35(7):1617-1635.

Selman, J. 1994. The fundamentals of X-ray and radium physics. Springfield, IL: Charles C Thomas.

Seta, K. and T. O'ishi. 1990. Distance meter utilizing the intermode beat of a He-Ne laser. *Applied Optics* 29:354-359.

Shahin, M. A., E. W. Tollner, R. W. McClendon and H. R. Arabnia. 2002a. Apple classification based on surface bruises using image processing and neural networks. *Transactions of the ASAE* 45(5):1619-627.

Shahin, M. A., E. W. Tollner, R. D. Gitaitis, D. R. Sumner and B. W. Maw. 2002b. Classification of sweet onions based on internal defects using image processing and neural network techniques. *Transactions of the ASAE* 45(5):1613-618.

Soatto, S. 2007. 3-D Shape Estimation and Image Restoration: exploiting defocus and motion blur. London: Springer-Verlag.

Sonka, M., V. Hlavac and R. Boyle. 1999. Image Processing, Analysis, and Machine Vision (2nd Ed.). California: Books/Cole Publishing Company.

Starck, J. L., F. Murtagh and A. Bijaoui. 1998. Image and Data Analysis: the Multiscale Approach. Cambridge, United Kingdom: Cambridge University Press.

Sun, Z., G. Bebis and R. Miller. 2005. On-road vehicle detection using evolutionary Gabor filter optimization. *IEEE Transactions on Intelligent Transportation Systems* 6(2):125-137.

Szeles, C., S.A. Soldner, S. Vydrin, J. Graves and D.S. Bale. 2007. Ultra High Flux 2-D CdZnTe Monolithic Detector Arrays for X-Ray Imaging Applications. *IEEE Transactions on Nuclear Science* 54(4):1350-1358.

Taibi, A., S. Fabbri, P. Baldelli, C. di Maggio, G. Gennaro, M. Marziani, A. Tuffanelli and M. Gambaccini. 2003. Dual-energy imaging in full-field digital mammography: a phantom study. *Physics in Medicine and Biology* 48(13):1945-1956.

Tao, Y. and J. G. Ibarra. 2000A. Thickness-compensated X-ray imaging detection of bone fragments in deboned poultry - Model analysis. *Transactions of ASAE* 43(2):453-459.

Tao, Y. and J. G. Ibarra. 2000B. Far-infrared and 3-D imaging for doneness assessment in chicken breast. In *Proc. of SPIE*, 4206: 185-192. Boston, MA.

Tao, Y., Z. Chen, H. Jing and J. Walker. 2001. Internal inspection of deboned poultry using x-ray imaging and adaptive thresholding. *Transaction of ASAE* 44(4):1005-1009.

Tao, Y., S. Lo, M. Freedman and J. Xuan. 2007. A preliminary study of content-based mammographic masses retrieval. In *Proc. of SPIE*, 6514, 65141Z. San Diego, CA.

Trahanias, P. and E. Skordalakis. 1990. Syntactic pattern recognition of the ECG. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12(7):648-657.

Tsai, R. Y. 1987. A versatile camera calibration technique for high-accuracy 3-D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation* 3(4):323-344.

Tsatsanis, M. K. and G. B. Giannakis. 1992. Object and texture classification using higher order statistics. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 14(7):733-750.

Tuceryan, M. and A. K. Jain. 1990. Texture Segmentation Using Voronoi Polygons. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12(2):211-216.

Tümer, M. B., L. A., Belfore, II and K. M. Ropella. 2003. A syntactic methodology for automatic diagnosis by analysis of continuous time measurements using hierarchical signal representations. *IEEE Transactions on Systems, Man, and Cybernetics* 33(6):951-965.

Ünsalan, C. 2007. Measuring Land Development in Urban Regions Using Graph Theoretical and Conditional Statistical Features. *IEEE Transactions on Geoscience and Remote Sensing* 45(12):3898-3999.

USDA, Economic Research Service (2008). Agricultural Outlook: Statistical Indicators. [Online] Available WWW: http://www.ers.usda.gov/Publications/AgOutlook/AOTables/

Vapnik, V. 1998. Statistical Learning Theory. New York, NY: John Wiley & Sons, Inc.

Valkenburg, R. J., A. M. McIvor and P. W. Power. 1994. An evaluation of subpixel feature localization methods for precision measurement. In *Proc. SPIE Videometrics* III, 2350: 229-238. Boston, MA.

Weng, J., P. Cohen and M. Herniou. 1992. Camera calibration with distortion models and accuracy evaluation. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 14(10):965-980.

Weston, J., S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio and V. Vapnik. 2001. Feature selection for SVMs. *Advances in Neural Information Processing Systems* 13:668-674.

Weston, J., A. Elisseeff, B. Schölkopf and M. Tipping. 2003. Use of the Zero-Norm with Linear Models and Kernel Methods. *Journal of Machine Learning Research* 3(7-8):1439-1461.

Witten, I. H. and E. Frank. 2005. Data Mining: Practical Machine Learning Tools and Techniques (2nd Ed.). San Francisco, CA.: Morgan Kaufmann.

Zentai, G., L. Partain, R. Pavlyuchkova, C. Proano, M. Schieber, K. Shah, P. Bennett, L. Melekhov and H. Gilboa. 2006. Comparison of Mercuric Iodide and Lead Iodide X-Ray Detectors for X-Ray Imaging Applications. *IEEE Transactions on Nuclear Science* 53(5):2506-2512.

Zhang, Z. 2000. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (11):1330-1334.

Zhong, J. and S. Sclaroff. 2003. Segmenting foreground objects from a dynamic textured background via a robust Kalman filter. In *Proc. 9th IEEE International Conference on Computer Vision*, 1: 44-50. Nice, France.

Zhou, M. and H. Wei. 2006. Face Verification Using GaborWavelets and AdaBoost. In *Proc. 18th International Conference on Pattern Recognition*, 1: 404-407. Hong Kong, China.