

DETECTION OF CITRUS DISEASES USING COMPUTER VISION TECHNIQUES

By

DAE GWAN KIM

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL  
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2011

© 2011 Dae Gwan Kim

To my parents and my advisor, Dr. Burks

## ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Thomas F. Burks whom provided invaluable guidance and encouragement during the course of this research. In addition, I would like to thank the following colleagues whom provided invaluable assistance to me during the course of my research; Tony Qin, Xuhui Zhao and Duke M. Bulanon. In addition, I wish to express my gratitude to several dear friends whose encouragements kept me going through the process of this research.

Finally, I would like to dedicate this research to my parents in Korea. They always encouraged me to pursue education, I am sure they would be pleasure with the results of this research and the completion of my PhD dissertation. I want to give my parents all the glory for this work.

# TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS.....	4
LIST OF TABLES.....	9
LIST OF FIGURES.....	11
ABSTRACT.....	15
CHAPTER	
1 INTRODUCTION.....	18
Background.....	18
Computer Vision Techniques for Citrus Disease Control.....	19
Objectives.....	20
Microscopic Imaging.....	21
Hyperspectral Imaging.....	21
2 LITERATURE REVIEW.....	24
Citrus Industry Overview.....	24
Citrus Canker.....	25
Citrus Greening.....	26
Black Spot.....	28
Recommendations on Management of Citrus Diseases.....	29
The Electromagnetic Spectrum.....	31
Machine Vision Based Crop and Fruit Status Monitoring.....	32
Image Processing Approaches.....	35
Spectral Images based Methods and Neural Network Classifiers.....	37
Near Infrared Monitoring Methods.....	40
Wavelet Transform.....	42
Control of Citrus Diseases using Image Processing.....	43
3 MATERIALS AND METHODS.....	46
Citrus Canker Samples.....	46
Citrus Leaf Samples.....	47
Black Spot Samples.....	48
Image Acquisition Systems.....	48
Charge-Coupled Device Camera System.....	48
Hyperspectral Image System.....	51
Hyperspectral Image System Calibration.....	52
Pre-processing Techniques.....	53

Visible Spectrum Image Pre-processing.....	53
ROI selection and color space conversion for canker detection.....	54
Edge detection for greening detection .....	55
Hyperspectral Image Pre-processing .....	55
Color Co-occurrence Methodology .....	57
Feature Extraction .....	60
Principal Component Analysis .....	62
Singular Vector Divergence .....	64
Correlation Analysis.....	66
Wavelet Transform .....	67
Pattern Recognition .....	69
Fisher`s Linear Discriminant Method.....	71
Neural Network Based on Back Propagation Network Method .....	71
Support Vector Machine Method .....	73
Statistical Classification .....	74
Hyperspectral Image Classification Methods .....	76
Spectral Angle Mapper.....	76
Spectral Information Divergence .....	77
<b>4 DETECTION OF CITRUS CANKER DISEASE .....</b>	<b>95</b>
Materials and Methods.....	98
Image Acquisition System .....	98
Image Analysis.....	99
Texture Classification .....	102
Results and Discussion.....	103
Selection of Texture Features .....	103
Classification of Citrus Peel Conditions .....	104
Stability Test of the Classification Model .....	106
Summary .....	107
<b>5 DETECTION OF CITRUS GREENING DISEASE ON ORANGE LEAVES.....</b>	<b>112</b>
Materials and Methods.....	115
Image Acquisition System .....	115
Image Processing.....	115
Features extraction .....	118
Classification algorithms .....	120
Results and Discussion.....	121
Classifications of Citrus Disease Conditions .....	121
The Confusion Matrix for Greening Positive vs. Negative .....	123
Stability Test of the Greening Classification Model .....	124
Summary .....	125
<b>6 DETECTION OF THE DISEASE USING PATTERN RECOGNITION METHODS</b>	<b>134</b>
Materials and Methods.....	136

	Image Acquisition System .....	136
	Image Analysis .....	137
	Classification Using Pattern Recognition Methods .....	140
	Classification Results.....	140
	Canker Disease Classification Based on Pattern Recognition Algorithms. ....	140
	Citrus Greening Disease Classification Based on Pattern Recognition Algorithms. ....	142
	Summary .....	143
7	CLASSIFICATION METHODS OF CITRUS GREENING BY HYPERSPECTRAL IMAGING .....	150
	Materials and Methods.....	153
	Citrus Leaf Samples .....	153
	Hyperspectral Image Acquisition .....	154
	Hyperspectral Image Analysis .....	155
	Hyperspectral image transformation using PCA .....	155
	Citrus greening classification algorithm using PCA.....	156
	Classification algorithm using wavelet transform .....	157
	Results and Discussion.....	158
	The Plot of Relative Reflectance Spectra.....	158
	Classification of Citrus Greening using PCA .....	158
	Classification of Citrus Greening using Wavelet Transform.....	160
	Summary .....	161
8	CITRUS BLACK SPOT DETECTION USING HYPERSPECTRAL IMAGING .....	170
	Methodology .....	172
	Samples .....	172
	Hyperspectral Image Acquisition .....	172
	Hyperspectral Image Processing.....	174
	Results and Discussion.....	175
	Spectral Characteristics of Black Spot and Other Conditions.....	175
	SID and SAM-based on Classifications .....	176
	The SID and SAM Threshold Values for Classification Accuracy.....	176
	Summary .....	177
9	HYPERSPECTRAL BAND SELECTION FOR DETECTION OF BLACK SPOT USING MULTISPECTRAL ALGORITHMS .....	189
	Materials and Methods.....	191
	Hyperspectral Image Acquisition .....	191
	Correlation Analysis Algorithms for Band Selection .....	192
	Principal Component Analysis Algorithms for Band Selection.....	193
	Singular Vector Decomposition Algorithms for Band Selection .....	194
	Result and Discussion .....	194
	Mean Relative Reflectance Spectra of Black Spot .....	194

Hyperspectral Band Selection using CA.....	195
Hyperspectral Band Selection using PCA .....	195
Hyperspectral Band Selection using SVD .....	196
The Hyperspectral Images Selected by Band Selection Methods .....	196
Classification of Black Spot using Band Ratio Images .....	197
Summary .....	200
10 CONCLUSION AND FUTURE WORK.....	221
Conclusions .....	221
Future Work .....	222
APPENDIX	
A  MATLAB CODE FILES FOR EDGE DETECTION.....	224
B  MATLAB CODE FILES FOR PATTERN RECOGNITION METHODS.....	226
C  MATLAB CODE FILES FOR PATTERN HYPERSPECTRAL IMAGING .....	234
D  MATLAB CODE FILES FOR PRINCIPAL COMPONENT ANALYSIS .....	235
E  MATLAB CODE FILES FOR WAVELELT TRANSFROM.....	236
LIST OF REFERENCES .....	241
BIOGRAPHICAL SKETCH.....	248

## LIST OF TABLES

<u>Table</u>	<u>page</u>
3-1 The number of black spot and other symptoms.....	94
4-1 Texture features selected by stepwise discriminant analysis .....	109
4-2 Classification results using model HSI_13 in Table 4-1 .....	110
4-3 Classification results in percent correct for all models in Table 4-1. ....	110
4-4 Classification results for shuffle data models in percent correct .....	111
5-1 Texture feature models selected by stepwise discriminant analysis for fall season.....	128
5-2 Texture feature models to all conditions except young flush for fall season .....	129
5-3 Classification summary in percent correct for all models in Table 5-1 .....	129
5-4 Classification summary in percent correct for all models in Table 5-2 .....	130
5-5 Classification result in percent correct for HSI_18 model in Table 5-1 .....	130
5-6 Classification result in percent correct for HSI_15 model in Table 5-1 .....	131
5-7 Classification result in percent correct for HSI_14 model in Table 5-1 .....	131
5-8 Confusion matrix in percent correct for HSI_18 model in Table 5-1 .....	132
5-9 Confusion matrix in percent correct for HSI_15 model in Table 5-1 .....	132
5-10 Confusion matrix in percent correct for HSI_14 model in Table 5-1 .....	132
5-11 Classification results for shuffle data about HSI_15 model in percent correct..	133
6-1 Texture feature models selected by SAS stepwise analysis for citrus canker ..	147
6-2 Canker disease classification results in percent correct for all models using FDA. ....	147
6-3 Canker disease classification results in percent correct for all models using BP neural network .....	147
6-4 Canker disease classification results in percent correct for all models using SVMs.....	148

6-5	Texture feature models selected by SAS stepwise analysis for citrus greening .....	148
6-6	Citrus greening classification results in percent correct for all models using FDA .....	148
6-7	Citrus greening classification results in percent correct for all models using BP.....	149
6-8	Citrus greening classification results in percent correct for all models using SVM.....	149
7-1	The number of citrus greening and other conditions.....	168
7-2	Classification result in percent correct using PCA based on method.....	169
7-3	Classification result in percent correct using WT based on method .....	169
8-1	The number of black spot and other conditions .....	186
8-2	The classification summary using 6 threshold values for differentiating black spot from other conditions using SID mapping of hyperspectral images .....	186
8-3	The classification summary using 6 threshold values for differentiating black spot from other conditions using SAM mapping of hyperspectral images .....	187
8-4	The best overall classification accuracy using the SID threshold value of 0.04	187
8-5	The best overall classification accuracy using the SAM threshold value of 0.09 .....	188
9-1	The classification summary for differentiating black spot from other diseases and normal conditions using PCA.....	218
9-2	The classification summary for differentiating black spot from other diseases and normal conditions using SVD.....	219
9-3	The classification summary for differentiating black spot from other diseases and normal conditions using CA.....	220

## LIST OF FIGURES

<u>Figure</u>		<u>page</u>
1-1	Overview of citrus disease detection system .....	23
2-1	The electromagnetic spectrum comprising the visible and non-visible range.....	45
3-1	Images of abnormal peel conditions .....	79
3-2	Images of nutritional deficiency .....	79
3-3	Images of citrus greening .....	80
3-4	Images of normal conditions.....	80
3-5	15 images of blotch mottle conditions.....	81
3-6	Representative images for each peel condition .....	82
3-7	Three faces of each fruit with 120° interval.....	82
3-8	Typical image system for acquiring RGB images from citrus samples .....	83
3-9	Digital microscope system for acquiring RGB images from citrus samples .....	83
3-10	Typical hyperspectral line-scan imaging system.....	84
3-11	Principle of the prism-grating-prism imaging spectrograph for acquiring spartial and spectral information from an object .....	84
3-12	Conceptual representation of a volume of hyperspectral image data .....	85
3-13	White paper printed with thin parallel lines 2 mm apart for geometric calibration .....	85
3-14	Spectral profiles of calibration lamps: xenon lamp and mercury-argon lamp.....	86
3-15	The relationship between the vertical band position and wavelength from two calibration lamps.....	86
3-16	ROI selection program developed in MATLAB R2007b.....	87
3-17	Typical ROI images for normal and diseased citrus peel conditions.....	87
3-19	Converted images of a leaf samples .....	88
3-20	Edge detected image of a leaf samples.....	88

3-21	Hyperspectral image pre-processing .....	89
3-22	The plot of reflectance factor for white panel .....	90
3-23	Nearest neighbor diagram .....	90
3-25	Feature vector, feature space and scatter plot .....	91
3-26	Good feature and bad feature.....	92
3-27	Distribution plots with Pattern types.....	92
3-29	Illustration of the calculation .....	93
3-30	Possible linear classifiers for separating the data.....	94
3-31	The angle between endmember spectra and target spectra as vectors in a space.....	94
4-1	Color image system for acquiring RGB images from citrus samples .....	109
4-2	Procedures for color image analysis.....	109
5-1	Digital microscope system for acquiring RGB images from citrus leaf samples	127
5-2	Gray-level dependence example: (a) SGDM for different orientations, (b) gray-level image .....	127
5-3	Procedures for color image analysis.....	127
5-4	Procedures for leaf edge detection.....	128
6-1	Color image system for acquiring RGB images from citrus samples .....	145
6-2	Digital microscope system for acquiring RGB images from citrus disease samples .....	145
6-3	Procedures for color image analysis for citrus canker .....	146
6-4	Procedures for color image analysis for citrus greening .....	146
7-1	Hyperspectral line-scan imaging system .....	163
7-2	A detailed illustration of Threshold based on classification algorithms .....	163
7-3	Comparison of the threshold value distribution between citrus greening and other conditions .....	164
7-4	The image-opening morphological filtering.....	164

7-5	A detailed illustration of the feature extraction using Wavelet Transform .....	165
7-6	Four-level decomposed images in each condition.....	165
7-7	Reflectance mean spectra and corresponding standard deviation of each ROI from citrus greening and other symptoms .....	166
7-8	Reflectance of spectra of citrus leaf samples with citrus greening and other conditions .....	167
7-9	Five score images obtained from principal component analysis using full spectral regions .....	168
8-1	Hyperspectral line-scan imaging system .....	180
8-2	A detailed illustration of SAM and SID classification algorithms .....	180
8-3	The mean and standard deviation from endmember spectra .....	181
8-4	Representative rule images from SID and SAM mapping the hyperspectral images.....	181
8-5	Reflectance mean spectra and corresponding standard deviation of each ROI from black spot, greasy spot, melanose, wind scar, and market .....	182
8-6	The reflectance spectra of the samples with black spot, normal and different symptoms over the wavelength range between 483 and 959 nm.....	183
8-7	Detailed comparison of the value distribution between black spot and non black spot region .....	184
8-9	Comparison with SID and SAM classification accuracy.....	185
9-1	Hyperspectral line-scan imaging system .....	203
9-2	Data analysis and classification algorithms for CA .....	203
9-3	First five score images obtained from principal component analysis for hyperspectral images of grapefruit samples .....	204
9-4	Data analysis and classification algorithms for PCA.....	205
9-5	Data analysis and classification algorithms for SVD.....	205
9-6	Representative ROI images for each black spot peel conditions.....	206
9-7	The reflectance spectra of the samples with black spot, normal and different symptoms over the wavelength range between 483 and 959 nm.....	206

9-8	The contour plot the correlation coefficients ( $r$ ) between two band ratio and fruit peel conditions.....	207
9-9	Mean weighting coefficients for the third principal component from 20 black spot ROIs .....	207
9-10	Singular value estimates of black spot conditions .....	208
9-11	The representative single band reflectance images with five different conditions at the wavelengths with (a) highest correlation coefficients and (b) second highest correlation coefficients by CA .....	209
9-12	Representative spectral reflectance images of diseased peel and normal conditions at four wavelength selected by PCA.....	210
9-13	Representative spectral reflectance images of diseased peel and normal conditions at four wavelength selected by SVD.....	211
9-14	The ratio images of black spot using wavelengths selected by PCA.....	212
9-15	The ratio images of black spot using wavelengths selected by SVD.....	212
9-16	The ratio images of black spot using wavelengths selected by CA .....	213
9-17	The histogram for ratio images of 802 nm/724 nm from PCA.....	213
9-18	The histogram for ratio images of 919 nm/572 nm from SVD.....	214
9-19	The histogram for ratio images of 854 nm/598 nm from CA .....	214
9-20	Identification of black spot lesions on the fruit peel based on two band ratio images using wavelengths selected by PCA (R802/R724).....	215
9-21	Classification accuracies obtained from the ratio image (802 nm/724 nm) selected by PCA.....	216
9-22	Classification accuracies obtained from the ratio image (919 nm/572 nm) selected by SVD.....	216
9-23	Classification accuracies obtained from the ratio image (854 nm/598 nm) selected by CA .....	217

Abstract of Dissertation Presented to the Graduate School  
of the University of Florida in Partial Fulfillment of the  
Requirements for the Degree of Doctor of Philosophy

## DETECTION OF CITRUS DISEASES USING COMPUTER VISION TECHNIQUES

By

Dae Gwan Kim

August 2011

Chair: Thomas F. Burks

Major: Agricultural and Biological Engineering

The citrus industry has led the agricultural economy of the state of Florida to prosperity. Florida has historically been the largest citrus producing state in the USA. Citrus fruits such as oranges, grapefruit, tangelos, navels, limes, tangerines and other specialty fruits are the chief crops of the state. The remarkable growth of the state economy has been partially based on the various citrus related industries. This situation has brought job opportunity for many people and important potential for the economic growth of the state. To maintain the prosperity of citrus industry, Florida has been concerned about disease control, labor cost, and global market.

During the recent past, citrus canker and citrus greening have become serious threats to citrus in Florida. These diseases can result in tree decline, death, yield loss and lost marketability. Likewise, the farmers are concerned about costs from tree loss, scouting, and chemicals used in an attempt to control the disease. An automated detection system may help in citrus greening and citrus canker prevention and, thus reduce the serious loss to the Florida citrus industry.

This thesis considers the development of disease detection approaches for these diseases (citrus greening and citrus black spot), when found in the presence of various

other citrus diseases. The detection approach consists of three major sub-systems, namely, image acquisition, image processing and pattern recognition. The image acquisition system consists of digital camera (standard lens on a microscope), frame grabber with high pixels embedded in the computer and supplemental lighting. The imaging processing sub-system includes image preprocessing for background noise removal, leaf boundary detection and image feature extraction. Pattern recognition approaches were used to classify samples among several different conditions on fruit for canker, black spot, and leaves for greening. The canker-based detection studies were conducted on grapefruit samples collected during the 2006-2007 harvest season. Fruit samples exhibiting canker and other topical peel condition were collected and RGB images were taken of the various conditions. Color texture features were extracted and discriminant analysis was used to classify grapefruit according to peel condition achieving overall accuracies of 96.5%.

The citrus greening-based detection studies were conducted on orange leaf samples collected during the spring of 2008. Leaf samples taken from the tree in an orange orchard exhibiting greening and other common leaf condition were collected and microscopic RGB images were taken. Color texture features were extracted at various levels of magnification and preliminary results demonstrated that 5× magnification was optimal.

The black spot-based detection studies were conducted on fruit samples collected during the spring of 2010. Hyperspectral images of fruit samples infected with black spots were acquired. In addition to black spot infections, fruit samples with other peel conditions such as greasy spot, wind scar, melanose, and market quality conditions

were also collected. For the hyperspectral image processing and analysis, unsupervised and supervised classification approaches were used to classify the fruit samples as black spot or non-black spot classes.

In order to evaluate the classification approaches, results were compared between classification methods for the citrus greening. Results demonstrated classification accuracy for citrus greening as high as 97.00%. Thus, this research mainly focused on demonstrating feasibility of citrus disease detection from visible symptoms on the fruit or leaf. These were offline approaches, not directly applicable to real-time technologies such as robotics and automated vehicles for agricultural application. However, they are a first step in demonstrating the potential for automated detection.

## CHAPTER 1 INTRODUCTION

### **Background**

Crops become stressed when any infections, such as viral infections, or physiological factors, such as air pollution, adversely affect growth, development, and yield. These stresses are expressed in various ways. For example, plant water stress can slow down photosynthesis, reduce evapotranspiration, and raise leaf surface temperature [4]. Other symptoms include morphological changes such as leaf curling, changes in leaf angle, wilting, chlorosis, discoloration of leaves and fruits and premature drop of fruits. Because of these observable symptoms, humans could easily assess the conditions of their crops. Traditionally, disease inspection is performed by trained human inspectors. However, human inspectors are highly variable and decisions are not always consistent between inspectors or from day to day. Moreover, commercial citrus is currently produced in large acreages with large distances between blocks. This needs significant demands for labor to manage diseases and pests. Consequently, labor saving tools are necessary to effectively scout for and manage diseases and pests. In recent years, the number of large scale farms has increased, while the labor force has shrunk due to laborers moving to other occupations. These conditions have stimulated development of automated scouts which can operate effectively for long periods of time. These systems can be more cost effective and accurate human scouts. In addition, automated disease detection can help farmers reduce cost.

Continual inflow of non-native disease also makes it impossible to use human inspectors. Every year, various crops, including citrus come into U.S. commerce from foreign countries. Many of these countries are home to insects and diseases that do not

occur in the United States. Non-native species can wreak havoc on local environmental and agricultural resources. Citrus diseases have entered into Florida from these international sources such as the import of Southeast Asia, Africa, South America, and in Australia. Scouting for infected plants and insects by human inspectors in the farm has limited usefulness. Moreover, the citrus diseases control is a repetitive manual task which is also very subjective in personal perception. In this type of Florida's environment, machine vision systems are ideally suited for routine inspection and quality assurance tasks.

### **Computer Vision Techniques for Citrus Disease Control**

Various technologies can be employed to provide intelligent systems, such as robotics and autonomous reasoning, and machine learning that have the potential to improve profitability in agricultural and food industries. Computers have been applied in the area of sorting, grading of fresh products, detection of defects such as cracks, dark spots and bruises on fresh fruits and seeds. Most symptoms of disease and defects on citrus fruit appear on the leaves of trees, and/or the peel of citrus fruit. The eye can detect some of these symptoms, but computer vision system has more accuracy and speed than the human eye. In outdoor scenes, illumination and color can change. This can confuse detection by human eye, and computer vision. Some image processing methods can eliminate these problems and find best detection results. Computer vision systems can be employed on autonomous robots, to provide un-manned disease scouting systems.

The computer vision and image processing techniques for disease detection used in this study have similar approaches. Figure 1-1 shows an overview of citrus disease detection system. Image data is acquired by various image acquisition equipments,

such as microscopic system, and hyper-spectral system, etc. systems can be used in a laboratory, and in some cases extended to the field. After collecting image data, proper image processing methods are applied. Original image data contains some unwanted factors such as noise; overexposure (underexposure) to the light, change of color etc. Image processing techniques can improve this and help increase classification accuracy. For classifying microscopic image data, image features are extracted by feature extraction methods like color co-occurrence method (CCM), wavelets transform, run length matrix, fractal dimension etc. For hyperspectral and multispectral algorithms, principal component analysis (PCA), singular vector decomposition (SVD), and correlation analysis (CA) are used to extract important band selection. Classification methods also have various techniques such as neural network, Mahalanobis minimum distance, self organizing map (SOM), support vector machine (SVM), etc. Each feature extraction and classification method has good points. It is important that proper extraction and classification methods be used because these methods affect the classification accuracy.

This dissertation will use these procedures for general disease detection using computer vision technologies. In the following chapters, these methods and steps are discussed in more detail.

### **Objectives**

The overall objectives of the research described in this dissertation were to develop a computer vision based method for detecting various diseases using color texture feature and spectral image analysis algorithms under a controlled lighting.

## **Microscopic Imaging**

The microscopic imaging objective was to investigate the potential of using color texture features in the visible region for detecting citrus canker and greening symptoms.

Specific goals were to:

- Compare the classification accuracies using various pattern recognition techniques and select optimal classifier.
- Use a digital color microscope system to collect RGB images from orange leaves with seven conditions (i.e., young flush, normal mature, blotchy mottle, greening islands, zinc deficiency, iron deficiency, and manganese deficiency)
- Use an image acquisition system to collect RGB images oranges with normal and five peel conditions (i.e., canker, copper burn, greasy spot, melanose, and wind scar).
- Determine image texture features based on the color co-occurrence method (CCM).
- Create a set of reduced feature data models through a stepwise elimination process and classify different citrus disease conditions.
- Compare the classification accuracies using various pattern recognition techniques and select optimal classifier.

## **Hyperspectral Imaging**

The hyperspectral imaging objective was to investigate the potential of using hyperspectral imaging in the visible and near-infrared region for detecting citrus

greening and black spot. Specific goals were to:

- Set up, calibrate and characterize a hyperspectral imaging system for acquiring spatially resolved, steady-state reflectance from the materials over the visible and short-wave near-infrared regions (500-1000 nm);
- Use a hyperspectral imaging system to collect the reflectance in the spectral region(500-1000 nm) from grapefruit leaves of two citrus greening diseases, three nutrition deficiency and two normal leaves (i.e., greening islands, blotch mottle, Iron Deficiency, Manganese Deficiency, Zinc Deficiency, normal mature and young flush),and grapefruits with normal and four diseased skin conditions (i.e., black spot, melanose, greasy spot, and wind scar);

- Use principal component analysis (PCA) and wavelet transform (WT) based algorithms for hyperspectral image processing and classification to identify citrus greening diseases, normal and nutrient deficiencies symptoms;
- Use the spectral information divergence (SID) and spectral angle mapper (SAM) based algorithms to differentiate citrus spot black from normal and other common diseased peel conditions.

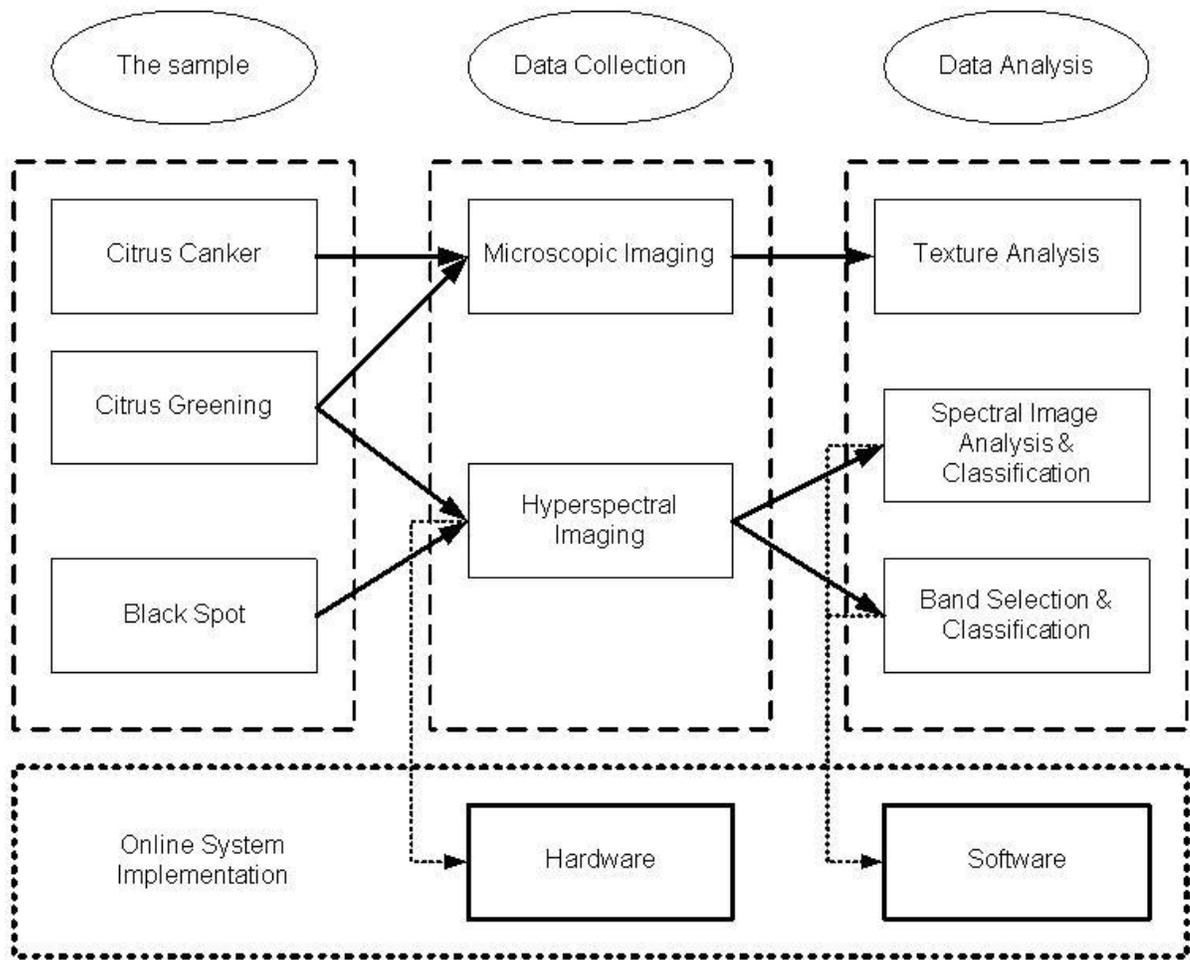


Figure 1-1. Overview of citrus disease detection system

## CHAPTER 2 LITERATURE REVIEW

The citrus market is one of the most important industries in the Florida economy. Citrus growers desire to exploit automation technology to help manage citrus groves and make the crop more economical. Automation systems offer farmers means to minimize agricultural chemicals inputs, and increase overall crop yield and profits. Automated agricultural systems plays an important part in managing the cost of production, since citrus production is an "input-intensive" crop and the Florida agricultural industry has a volatile trend of cost per unit area (Sevier and Lee, 2003). These agricultural systems require more knowledge of in-field variability and its relationship to crop yield, as well as, the influence of disease and pest on crop damage and ultimately profitability. In the past decade, various researchers have used image processing and pattern recognition techniques in agricultural applications, such as detection of weeds in a field, sorting of fruits and vegetables, and pest and disease detection, etc.

This chapter will introduce previous research in agricultural based machine vision, image processing, and classification methods for agricultural applications, and also introduce background information for citrus diseases.

### **Citrus Industry Overview**

Florida leads the U.S. citrus production and accounts for a major part of the U.S. citrus industry. According to the Citrus production report (USDA, 2005-06), Florida accounts for more than 70% of U.S. citrus production.

Florida produces a number of different citrus products. Orange, grapefruit, and lemons are major citrus crops, with lesser production in tangerines, limes, and an

increasing variety of specialties. Oranges account for more than 80% of yield (Hodges et al., 2001).

Orange and grapefruit account for about 90% of U.S. production (Jacobs et al., 1994). Florida, California/Arizona, and Texas are the major states for U.S. citrus productions. Florida accounted for more than three quarters of all U.S. citrus, while California/Arizona and Texas produced the remainder. Citrus is consumed not only as fresh fruit but also in products that use citrus fruits. Florida has 52 citrus processing plants (Hodges et al., 2001). They produce various orange juice items such as chilled, canned, and concentrated juice. Processing also generates a number of byproducts like food additives, cattle feed, and cosmetics. These plants also produced many byproducts like citrus pulp, molasses, and the essential oil-limonene (Hodges et al., 2001).

### **Citrus Canker**

Polek (2006) described the origin of citrus canker and characteristics important for canker identification. Citrus canker occurs in various countries around the world, Asia, United States, Indian Ocean and Pacific islands, the Middle East, and Australia. Citrus canker occurs not only in countries with warm and tropical climates but also in dry climates. The major cause of citrus canker is pathotypes or variants of bacterium, *Xanthomonas axonopodis* (formerly *campestris*) pv.citru(Xac). Many citrus growing counties have regulated this bacterium and citrus products infected with citrus canker.

Gottwald et al. (2002) presented a history and overview of citrus canker. Florida has a long history of fighting citrus canker. Citrus canker first appeared around 1912 in Florida with outbreaks in the Florida peninsula. After about 80 years, citrus canker spread over the state, and Florida Department of Agriculture and Consumer Services

(FDACS), the Division of Plant Industry (DPI), and the USDA Animal and Plant Health Inspection Service (APHIS) established The citrus canker detection program , a cooperative state/federal citrus canker eradication program (CCEP).

After becoming infected with citrus canker, leaf and fruits will show several symptoms. The surface of the leaf may have spots and the peel become blemished with lesions. When the conditions are fully developed, the fruit are dropped and the tree dies. Leaf lesions appear about 7 to 10 days after infection. Lesions on fruit and twigs generally appear like cork and sometimes lesions appear like a blister or eruption on leaf and fruits.

Once fruit becomes symptomatic, the marketability to many parts of the world becomes limited. This is most problematic for fresh market citrus, while citrus for juice is not affected. The disease primarily spread during spring and summer rains combined with wind speeds in excess of 18 mph.

### **Citrus Greening**

In 2005, citrus greening, Huanglongbing (HLB), was first discovered in Florida groves. This disease is a greater threat to Florida groves than citrus canker. Florida citrus growers are desperately fighting this plant disease which has the potential to destroy the state's \$9 billion commercial citrus industry (Defending Florida's Citrus Industry from an Emerging Disease, The American Phytopathological Society, 2008). Citrus greening has spread from eight to 23 counties since it was first found in Florida just a little more than three years ago. Once citrus trees are infected, the fruit yield, and quality are greatly reduced. The trees also become susceptible to other diseases and health problems.

HLB threatens the Florida citrus industry. The diseases effects are:

- Destroys production, appearance and value of citrus trees.
- Produces bitter, inedible, misshapen fruit.
- Is fatal to citrus trees.

While canker blemishes the fruit peel affecting marketability, citrus greening can kill the whole grove in a matter of a few years. Crops in Asia and Africa have been severely impacted by HLB and have caused a loss in citrus production in Brazil. In some areas of Brazil, citrus greening has affected as much as 70% of the fruit yield (ScienceDaily, 2007). Citrus greening reduces citrus production period from fifty years to fifteen years, and to date no cure for citrus greening has been found.

Citrus greening was one of the most serious diseases that affect citrus (Chung and Brlansky, 2006). A number of countries in Asia have been affected by this disease in the past, while more recently Brazil (2004) and Florida (2005) were infected. Early symptoms of HLB are yellowing leaves that may appear on only a single branch. A condition termed, blotchy mottle, appears on the leaf surface. Infected trees become non-productive in 2 to 3 years. After early stages of HLB, the leaves became small and exhibit nutritional deficiency symptoms frequently. Also, the value of citrus fruit is reduced as they become small, sparse and have abnormal appearance and color. Moreover, the affected citrus often contains aborted seeds and juice quality is low.

Nutritional deficiencies of citrus each have distinctive feature on the leaf surface. Leaves with iron deficiency have the dark green veins within the yellow leaf blade. The Leaves with manganese deficiency have symmetrical yellowing across the mid-vein and a dark triangle at the leaf base. The leaves with zinc deficiency have fairly symmetrical yellowing across the mid-vein (Polek, 2007). Leaves with citrus greening symptoms are

similar to other conditions and diseases, but a blotchy mottle pattern with asymmetrical light greening and dark green patches, is the most distinctive symptom. Green islands on the leaf have asymmetrical patterns on the opposite sides of the mid-vein.

### **Black Spot**

Citrus black spot (CBS) is the latest exotic disease introduced in Florida. This fungal disease has potential as a greater threat to Florida groves similar to citrus canker and greening because it is one of the most well-known fungal diseases worldwide. Once citrus trees are infected with citrus black spots, the fruit yield, and quality are greatly reduced. Besides, fruits infected with citrus black spots are not acceptable for fresh market (Chung, 2009). Therefore it is important to control this disease to achieve profitable production.

Citrus black spot is a kind of fungal disease caused by *Guignardia citricarpa* (sexual stage) and *Phyllosticta citricarpa* (asexual stage). Citrus black spot can be identified by fruit symptoms which cause cosmetic lesions on the fruit peel that are the most conspicuous symptom of infection (Dewdney, 2010). Fruit symptoms can be quite variable. Black spot lesions begin as small orange or red spots with black margins and enlarge to become necrotic lesions. Other symptoms of black spot on fruit include hard spot lesions, virulent spot, cracked spot, and false melanose (The Institute of Food and Agricultural Sciences Extension at University of Florida, 2010). Detecting fruits infected with black spots can help in controlling the spread of this disease specifically in areas that are black spot free. The design and implementation of technologies that can efficiently detect black spot symptoms will greatly aid in the control effort.

## **Recommendations on Management of Citrus Diseases**

Citrus disease management has been introduced by several researchers. Morris and Muraro (2008) reported controlling the spread of greening in Brazil. Recent experience in Brazil has shown that when greening is detected early enough and control practices are following diligently, groves can remain productive with disease incidence at low levels. Early detection is the key for controlling the spread of greening, followed by effective control practices. In Brazil, after three years of intensive spraying for psyllids, scouting for the disease, and aggressive removal of infected trees, the best managed groves have below 1% infection. Other diseases are causing more tree losses than greening in these groves. Moreover, control of inoculum by removing infected trees is the most critical component of a successful greening control program. Thus, the sooner infected trees are removed, the better the chances of reducing future infections. An experienced scouting crew can identify trees accurately, enabling the removal of infected trees without the turn-around time it takes to send infected leaves to a lab for confirmation. Thus, the number of trees with suspect samples sent to a lab for verification from these programs is small percentage of the trees removed. In Brazil, the objective is to remove infected tree the day they are identified.

Yates et al (2008) suggested early scouting for citrus greening management. Scouting for citrus greening has become a necessary practice for citrus production since the first positive confirmation of greening in Florida in August 2005. They emphasized the need of scouting frequency for citrus greening management. Citrus groves should be inspected for greening every two to three months. If greening has previously been found in a grove or has been confirmed nearby, scouting more than four times a year is strongly recommended. Greening symptoms are most visible during

the fall and winter months, but can be observed year round. During the spring flush, scouting becomes more difficult because new leaves typically do not express greening symptoms, and the older, symptomatic leaves become hidden by the new flush.

They introduced various methods such as All-terrain vehicles (ATVs), elevated platform and walking through groves to scout for greening. Multiple types of elevated platforms are available, including tractor- or truck-mounted platforms. They suggested other managements such as removal of infected trees, integrated pest management, and use of disease-free nursery trees.

Brlansky et al. (2010) recommended the scouting for greening-infected trees in Florida. The scouting for greening should be done routinely so that infected trees can be removed. It is recommended that scouting be conducted four or more times per year. The frequency of scouting may be higher in areas previously determined to have HLB-positive trees. Symptoms are the easiest to find from October to March. However, symptoms may be present at other times of the year. The current methods used to scout are walking, all-terrain vehicles and on vehicle-mounted platforms. Other management practices include removal of infected trees. For removal of infected trees, symptomatic tree numbers and the rows in which they are found should be marked with colored flagging tape and GPS coordinates taken or the sites marked on a map to facilitate relocation. In some cases, an HLB PCR diagnostic test may be necessary to confirm the disease. Removal of infected trees is the only way to ensure that they will not serve as a source of the bacteria for psyllid acquisition and subsequent transmission. Prior to removal, the infected tree should be treated with a foliar insecticide (such as Danitol, fenpropathrin) to kill all adult psyllids feeding on that tree.

Failure to control these psyllids will result in the infected psyllids dispersing to new plants once the diseased tree is removed.

To control citrus black spot, fungicides are required to control citrus black spot in countries where it is endemic (Chung et al., 2009). Protective treatments using copper or strobilurin fungicides or mancozeb must be properly timed, and up to 5 sprays may be required during the period of susceptibility. Removal of dead leaves in groves is an effective practice and reduces inoculum potential. Long distance spread of citrus black spot occurs via infected nursery stock, and steps to avoid movement of infected trees help limit spread of the disease to new areas. Little effort has been made toward developing varieties with tolerance or resistance to citrus black spot.

### **The Electromagnetic Spectrum**

Sun (2008) summarized the electromagnetic spectrum. The electromagnetic spectrum is useful in image formation. The images are derived from electromagnetic radiation in both visible and non-visible ranges. Radiation energy travels in space at the speed of light in the form of sinusoidal waves with known wavelengths. Figure 2-1 shows the electromagnetic spectrum of all electromagnetic waves.

The electromagnetic spectrum arranged from shorter to longer wavelengths provides information on the frequency as well as the energy distribution of the electromagnetic radiation. The gamma rays with wavelengths of less than 0.1 nm constitute the shortest wavelength of the electromagnetic spectrum. Traditionally, gamma radiation is important for medical and astronomical imaging, leading to the development of various types of anatomical imaging modalities such as CT, MRI, SPECT and PET.

Ultraviolet (UV) light is of shorter wavelength than visible light. Similar to IR, the UV part of the spectrum can be divided, this time into three regions: near ultraviolet (NUV) (300 nm), far ultraviolet (FUV) (30 nm), and extreme ultraviolet (EUV) (3 nm). NUV is closest to the visible band while EUV is closest to the X-ray region and therefore is the most energetic of the three types. FUV, meanwhile, lies between the near and extreme ultraviolet regions, and is the least explored of the three. There are many types of CCD camera that provide sensitivity at the near-UV wavelength range. The sensitivity of such a camera usually peaks at around 369 nm while offering coverage down to 300 nm.

Located in the middle of the electromagnetic spectrum is the visible range, consisting of narrow portion of the spectrum with wavelength ranging from 400 nm (blue) to 700 nm (red). The popular charge-coupled device (CCD) camera system works in the visible range. Infrared (IR) light lies between the visible and microwave portions of the electromagnetic band. As with visible light, infrared has wavelengths that range from near (shorter) infrared to far (longer) infrared. The latter belongs to the thermally sensitive region, which makes it useful in imaging device is the indium gallium arsenide (InGaAs)-based near-infrared (NIR) camera, which gives the optimum response in the 900-1700 nm band. At the other end of the spectrum, the longest waves are radio waves, which have wavelengths of many kilometers. The well-known ground-probing radar (GPR) and other wave-based imaging modalities operate in this frequency range.

### **Machine Vision Based Crop and Fruit Status Monitoring**

Schertz and Brown (1968) proposed fruit identification using machine vision systems nearly forty years ago. Using light reflectivity differences between leaves and

fruits in the visible or infrared portion of the electromagnetic spectrum, the information about location of fruits might be determined.

Powell et al. (2005) developed machine vision for an autonomous vehicle. The development is related in various test situation such as straight, curved and sine wave lines. The autonomous car was tested in a controlled environment. The result about straight and sweeping curve paths are satisfied, and nonlinear test such as sine wave lines also had good result. However, the sine wave path was less accurate. The overall result is satisfied and the system could identify test situation.

Ng et al. (1989) measured corn (*Zea mays* L.) kernel damage using machine vision algorithms. Damage was divided into two classes, mechanical damage and mold damage. Each damage had a different analysis method. For mechanical damage, both single-kernel and batch analysis was used for evaluating damage area using green dye. In mold damage; only single-kernel analysis was used. The machine vision system had high classification (mechanical damage: 99.5% and mold damage: 98.7%) accuracy for both damages. This research demonstrated that batch analysis method was faster than single-kernel analysis.

Takashi et al. (2003) developed a machine vision system for crop status. Crop status is a vital part for farm management and crop cultivation. The system used several specific functions, like growth curve (the Gompertz curve) and exponential function for estimating vegetation cover area and the actual plant status. The result was not acceptable for every plant status, but the research demonstrated that image processing could be very useful to build machine vision systems for crop status detection.

Jimenez et al. (2000) surveyed several computer vision approaches for locating fruit in trees. Their work reported that most detection vision systems had three sensor types: spectral, intensity and distance, and two analysis methods: local and shape. Vision systems can use different combinations of possible types of sensing and analysis approaches. Each combination has their own advantages. The method using local analysis based on spectral images obtained the best rates for detection (100%), but required known conditions for background color. Another method using shape analysis based on intensity or spectral images was found to be good for detecting fruits regardless of its color, however there were frequent false detections. Another method using range image and shape-based approaches had good results (above 80%). In addition, this algorithm did not generate false detections.

Regunathan et al. (2005) identified fruit count and size using machine vision and an ultrasonic sensor. The ultrasonic sensor gave distance information, while color images were obtained for counting fruit. An image processing method converted red-green-blue (RGB) data to hue-saturation-intensity (HLS) data format. The detection system combined two sensors to estimate the actual size of fruits. Three different classification methods were used: Bayesian and neural network and Fisher linear discriminant giving good results for fruit size and counts, but root mean square error the neural network method was better than others for fruit size, while the error for neural network was only 2.6% for fruit count and 97.4% for size estimation (0.4 cm).

Ashish et al. (2007) detected citrus greening using a spectral method for the detection of greening. To distinguish greening and healthy trees, discriminability, spectral derivative analysis and spectral ratio analysis were used. The discriminability of

wavelengths had a wide range. In visible wavelengths of 695 to 705 nm, the discriminability had the best result of 0.83 to 0.86. In spectral derivative analysis, the wavelengths of 480 nm, 590 nm, 754 nm, 1041 nm, and 2071 nm was suitable to differentiate greening. The spectral ratio analysis was better for understanding of the spectral properties of greening-infected in citrus canopy. The reflectance at 530-564 nm was higher in the spectral ratio.

### **Image Processing Approaches**

Dave et al. (1995) compared three image processing methods for identifying species of plants. Three methods were investigated, Fourier transform, fractal dimension, and co-occurrence matrix methods. A fast Fourier transform method converted spatial domain image data to frequency domain data. Using frequency domain data, the power spectrum inside the annular band of radius  $R_1$  to  $R_2$  was calculated. The fractal dimension method used fractal geometry information from the input image. After converting RGB to gray image, geometrical properties, such as shape, line and surface area were calculated. The co-occurrence matrix method created spatial-dependence matrices calculated from the gray scale image. All texture features were obtained from these matrices. The co-occurrence matrix method gave the best accuracy, and the execution time also was less than others.

Du et al (2006) described five different texture feature methods to investigate tenderness of cooled pork ham. In this work, First-order gray-label statistics (FGLS), Run length matrix (RLM), Gray-level co-occurrence matrix (GLCM), fractal dimension (FD), and wavelet transform (WT) were investigated. After extracted texture feature, statistical analysis was studied using the CORR procedure in SAS package. Moreover, for further analysis, a partial least squares regression (PLSR) technique was applied.

The WT method had multi-scale representation. These characteristics presented the best results about the tenderness of cooked pork ham than the traditional methods (FGLS, GLCM, and RLM).

Burks et al. (2000) used the color co-occurrence method and neural network for classifying weed species. Thirty-three texture features were extracted using the co-occurrence matrix method (CCM) in input image data. For analyzing texture features, SAS discriminant procedure STEPDISC was used. The back-propagation neural network method gave good accuracy rates (96.7%) as well as high individual species classification accuracies (90.0%).

Pydipati et al. (2006) identified citrus disease using the co-occurrence matrix method, (CCM) texture feature method, and a discriminant analysis. Before extracting image features, the input image was converted from RGB to HSI. HSI data consisted of three texture sets (hue, saturation, and intensity). This image processing method based on the co-occurrence matrix method (CCM) method gave high accuracy rates (above 95.0%) in an identification of citrus disease.

Zhang et al. (2005) proposed a new method based on wavelet package transform for extraction of image texture features. The wavelet package transform was compared with color co-occurrence matrix. Original images, stone, door, brick, bark, noise, ripple and leaf images, were preprocessed. Once converting the same size gray-scale images, image data sets were compared. From the classification results, an approach based on wavelet package transform had better accuracy of classification than co-occurrence matrix. In this paper, the wavelet transform methods showed better execution efficiency than traditional method based on CCM method.

Pydipati (2004) demonstrated color texture can detect citrus disease on leaves. Algorithms based on image processing techniques for feature extraction and classification were developed. Manual feeding of datasets, in the form of digitized RGB color images were used for feature extraction and training of the SAS statistical classifier. The whole procedure was replicated for three alternate classification approaches to include; Statistical classifier using the Mahalanobis minimum distance method and neural network based classifier using radial basis functions. Comparison of the results obtained from the three approaches was completed and the best approach for the problem at hand was found to be the Mahalanobis based classifier.

### **Spectral Images based Methods and Neural Network Classifiers**

In recent years, hyper-spectral imaging has been used in the food processing and inspection application. Hyper-spectral imaging gives useful information about the full spectral response of the target image across a broad spectral band. Support vector machine (SVM) are a set of supervised generalization linear classifiers that have often been found to provide higher classification accuracies than other widely used pattern classification techniques, such as multilayer perception neural networks (Burges, 1998). The growing interest in SVM has demonstrated its potential by their successful implementation in food and grain analysis, such as classification of starch and wheat.

Jiang et al. (2007) used hyper-spectral fluorescence imaging to analyze the differences between walnut shells and meat. Hyper-spectral fluorescence imaging system scanned samples at 79 different wavelengths ranging from 425nm to 775nm with 4.5 nm increments. Data redundancy of data was reduced by principal component analysis (PCA). This investigation used two statistical pattern recognition methods: Gaussian Mixture Model (GMM) based Bayesian classifier and Gaussian kernel based

Support Vector Machine (SVM). It was found that the SVM method with Gaussian kernel performed more effectively than PCA and FDA reached in classification of walnut meat and shells by hyper-spectral imagery. The overall recognition rate of Gaussian – SVM achieved above 90%, and was about 3% more than that of PCA and 5% more than that of PDA.

Zhang et al. (2007) suggested a creative classification approach for distinguishing healthy and fungal infected wheat kernels during storage. The research showed the potential use of NIR hyper-spectral imaging in grain quality assessment. The research used NIR hyper-spectral imaging and support vector machine (SVM) for identifying the fungi that caused the infection. In this study, 2160 kernels were randomly selected, with 540 kernels from each group of *A. niger*, *A. glaucus*, *Penicillium* spp. With the NIR hyper-spectral imaging and SVM, the overall classification accuracy was 94.8%, with 531 kernels correctly classified and 29 kernels not. This method could classify healthy wheat kernels from infected ones with 100% accuracy, and *Penicillium* spp. infected kernels accuracy of 99.3%.

Mehl et al. (2002) developed a simple multispectral detection system utilizing only three channels in the visible spectral range. They used hyper-spectral imaging to design the multi-spectral image system for rapid detection of apple contaminations. Apple cultivars selected were: Red Delicious, Golden Delicious, and Gala. Despite their color differences, it was possible to use the same configuration. In this study, through hyper-spectral imaging, a rapid multi-spectral imaging analysis for food safety and food quality was designed.

Lee et al. (2005) used the hyper-spectral imaging technique to detect defects on apple peel after harvest. They developed a proper wavelength selection method for detecting the defects. They used principal component analysis (PCA) to detect the fecal contamination on the poultry or apples from hyper-spectral images. They used the correlation analysis method using wavelength differences and band ratio from correlation. For the ratio, the correlation coefficient was 0.91 for the ratio of 683.8nm and 670.3 nm. For the difference of wavelength a high correlation was found at 827.9 nm and 737.8 nm, and the correlation coefficient value was 0.79. The band-ratio by the combination of two spectral images was 0.8. It was the same for the image analysis using the wavelength which was selected from the correlation analysis method. They demonstrated that the correlation analysis method was feasible for selecting the wavelengths to detect defects on apple.

Cheng et al. (2003) presented a new approach combining principal component analysis (PCA) and fisher's linear discriminant (FLD) method. This method maximized the representation and classification effects on the new feature bands extracted from hyperspectral imaging. In this research, the new projected features generated by PCA method gave good results for pattern representation and performed well for obviously separated patterns. For similar patterns, however, FLD method obtained better classification results, but FLD was more sensitive to noise and less stable than PCA. Therefore, they proposed an integrated PCA-FLD method to overcome the drawbacks of the previous two methods, which needed more flexibility in dealing with different sample patterns by adjusting K value properly.

Kim (2002) researched a method for using hyper-spectral data to identify wavebands to be used in multispectral detection systems, and evaluated spatial and spectral responses of hyper-spectral reflectance images of fecal-contaminated apples. To detect fecal-contaminated apples, he presented a systematic method using hyperspectral reflectance imaging technique in conjunction with the use of PCA to define several optimal wavelength bands.

Fukagawa et al. (2003) developed a monitoring system of crop status for precision farming. A multi-spectral imaging sensor, which can get three wave length images (R, G and NIR) simultaneously, was used as the imaging sensor for this system. Leaf height and the number of stems were estimated by Vegetation Cover Rate (VCR). The correlation coefficient between the leaf height and VCR in transplanting variety and seeding variety indicated 0.64, 0.69 respectively. Consequently, the numbers of stems can be estimated by VCR. SPAD value was estimated by LCI is defined as corrected R gray level. The correlation coefficient between the leaf height and VCR in transplanting variety and seeding variety were each 0.79, 0.60. Additionally, LCI map was generated using position and posture data of the helicopter.

### **Near Infrared Monitoring Methods**

Shimada et al. (2008) developed a personal remote sensing system, which consisted of the following components for an extremely narrow area like a Japanese paddy field (100\*100m): radio controlled helicopter, two digital still cameras, network camera board, wireless LAN, notebook computer for the ground station and image processing software. The goal of this project was to develop a low-cost remote sensing system. red-green-red (RGB) and near infrared (NIR) data were acquired with a 4 band camera, system that can calculate normalized difference vegetation index (NDVI) and

so on. The R sensitivity characteristics of the NIR camera, and it became possible for more accurate exposure value compensation to be done. The precision of NDVI improved over former systems as a result.

Yang et al. (2001) developed an infrared imaging and wavelet-based segmentation method for apple defect detection. They proposed that the reflectance spectrum of apple surface in the near-infrared region (NIR) provided effective information for a machine vision inspection system. The differences in light reflectance of the apple surfaces caused the corresponding pixels of bruised areas and good areas to appear different in a NIR apple image. Segmenting the defective areas from the non-defective in apple images was a critical step for apple defect detection. In their work they used a 2-D multi resolution wavelet decomposition to generate 'wavelet transform vectors' for each pixel in the NIR apple images. These vectors were combined and weighted by dynamic modification factors to produce the pixel vectors. The pixels with similar properties are labeled as one class, to separate the defective areas from the good area of apples in the NIR image. They reported 100% accuracy of detecting good apples and 94.6% accuracy of detecting defective apples.

Kawamura et al. (2003) constructed an on-line near-infrared (NIR) spectroscopic sensing system on an experimental basis. This system enables NIR spectra of unhomogenized milk to be obtained during milking over a wavelength range of 600 nm to 1050 nm. In this study, the NIR sensing system could be used to assess milk quality in real time during milking. The system can provide dairy farmers with information on milk quality and physiological condition of individual cows and therefore give them feedback control for optimizing dairy farm management.

## Wavelet Transform

Ionescu and Llobet (2002) used the classification method based on the discrete wavelet transform (DWT) to discriminate between gases and vapours. They compared the discrete wavelet transform (DWT) with the fast Fourier transform (FFT) to extract the texture features from the sensor working temperature. In research, they showed that the discrete wavelet transform (DWT) outperformed the fast Fourier transform (FFT). DWT gave both frequency and information about temperature in the sensor, but FFT provided only frequency information for the complete duration of the signal. Therefore, the DWT analysis helped to extract more important features from the sensor response than the FFT analysis. Through this comparative study, the DWT demonstrated a good linear separation in feature space between mixed vapours. Moreover, algorithms based on DWT can be developed easily and promise for low-cost to operate gas monitors.

Arivazhagan and Ganesan (2003) introduced the texture analysis using the wavelet transform. The texture features were extracted using the co-occurrence method from decomposed images and originals. This research shows texture segmentation algorithms; the images were read and decomposed sub-image using the discrete wavelet transform (DWT). The co-occurrence matrices were implemented for original images and sub-image blocks. The wavelet co-occurrence features (WCF) such as, contrast, cluster shade, and cluster prominence were calculated from the matrices for texture segmentation and the difference between the sums of WCFs of sub-images. After this analysis, post-processing (disk filtering and thresholding techniques) was performed in the segmentation band to remove the noise, and the skeletonizing algorithm was applied to get segmented line of one pixel thickness. The thinned lines are exactly lined up with texture boundaries.

Ruiz, Fdez-Aarria and Recio (2004) studied several texture methods for the classification of remote sensing images with different types of four areas (three forests and one urban). In this research, four approaches were applied for texture analysis and feature extraction: (1) Grey level co-occurrence matrix, (2) Energy filters and edgeness, (3) Gabor filters, (4) Wavelet transform. For comparison of these methods, several combinations of texture variables and different methods were tested, and the method used in the classification step was the maximum likelihood classifier. The table of overall accuracy in this paper showed that the combination of four texture analysis method produced a significant increase in the overall accuracy levels, and the wavelet transform and the grey level co-occurrence matrix also increased the overall accuracy.

### **Control of Citrus Diseases using Image Processing**

Camargo and Smith (2009) developed a visible spectrum imaging system and color image processing to identify banana leaves that were infected with Black Sigatoka. Color images of infected banana leaves were used to develop the image processing which includes color transformation, histogram multi-thresholding, and segmentation. They demonstrated that the algorithm was able to identify the diseased regions in most of the images tested. A machine vision system in the visible region was used by Blasco et al. (2003) to detect external blemishes in apples and peaches. They achieved 93% accuracy in blemish detection. Visible imaging systems are sufficient if the symptoms occur in the visible region, but most of the diseases and conditions could be identified more effectively by exploiting regions beyond the visible. Zarco-Tejada et al. (2005) used a compact airborne spectrographic imager and hyperspectral sensors to monitor the temporal and spatial chlorosis condition of *Vitis vinifera L.* Their results showed that they could estimate chlorophyll content in the narrow band hyperspectral

indices calculated in the 700-750nm. A ground-based hyperspectral imaging system for characterizing vegetation spectral features was developed by Ye et al. (2008). They used a hyperspectral line sensor with a wavelength range of 360-1010 nm to demonstrate the potential of the system to monitor vegetation variability in crop systems. Several researcher (Sepulcre-Canto,2007, Muhammed,2005, and Gowen et al,2007) have reported changes in spectral signatures due to deficiencies of nutrients and damage by pests and environmental factors. In addition, the spectral signatures is influenced by the amount of pigments, leaf angle, leaf surface texture, diseases and stress, plant growth stage, and measurement condition.

While researchers have demonstrated the potential of imaging technologies for disease detection, not much work can be found on the automatic detection citrus greening.

Mishra et al. (2007) investigated the spectral characteristics of leaves infected with HLB and normal leaves using a spectroradiometer (350-2500 nm). The spectral bands of green to red wavelength and the near infrared band were found to have the potential to discriminate HLB-infected leaves from normal leaves by using discriminability, spectral derivative analysis, and spectral ratio analysis. These wavelengths include 530-564 nm, 710-715 nm, 1041 nm, and 2014 nm. In addition, they also developed a four-band active optic sensor to identify young leaf flushes to control the spot spraying of citrus trees. It is known that the Asian citrus psyllid, vector of HLB disease feeds on young leaves. Although they have characterized the spectral reflectance property of HLB infected leaves and discriminate it from normal leaves, other conditions such as nutrient deficiency of leaves could look similar to HLB infected leaves.

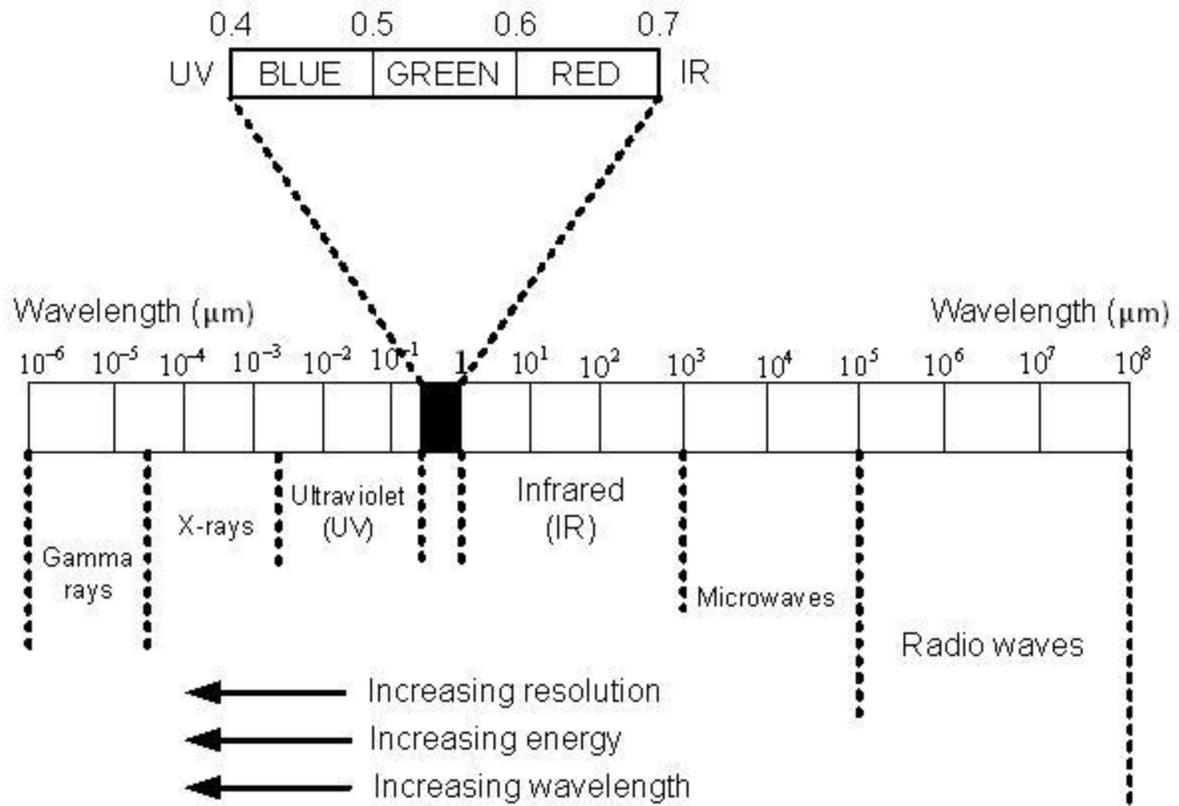


Figure 2-1. The electromagnetic spectrum comprising the visible and non-visible range.

## CHAPTER 3 MATERIALS AND METHODS

This chapter describes the citrus leaf and fruit samples that were collected for this study. It also discusses the features of the two typical optical hardware systems used to collect data. Finally it describes the data analysis methods used to detect citrus disease symptoms.

### **Citrus Canker Samples**

'Ruby Red' grapefruit was selected for this study because it is one of the more popular grapefruit varieties in Florida, and it is susceptible to canker and other common peel diseases. Fruit samples were handpicked from a grapefruit grove near Punta Gorda, Florida, during the spring 2007 because the harvest season is October to June in Florida and Texas. Dr. Tim Schubert and Dr. Gordon Bunn advised to select grapefruit samples. Grapefruit samples of normal market quality and five fruit with abnormal peel conditions were collected because these are very typical peel conditions during the spring. The peel conditions considered were;

- Normal
- Canker, copper burn, greasy spot, melanose, and wind scar

Representative images for each peel condition are shown in Figure 3-1.

Thirty samples for each condition were selected, giving a total of 180 grapefruits tested in this study. All grapefruits were washed and treated with chlorine and sodium o-phenylphenate (SOPP) at the Indian River Research and Education Center of University of Florida in Fort Pierce, Florida. The samples were then stored in an environmental control chamber at 4 °C and were removed from cold storage about 2 hours before imaging to allow them to equilibrate to room temperature.

## Citrus Leaf Samples

Citrus leaves from 'Valencia' orange trees can serve as indicators of common diseases and nutrient deficiencies. Various colors of leaves present the symptoms and reflect the tree health. The leaf samples used in this study were collected from two 'Valencia' orange groves near Immokolee in Southwestern Florida. Dr. Mongi Zekri is the southwest Florida Multi-County Citrus Agent. He provided technical advices for distinction of citrus greening symptoms and season of sample collection. The samples were collected in late spring, summer and fall of 2008. The degree of symptoms varied between leaf samples. The time of season also affects the appearance of the disease symptoms. Hence, in this study, samples were collected at optimal times for symptom detection. In this study, seven different classes of citrus leaves that included symptoms of greening blotchy mottle, greening island, iron deficiency, magnesium deficiency, zinc deficiency, healthy young leaves, and healthy old leaves were selected. The early symptoms of blotchy mottling may be confused by other chlorosis symptoms common to citrus, including nutritional deficiencies, and those which result from some insects and diseases. These leaf samples had 3 different symptom categories:

- Nutritional deficiency: Iron Deficiency, Manganese Deficiency, and Zinc Deficiency
- Citrus Greening symptoms: Blotchy Mottle and Green Islands
- Normal conditions: Normal Young and Normal Mature

Images of Leaf samples are shown in Figures 3-2, 3-3, and 3-4. Figure 3-5 shows various greening symptoms on Leaves used in this study.

Leaf samples were clipped with petioles intact and then sealed in plastic bags to maintain the moisture level of the leaves. Sixty samples of each of the seven classes of leaves were collected. The samples were brought to the laboratory, then sealed in new

bags with appropriate labels and put in environmental control chambers maintained at 4° C. The leaf samples were then taken to an imaging station and images of the upper leaf surface were acquired.

### **Black Spot Samples**

'Valencia' oranges were handpicked from citrus groves near Immokalee in Southwestern Florida in the April of 2010 because symptoms of black spot were first detected in 'Valencia' sweet oranges in the Immokalee area on March 8, 2010. The fruit samples included marketable fruit and those with symptoms of black spot or greasy spot, melanose, or wind scar since these symptoms were progressive during the spring. Dr. Mark Ritenour has contributed to suggestions about sample collection and technical advices. Representative images for each peel condition are shown in Figure 3-6.

All the samples were washed with soap to remove dirt from the grove before imaging the fruit samples. To maximize the collection of hyperspectral images of the limited number of black spot samples, three faces of each fruit (with 120° interval) were collected as shown in Figure 3-7. Table 3-1 demonstrated the sample number for this study. Thus, a total of 525 samples were selected in this study.

### **Image Acquisition Systems**

#### **Charge-Coupled Device Camera System**

In general, a charge coupled device (CCD) camera is equipment which is designed to convert optical brightness into electrical amplitude signals using a plurality of CCDs, and then reproduce the image of a subject using the electric signals without time restriction. CCD vision system is suitable for surface imaging.

Figure 3-8 shows a typical set-up. CCD camera system to collect images uses the computer system installed with image capture software and a 24-bit color frame grabber board with 480x640 pixel resolution.

This system consisted of the following major components.

- Lighting system: Two 13W cool white fluorescent bulbs with reflectors were used.
- Color CCD Camera: The camera consisted of 3-CCD RGB color camera (CV-M90, JAI, San Jose, CA, USA) and a zoom lens (Zoom 7000, Navitar, Rochester, NY, USA).
- Computer System: Coreco PC-RGB 24 bit color frame grabber with 480x640 pixels embedded in the CPU of the computer.

To minimize specular reflectance and shadow and maximize the contrast of the images, the setup of the lighting system was designed. The height of the camera and its focus were adjusted to contain the image of the whole fruit, with an approximate 100 mmx100 mm field of view. Automatic white balance calibration was conducted using a calibrated white balance card before acquiring images from fruit samples. The digital color images were saved in uncompressed BMP format.

Compared to a typical computer vision system, a digital microscope system has several benefits and advance for collecting sample images. This system can easily adjust the magnification using a zoom lens. The magnification can be changed without losing sight of the target you are observing. In addition, optimal magnification to provide the clearest image can be set easily. Integrated illumination system required no setup time. The field view and focus of camera can be simply established. CCD camera system captured images and saved on has 160 GB hard disk drive which can store 575,000 compressed images.

In this research, a digital microscope system (VHX-600K, Keyence, JAPAN) was used for acquiring RGB images from citrus leaf samples, and it is shown in Figure 3-9. These descriptions were extracted from a user`s manual book, released by the Keyence Corporation, Osaka, Japan.

This system is made up of a CCD camera and a controller. The camera unit consist of a high pixel color CCD and light. The controller has various functions such as display, record, measurement, and etc. The stand device offers the user quick observation, analysis and data processing. The imaging system consisted of several parts: High-intensity halogen lamp (12V, 100W), Zoom lens (C-mount lens, OP-51479), 2.11-million-pixel CCD image sensor (1/1.8-inch), 15-inch Color LCD monitor (TFT, 1600x1200, UXGA), Console installed with a hard disk drive (Image format: JPEG and TIFF, Storage capacity: 700MB), and CD-RW drive units, and Control Panel.

- **Color CCD Camera.** The camera uses a pixel-shift technology. This technology allows the maximum resolution to reach 18 million pixels. There are four options for the number of pixels (18 million/8 million/4 million/2.11 million) based on the type of observation to be selected. Lowest resolution (2.11 million, 1/1.8-inch) was used for capturing a target image in this research. Automatic white balance calibration was conducted using a calibrated white balance function before acquiring images from leaf samples.
- **Hard Disk.** The image data can be stored on the built-in hard disk in the controller. It can store up to 575,000 pictures. The leaf sample images were saved in uncompressed JPEG format (1200x1600, 8bit).
- **Built-in Light.** The build-in light system irradiated rays of light directly from the lens. The lighting system was designed to maintain optimum illumination intensity and minimize specular reflectance and shadow. The light system was consisted of a 12V, 100W, Halogen lamp.
- **Zoom Lens.** Digital microscopes came with various types of zoom lenses allowing continuous adjustment of magnification. The lens power and focus were adjusted easily to maintain the image of the whole leaf, with center on the vein.
- **LCD Monitor.** A 15-inch, built-in 1600x1200 pixel high resolution liquid crystal monitor displayed the magnified image.

- **Operation Console.** This console can be used to quickly and easily perform the major observation task such as, adjusting the brightness, colors, and shutter.

### **Hyperspectral Image System**

This section introduces a new technique using hyperspectral imaging in line-scan mode for the rapid acquisition of spatially-resolved scattering profiles as the wavelengths of 450-1000 nm. The instrument set-up and calibration, data analysis, and system validation are presented. The hyperspectral imaging technique is faster and simpler than a solid-state CCD array camera system, and is capable of measuring the optical properties of turbid or opaque sample fruits over a broad spectral range simultaneously.

A hyperspectral line-scan imaging system, as shown in Figure 3-10, was used for acquiring spectral images from citrus leaf and fruit samples. The imaging system consisted of an electron-multiplying charge-coupled-device (EMCCD) camera (Luca, Andor Technology Inc., CT, USA) with imaging spectrograph (ImSpector V10E, Spectral Imaging Ltd., Oulu, Finland) and a C-mount lens (Rainbow CCTV S6X11, International Space Optics, S.A., Irvine, CA, USA), a pair of halogen line lamp (21V, 150W) powered with a DC voltage regulated power supply (Dolan-Jenner Industries, Inc., Lawrence, MA, USA), and a programmable motorized positioning table (XN10-Xslide, Velmex Inc., Bloomfield, NY, USA). This equipment was placed inside in a dark box to eliminate stray external light.

The EMMCCD camera has 1004x1002 pixels and is equipped with a Peltier cooling device to cool the CCD detector to -20°C to improve the dynamic range and the signal-to-noise ratio of the CCD detector. The use of high-performance CCD camera with a large dynamic range is necessary because light attenuation in opaque fruit is so

significant that the diffuse reflectance profile changes dramatically in a short distance. The imaging spectrograph shown in Figure 3-11 is based on prism-grating-prism principle, and does not have moving mechanical components. It has a slit 2.8 mm long and 30  $\mu\text{m}$  wide, and line-scans the sample. If the incoming radiation passes the prism-grating-prism unit, it is dispersed light is projected onto the pixels of the CCD detector, creating a special 2-D image: one dimension represents spatial information and the other dimension spectral (Figure 3-11). When the sample is moved perpendicularly to the scanning direction by the motorized positioning table, one thousand seven hundred and forty line scans were performed for each sample, and four hundred pixels covering the scene of the sample at each scan were saved. Therefore, a three-dimensional image cube associated with a spectral curve is created (Figure 3-12).

The hyperspectral imaging software to transfer data and parameterization was developed using the Andor Software Development Kit (SDK, Luca, Andor Technology Inc., CT, USA) for the hyperspectral line scan imaging system (Kim et al., 2001). An Hg-Ne spectral calibration lamp (Oriel Instruments, Stratford, CT, USA) was used to investigate spectral calibration of the system. Because of low light output in the visible region less than 450 nm, and low quantum efficiency of the EMCCD in the NIR region beyond 930 nm, the wavelength range between 400 nm and 900 nm was used (totaling 92 bands with a spectral resolution of 5.2 nm).

### **Hyperspectral Image System Calibration**

Before the hyperspectral imaging system is used for imaging, proper calibration is required. Calibration requirements may vary depending on the application. Spectral and geometric calibrations are generally required for all applications. Spectral calibration ensures that each pixel on the CCD area array is assigned to an appropriate

wavelength, whereas geometric calibration corrects any distance distortions for individual pixels of the image. Although a complete commercial hyperspectral imaging system is usually pre-calibrated spectrally and geometrically, most hyperspectral imaging system built with components from different dealers is assembled by the researchers. Thus, two forms of calibration and necessary, spectral and geometric calibration must be performed.

As shown in Figure 3-13, geometrical calibration was done with a white grid paper printed with lines 2 mm apart. In system calibration, full-scale (i.e., pixel by pixel) calibrations were not necessary because spatial and spectral distortions such as smile and keystone for any given spatial pixel from the scanning line or any wavelength over the spectral region of 500-1000 nm were occurred in one pixel.

Spectral calibrations were performed using spectral calibration lamps (i.e., xenon lamp (model 6033), mercury-argon lamp (model 6035), New port, Irvine, CA, USA). Figure 3-14 shows the spectral output of 6033 xenon lamp and of 6035 mercury-argon calibration lamps. The spectral peaks from each lamp and their pixel position in the original images were identified. The relationship between the vertical band position and wavelength from the two calibration lamps was established using a linear regression function (Figure 3-15).

## **Pre-processing Techniques**

### **Visible Spectrum Image Pre-processing**

The images were taken for each sample class and were stored in JPEG format. The images from each class were divided into two datasets consisting of samples for training and testing. The samples were first arranged in ascending order for the time the

images were acquired. This approach minimizes negative time dependant variability, and reduces potential for data selection bias between the training and test datasets.

All algorithms for image segmentation and texture feature generation were developed in MATLAB. In the initial step, the RGB images of all leaf samples were obtained. For reducing the computational burden with minimal loss of texture feature quality, the image resolution was reduced from 1600x1200 pixels to 800x600 pixels and the reduced images were then converted from eight bit to six bit per channel RGB format. The subsequent steps were repeated for each image in the dataset.

### **ROI selection and color space conversion for canker detection**

ROI images were first extracted from the original RGB color images with the dimension of 480x640 pixels, generating small images covering the interested areas (i.e., normal peels or various diseases) on the fruit surface. Referring to Figure 3-16, the ROI selection was started manually by determining a point on the original image, and then was finished by a Matlab program for extracting a square portion with the dimension of 64x64 pixels centered on the determined point. This approach obtains the useful image data and significantly reduces the computational burden for the following data analysis procedures. Representative ROI images for each fruit peel condition used in this study are shown in Figure 3-17.

The ROI images were then converted from the original eight bit per channel red, green, blue (RGB) color representation to a six bit per channel hue, saturation, and intensity (HSI) color representation to facilitate the SGDM calculation. Intensity is calculated using the mean value of the three RGB values. The hue and saturation values are determined using a geometrical transformation of the International Commission on Illumination's (CIE) chromaticity diagram (Ohta, 1985). In this process,

the CIE chromaticity diagram represents a two-dimensional hue and saturation space (Wyszecki and Stiles, 1992). The RGB values determine the chromaticity coordinates on the hue and saturation space, which are then used to geometrically calculate the value of hue and saturation.

### **Edge detection for greening detection**

After the images were reduced, edge detection of the leaf was completed on each image of the leaf sample using MATLAB program file. Figure 3-18 exhibits a detailed edge detection process. First, as shown in Figure 3-19, each RGB image was converted to a gray image and then a binary image. Next, edge of a binary image was detected by the command 'imerode' and 'imdilate' in MATLAB.

Once the edge detection was finished, the image was scanned from left to right for each row in the pixel map, and the area outside the leaf was zeroed to remove any background noise. In next step, the images were converted from RGB format to HSI format. A sample edge detected image of the leaf sample was shown in the following Figure 3-20.

### **Hyperspectral Image Pre-processing**

Figure 3-21 shows the details of pre-processing steps. Image preprocessing was first performed for the original hyperspectral reflectance images to fulfill flat-field correction, spatial data reduction and image masking, resulting in normalized and masked hyperspectral data with dimension of 870x200x92 (92band).

Flat-field corrections were performed on the hyperspectral images to obtain the relative reflectance prior to image analysis and image processing for classification. Equation (3-1) was used for the flat-field correction to obtain the relative reflectance R

for the 92 spectral bands. Flat-field correction technique can reduce uneven illumination and distortion.

$$R(w) = \frac{R_{\text{sample}}(w) - R_{\text{dark}}(w)}{R_{\text{White}}(w) - R_{\text{dark}}(w)} \times w \quad (3-1)$$

where  $R(w)$  is the relative reflectance,  $R_{\text{sample}}(w)$  is the original sample image,  $R_{\text{White}}(w)$  is the reference image acquired from a white Spectralon panel,  $R_{\text{dark}}(w)$  is the dark images obtained with cap covering the camera lens, and  $w$  is the wavelength,  $r$  is reflectance factor. Referring to Figure 3-22, the actual reflectance factor for the Spectralon panel is about 99% in the wavelength range measured by the hyperspectral imaging system, but a reflectance factor was assumed to be 100% in this study for simplicity. The relative reflectance was adjusted to the value in the range of 0 to 10,000 so the resulting image would be in the range of the original data from 0-16383 (the 14-bit EMCCD). The adjusted value could reduce rounding errors for further data analysis.

Next, the mask template of the relative reflectance image was created by the finding the threshold value. To reduce processing time, the leaf area was separated from the background by creating a mask. The mask was created using a threshold value determined from the hyperspectral image which gave the largest contrast between the leaf surfaces with the background. By using the mask, the image resolution was reduced to half of the original size. The reduction of the images serves several important purposes: 1) it reduces the computational burden of the redundant features, 2) it tends to improve the performance of classification algorithms, and 3) it reduces memory and storage demands. The reduction brought comparable spatial resolutions for vertical and horizontal dimensions.

## **Color Co-occurrence Methodology**

Image data includes large amounts of information, such as color, light, texture and shape. These properties are used in image processing and computer vision algorithm. In these properties, color, shape and light can be changed by surroundings. For example, chameleons have a special ability to match their skin color to surroundings. This ability confuses a prey to escape dangerous situation. If background of an input image becomes too dark or too light, the object in the image cannot be perceptible in its background. Images of real objects do not have uniform properties, but texture can give information about the image through repeating pattern. This characteristic of texture segmentation is very important in machine vision and image processing, and a variety of texture analysis methods have been applied in various fields of study.

Tuceryan (1998) introduced various aspects of texture analysis. According to the paper, there are three main applications, inspection, medical image analysis, and document processing. "Inspection means the detection of defects in texture images or textile inspection. Medical image analysis has involved the automatic extraction of features from images which is used for a variety of classification, such as distinguishing normal tissues from abnormal tissue. Document processing has applications ranging from postal address recognition to analysis and interpretation of maps." This paper also presented three texture methods; statistical, geometrical, and model based methods. Statistical methods were proposed early and used widely. This study has used spatial distribution of gray values. The key word of geometrical methods was "texture elements" or primitives. After extracting texture elements in image, the texture features were utilized. Model based methods is "the construction of an image model that can be used not only to describe texture, but also to synthesize it"

In this thesis, the image analysis method was the color co-occurrence matrix (CCM) statistical method. Several researchers have applied this method to agricultural application. Bucks (2000) used color co-occurrence method (CCM) for detection of weed, and Pydipati et al. (2005) applied disease detection in citrus trees. Ondimu et al (2008) compare the plant water stress in sunagoke moss using color co-occurrence matrix. These papers showed good result and high accuracy in classification.

Before applying CCM method to input images, the original image consisting of a red, green, and blue (RGB) color space are converted to a HSI color space. HSI space is distributed into hue, saturation, and intensity components. Most of image processing engines and methods are based on HSI color space system. This color system has strong tolerance for a change of a light on an image or a reflection. This characteristic of HSI can help image processing be less sensitive to illumination of surroundings.

Each pixel map was used to generate a color co-occurrence matrix after the HSI image was completed, resulting in three CCM matrices. That is, one CCM matrix for each of the HSI pixel maps. Through the use of spatial gray-level dependence matrices (SGDM's), the color co-occurrence texture analysis method was developed. The spatial gray-level dependence matrices (SGDM's) is related with gray level co-occurrence matrix (GLCM) because of second-order statistics. Haralick et al. (1979) proposed the use of gray level co-occurrence matrix (GLCM) method. As his paper, the matrix  $p_a(i, j)$  is consisted of distance ( $a$ ), gray-level. The operator  $a$  means distance between gray-level  $i$  from  $j$ .

As an example, consider the following  $4 \times 4$  image containing 4 different gray values:

$$I_{x,y} = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 3 & 3 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 2 & 2 \end{bmatrix} \quad (3-2)$$

This  $4 \times 4$  matrix was converted to SGDM matrix, " $P_a, a = (1,0)$ " as follows:

$$I_{x,y} = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 3 & 3 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 2 & 2 \end{bmatrix} \rightarrow P_a(i,j) = \begin{bmatrix} 2 & 1 & 2 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3-3)$$

The matrix  $P_{(1,0)}(0,0)$  means the number of times a pixel with a vector (1, 0) from gray-level "0" to "0".

Shearer (1986) illustrated the SGDM presented by the function  $P(i, j, d, \Theta)$ . It is similar to function  $P_a(i, j)$ , but this function does not use vector (i,j) but distance (d) and an orientation angle ( $\Theta$ ).

As shown in Figure 3-23, all the neighbors from 1 to 8 are numbered in a clockwise direction. Equation 3-4 presents an example image matrix.

$$I_{x,y} = \begin{bmatrix} 0 & 0 & 2 & 3 \\ 0 & 0 & 3 & 3 \\ 3 & 2 & 1 & 1 \\ 2 & 2 & 1 & 1 \end{bmatrix} \rightarrow P(i,j, 1,0) = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 1 & 2 & 1 & 2 \\ 1 & 0 & 2 & 1 \end{bmatrix} \quad (3-4)$$

Haralick and Shanmugam (1974) developed a set of 16 texture features associated with the SGDM's. After these features were founded, Shearer (1986) expanded these features to use the hue, saturation and intensity color features. Also, Shearer reduced 16 features to 11 features using the method developed by Haralick and Shanmugam. For classifying cancer tissue, Donohue et al. (2001) suggested that image contrast and modus texture features were added to original 11 texture features. Therefore, the color co-occurrence matrices (CCM) consisted of these matrices, once

each for the hue, saturation and intensity features. This resulted in 39 color texture features.

### Feature Extraction

Shearer and Holmes (1990) defined related equations with a brief description as pertains to intensity. Shearer (1986) also applied to saturation with similar descriptions. Hue values are different from intensity and saturation values, so the values are treated very differently in the analysis by Shearer (1986). The descriptions of texture feature equations below were found in Burks (1997). The CCM matrices are normalized using the equation as follows;

Matrix Normalization:

$$p(i, j) = \frac{p(i, j, 1, 0)}{\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p(i, j, 1, 0)} \quad (3-5)$$

Marginal probability matrix:

$$p_x(i) = \sum_{j=0}^{N_g-1} p(i, j) \quad (3-6)$$

Sum and difference matrices:

$$p_{x+y}(k) = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p(i, j) \quad (3-7)$$

$$K = i + j ; \text{ for } k = 0, 1, 2, \dots, 2(N_g - 1) \quad (3-8)$$

$$p_{x-y}(k) = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p(i, j) \quad (3-9)$$

$$K = |i - j| ; \text{ for } k = 0, 1, 2, \dots, 2(N_g - 1) \quad (3-10)$$

where

$P(i, j)$ =the image attribute matrix, and

$N_g$ =the total number of attribute levels

After normalized CCM matrices, texture features are extracted using equations as follows;

Texture features:

The angular second moment (I1) is a measure of the image homogeneity

$$I_1 = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} [p(i,j)]^2 \quad (3-11)$$

The mean intensity level (I2) is a measure of image brightness derived from the co-occurrence matrix.

$$I_2 = \sum_{i=0}^{N_g-1} i p_x(i) \quad (3-12)$$

Variation of image intensity is identified by the variation textural feature (I3).

$$I_3 = \sum_{i=0}^{N_g-1} (i - I_2)^2 p_x(i) \quad (3-13)$$

Correlation (I4) is a measure of the intensity linear dependence in the image.

$$I_4 = \frac{\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} i j p(i,j) - I_2^2}{I_3} \quad (3-14)$$

The product moment (I5) is analogue to the covariance of the intensity co-occurrence matrix.

$$I_5 = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (i - I_2)(j - I_2) p(i,j) \quad (3-15)$$

Contrast of an image can be measured by the inverse difference moment (I6).

$$I_6 = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \frac{p(i,j)}{1+(i-j)^2} \quad (3-16)$$

The entropy feature (I7) is a measure of the amount of order in an image.

$$I_7 = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p(i,j) \ln p(i,j) \quad (3-17)$$

The sum and difference entropies (I8 and I9) are not easily interpreted, yet low entropies indicate high levels of order.

$$I_8 = \sum_{k=0}^{2(N_g-1)} p_{x+y}(k) \ln p_{x+y}(k) \quad (3-18)$$

$$I_9 = \sum_{k=0}^{N_g-1} p_{x-y}(k) \ln p_{x-y}(k) \quad (3-19)$$

The information measures of correlation (I10 and I11) do not exhibit any apparent physical interpretation.

$$I_{10} = \frac{I_7 - HXY1}{HX} \quad (3-20)$$

$$I_{11} = [1 - e^{-2(HXY2 - I_7)}]^{1/2} \quad (3-21)$$

Where,

$$HX = - \sum_{i=0}^{N_g-1} p_x(i) \ln p_x(i) \quad (3-22)$$

$$HXY1 = - \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p(i,j) \ln[p_x(i)p_x(j)] \quad (3-23)$$

$$HXY2 = - \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p_x(i)p_x(j) \ln[p_x(i)p_x(j)] \quad (3-24)$$

The contrast feature (I12) is a measure of salience in object recognition.

$$I_{12} = \sum_{|i-j|=0}^{N_g} (i-j)^2 \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p(i,j) \quad (3-25)$$

The modus feature (I13) is a measure of maximum value in images.

$$I_{13} = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \max [p(i,j)] \quad (3-26)$$

Intensity texture feature equations are presented in Table 3-1.

### Principal Component Analysis

Principal component analysis (PCA) is an unsupervised method to find the optimal features from the data. PCA projects d-dimensional data onto lower-dimensional subspace in a way that is optimal in a sum-squared error sense. Duda et al.(2000) presented the basic approach in *karhunen-loève transform* in their book.

Given n samples of m-dimensional data, represented as the m-by-n matrix.

$X = [x_1, x_2, \dots, x_n]$ , the sample mean is

$$m = \frac{1}{n} \sum_{j=1}^n x_j \quad (3-27)$$

where  $x_j$  is the jth column of X.

Let e be a unit vector in the direction of the line. Then the equation of the line can be written as

$$\mathbf{x} = \mathbf{m} + \mathbf{a}\mathbf{e}, \quad (3-28)$$

where the scalar  $\mathbf{a}$  (which takes on any real value) corresponds to the distance of any point  $\mathbf{x}$  from the mean  $\mathbf{m}$ . If we represent  $x_k$  by  $m + a_k e$ , we can find an "optimal" set of coefficients  $a_k$  by minimizing the square-error criterion function. The squared-error criterion function can be written as

$$\begin{aligned} J_1(a_1, \dots, a_n, e) &= \sum_{k=1}^n \|(m + a_k e) - x_k\|^2 = \sum_{k=1}^n \|a_k e - (x_k - m)\|^2 \\ &= \sum_{k=1}^n a_k^2 \|e\|^2 - 2 \sum_{k=1}^n a_k e^t (x_k - m) + \sum_{k=1}^n \|x_k - m\|^2 \end{aligned} \quad (3-28)$$

Recognizing that  $\|e\| = 1$ , partially differentiating with respect to  $a_k$ , and setting the derivative to zero, we obtain

$$a_k = e^t (x_k - m) \quad (3-29)$$

This brings us to the more interesting problem of finding the best direction  $\mathbf{e}$  for the line. The sample covariance matrix arises here when we substitute  $a_k$  found in Equation 3-29 into Equation 3-28 to obtain

$$\begin{aligned} J_1(e) &= \sum_{k=1}^n a_k^2 - 2 \sum_{k=1}^n a_k^2 + \sum_{k=1}^n \|x_k - m\|^2 \\ &= -\sum_{k=1}^n [e^t (x_k - m)]^2 + \sum_{k=1}^n \|x_k - m\|^2 \\ &= -\sum_{k=1}^n e^t (x_k - m) (x_k - m)^t e + \sum_{k=1}^n \|x_k - m\|^2 \\ &= -e^t \mathbf{S} e + \sum_{k=1}^n \|x_k - m\|^2 \end{aligned} \quad (3-30)$$

Clearly, the vector  $\mathbf{e}$  that minimizes  $J_1$  also maximizes  $e^t \mathbf{S} e$ . We use the method of Lagrange multiplier to maximize  $e^t \mathbf{S} e$  subject to the constraint that  $\|e\| = 1$ . Letting  $\lambda$  be undetermined multiplier, we differentiate

$$\mathbf{u} = e^t \mathbf{S} e - \lambda (e^t e - 1) \quad (3-31)$$

with respect to  $\mathbf{e}$  to obtain

$$\frac{\partial \mathbf{u}}{\partial \mathbf{e}} = 2\mathbf{S}e - 2\lambda e \quad (3-32)$$

Setting this gradient vector equal to zero, we see that  $e$  must be an eigenvector of the covariance matrix:

$$Se = \lambda e \quad (3-33)$$

In particular, because  $e^t Se = \lambda e^t e = \lambda$ , it follows that to maximize  $e^t Se$ , we want to select the eigenvector corresponding to the largest eigenvalue of the covariance matrix. In other words, to find the best one-dimensional projection of the data (best in the least-sum-of-squared-error sense), we project the data onto a line through the sample mean in the direction of the eigenvector of the covariance matrix having the largest eigenvalue.

The  $d$ -dimensional mean vector  $\mu$  and  $d \times d$  covariance matrix  $\Sigma$  are computed for the full data set. Next, the eigenvectors and eigenvalues are computed, and sorted according to decreasing eigenvalue. Call these eigenvectors  $e_1$  with eigenvalue  $\lambda_1$ ,  $e_2$  with eigenvalue  $\lambda_2$  and so on, and choose the  $k$  eigenvectors having the largest eigenvalues. Often there will be just a few large eigenvalues, and this implies that  $k$  is the inherent dimensionality of the subspace governing the "signal" while the remaining  $d - k$  dimensions generally contain noise. We form a  $d \times k$  matrix  $A$  whose columns consist of the  $k$  eigenvectors. The representation of data by principal components consists of projecting the data onto the  $k$ -dimensional subspace according to

$$x' = F_1(x) = A^t(x - \mu) \quad (3-34)$$

### Singular Vector Divergence

The procedures for deriving singular vector divergence (SVD) are summarized in this section. Wall et al. (2003) described the mathematical definition of singular vector divergence (SVD). The equation for singular value decomposition of  $X$  is given by

$$X = USV^T \quad (3-35)$$

$X$  means an  $m \times n$  matrix of real-valued data and rank  $r$ , where without loss of generality  $m \geq n$ , and therefore  $r \leq n$ . In the case of microarray data,  $x_{ij}$  is the expression level of the  $i^{\text{th}}$  gene in the  $j^{\text{th}}$  assay. The elements of the  $i^{\text{th}}$  row of  $X$  form the  $n$ -dimensional vector  $\mathbf{g}_i$ , which we refer to as the transcriptional response of the  $i^{\text{th}}$  gene. Alternatively, the elements of the  $j^{\text{th}}$  column of  $X$  form the  $m$ -dimensional vector  $\mathbf{a}_j$ , which we refer to as the expression profile of the  $j^{\text{th}}$  assay.

$U$  is an  $m \times n$  matrix. The columns of  $U$  are called the left singular vectors,  $\{\mathbf{u}_k\}$ , and form an orthonormal basis for the assay expression profiles, so that  $\mathbf{u}_i \cdot \mathbf{u}_j = 1$  for  $i = j$ , and  $\mathbf{u}_i \cdot \mathbf{u}_j = 0$  otherwise.  $S$  is an  $n \times n$  diagonal matrix. The elements of  $S$  are only nonzero on the diagonal, and are called the singular values.

Thus,  $S$  is diagonal matrix,

$$S = \begin{bmatrix} S_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & S_n \end{bmatrix} \text{ and } s_k > 0 \text{ for } 1 \leq k \leq r, \text{ and } s_i = 0 \text{ for } (r+1) \leq k \leq n. \quad (3-36)$$

$V^T$  is also an  $n \times n$  matrix. The rows of  $V^T$  contain the elements of the right singular vectors,  $\{\mathbf{v}_k\}$ , and form an orthonormal basis for the gene transcriptional responses.

One important result of the SVD of  $X$  is that

$$X^{(l)} = \sum_{k=1}^l \mathbf{u}_k s_k \mathbf{v}_k^T \quad (3-37)$$

is the closest rank- $l$  matrix to  $X$ . The term “closest” means that  $X^{(l)}$  minimizes the sum of the squares of the difference of the elements of  $X$  and  $X^{(l)}$ ,  $\sum_{ij} |x_{ij} - x^{(l)}_{ij}|^2$ .

One way to calculate the SVD is to first calculate  $V^T$  and  $S$  by diagonalizing  $X^T X$ :

$$X^T X = V S^2 V^T \quad (3-38)$$

and then to calculate  $U$  as follows:

$$U = X V S^{-1} \quad (3-39)$$

where the  $(r+1), \dots, n$  columns of  $V$  for which  $s_k = 0$  are ignored in the matrix multiplication of Equation 3-39. Choices for the remaining  $n-r$  singular vectors in  $V$  or  $U$  may be calculated using the Gram-Schmidt orthogonalization process or some other extension method. In practice there are several methods for calculating the SVD that are of higher accuracy and speed. Section 4 lists some references on the mathematics and computation of SVD.

### Correlation Analysis

Gonzalez et al. (2008) described that the correlation analysis measures the relationship between two variables. The correlation coefficient, defined as

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (3-40)$$

Where  $\sigma_{xy}$ ,  $\sigma_x$  and  $\sigma_y$  are

$$\sigma_{xy} = \varepsilon[(x - \mu_x)(y - \mu_y)] = \sum_{x \in X} \sum_{y \in Y} (x - \mu_x)(y - \mu_y)P(x, y) \quad (3-41)$$

$$\sigma_x = \sqrt{\varepsilon[(x - \mu_x)^2]} = \sqrt{\sum_{x \in X} \sum_{y \in Y} (x - \mu_x)^2 P(x, y)} \quad (3-42)$$

$$\sigma_y = \sqrt{\varepsilon[(y - \mu_y)^2]} = \sqrt{\sum_{x \in X} \sum_{y \in Y} (y - \mu_y)^2 P(x, y)} \quad (3-43)$$

$\rho$  is a normalized covariance, and must always be between -1 and +1. If  $\rho = +1$ , then  $x$  and  $y$  are maximally positively correlated, while if  $\rho = -1$ , they are maximally negatively correlated. If  $\rho = 0$ , the variables are uncorrelated. It is common practice to consider variables to be uncorrelated for practical purpose if the magnitude of their correlation coefficient is below some threshold, such as 0.05, although the threshold that makes sense does not depends on the actual situation.

If  $x$  and  $y$  are statistically independent, then for any two functions  $f$  and  $g$  we obtain

$$\varepsilon[f(x)g(x)] = \varepsilon[f(x)]\varepsilon[g(x)] \quad (3-44)$$

a result which follows from the definition of statistical independence and expectation.

Note that if  $f(x) = x - \mu_x$  and  $g(x) = y - \mu_y$ , this theorem again shows that

$\sigma_{xy} = \varepsilon[(x - \mu_x)(y - \mu_y)]$  is zero if  $x$  and  $y$  are statistically independent.

### Wavelet Transform

The wavelet transform (WT) in digital image processing is one of the practical tools to transform images and signals. The Haar transform, which is the simplest orthogonal wavelet transform, was selected in this study. The basic idea of the Haar transform for 2-D multiresolution data is that it is computed by iterating difference and averaging between odd and even samples of input data. Since the data is in 2D, the average and difference in both the horizontal and vertical directions are computed. The main formula for the wavelet transform of image  $f(x, y)$  of size  $M \times N$  can be described by the following equation (Gonzalez, 2008):

$$W_\varphi(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \varphi_{j_0, m, n}(x, y) \quad (3-45)$$

$$W_\psi^i(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \psi_{j_0, m, n}^i(x, y), \quad i = \{H, V, D\} \quad (3-46)$$

Where

$$\varphi_{j_0, m, n}(x, y) = 2^{j/2} \varphi(2^j x - m, 2^j y - n) \quad (3-47)$$

$$\psi_{j_0, m, n}^i(x, y) = 2^{j/2} \varphi(2^j x - m, 2^j y - n), \quad i = \{H, V, D\} \quad (3-48)$$

$\{H, V, D\}$  are horizontal along the  $x$  axis, vertical along the  $y$  axis, and diagonal along  $y=x$ .

To obtain the image features using the Haar wavelet transform, the images were taken of the top surface for each leaf class and centered on the mid. Algorithms for texture feature generation were developed in MATLAB. Due to the minimal loss of

texture feature quality and orthogonal wavelet transform, the image resolution was rescaled to 256 × 256 pixels. The subsequent steps were repeated for each image in the dataset. After the images were rescaled, decomposition of the images was completed on each image of the leaf sample using the MATLAB. As shown in Figure 3-24, each image was decomposed to four-level wavelet transform. The four-level wavelet transform produce a total of 13 sub bands, including vertical, horizontal and diagonal bands. In the next step, the texture features were then extracted from each sub-image.

Once the 4-level decompositions were generated for each image, texture features were extracted using energy of each sub bands in the image. The average energy of approximation and detail sub-image of four level decomposed images are calculated as features using the formulas given as follows:

$$E = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N W_i^2(i, j) \quad (3-49)$$

Where M,N denote the size of sub-image, w(i,j) denotes the value of pixel of image

Therefore, a total of 13 sub bands were obtained and each image has 13 texture features. In the classification phase, Fisher linear discriminant was used for Wavelet Transform based classification. Fisher's linear discriminant method is that a line projects high-dimensional data and classification in a one-dimensional space. The main formula for calculating Fisher's linear discriminant is giving by:

$$J(w) = \frac{w^T S_w w}{w^T S_B w} \quad (3-50)$$

where  $S_B$  is the "between classes scatter matrix" and  $S_w$  is the "within classes scatter matrix"

The definition of scatter matrix is:

$$S_B = \sum_c N_c (\mu_c - \bar{x})(\mu_c - \bar{x})^T \quad (3-51)$$

$$S_W = \sum_c \sum_{i \in c} N_c (x_i - \mu_c)(x_i - \mu_c)^T \quad (3-52)$$

$$\mu_c = \frac{1}{N_c} \sum_{i \in c} x_i \quad (3-53)$$

where

$$\bar{x} = \frac{1}{N} \sum_i x_i = \frac{1}{N} \sum_c N_c \mu_c \quad (3-54)$$

and  $N_c$  is the number of cases in class  $c$ .

### Pattern Recognition

Pattern recognition is a part of artificial intelligence. Artificial intelligence is an approach that mimics human's intelligence to build learning ability, reasoning, and perception. The research about artificial intelligence has developed into a discipline known as intelligent system.

Bishop (2006) presented pattern recognition techniques consisted of features and patterns. A feature is specific aspect, quality and characteristic of some objects. The feature can be color, a symbol sign, and numerical value, such as distance, height, and weight. If features have two or more numerical values, the features can express d-dimension row features called as a feature vector, and this d-dimension space defined as a feature vector is a feature space (Bishop, 2006).

An object can be expressed as points modeled by a feature vector. Such as, a plot of features expressed as points in space of a scatter plot. This plot can be expressed visually to 3-dimensional a space. If the feature space is 4 or more dimension, the feature vector cannot be plotted, but still exists in a n-dimensional space. Figure 3-25

below shows examples of various features. Pattern means traits or features of an individual object, and is defined as a set of feature together. Features and patterns are similar concepts, but features form pattern. In pattern recognition, the pattern is expressed by  $\{x, \omega\}$ ,  $x$  is feature vector observed and  $\omega$  is the unique class of the feature vector. This class is also called category, group, or label. The feature vector selected to represent the class is very important and affects the selection of pattern recognition algorithms and the cognitive approach. Hence, the feature represents characteristics of the classes that make them distinguishable. In other words, samples from one should have similar features within the class, while samples from another class should have different feature values. Figure 3-26 presents good feature separation on the left and poor separation on the right. The feature vector with pattern can be classified by its distribution type as follows.

- Linear distribution.
- Nonlinear distribution.
- High correlation distribution.
- Multi-class distribution.

Figure 3-27 shows distribution plots of pattern types. When feature patterns are distributed by linear distribution type, the class can be classified more easily.

There are various approaches in pattern recognition, and these approaches have demonstrated success in a variety of research areas such as aerospace, defense, medical, neurobiology, and linguistics etc. In this study, three pattern recognition approaches were used for finding an optimal pattern recognition approach.

- Linear models for classification: Fisher's linear discriminant analysis method.

- Neural Networks for classification: Back propagation based on neural network method.
- Nonlinear Kernel for classification: Support Vector Machine (SVM).

### Fisher`s Linear Discriminant Method

The central idea of Fisher's linear discriminant method is that a line projects high-dimensional data and classification in a one-dimensional space.

M. Welling (2006) introduced Fisher's linear discriminant analysis where the formula for calculating Fisher's linear discriminant is giving by:

$$J(w) = \frac{w^T S_w w}{w^T S_B w} \quad (3-55)$$

where  $S_B$  is the "between classes scatter matrix" and  $S_w$  is the "within classes scatter matrix"

The definition of scatter matrix is:

$$S_B = \sum_c N_c (\mu_c - \bar{x})(\mu_c - \bar{x})^T \quad (3-56)$$

$$S_w = \sum_c \sum_{i \in c} N_c (x_i - \mu_c)(x_i - \mu_c)^T \quad (3-57)$$

$$\mu_c = \frac{1}{N_c} \sum_{i \in c} x_i \quad (3-58)$$

where

$$\bar{x} = \frac{1}{N} \sum_i x_i = \frac{1}{N} \sum_c N_c \mu_c \quad (3-59)$$

and  $N_c$  is the number of cases in class  $c$ .

### Neural Network Based on Back Propagation Network Method

The main idea of a back-propagation network is that the connecting link weights between hidden layers are found by back-propagating the errors from the output layer. Back-propagation networks typically use multilayer networks. Figure 3-28 illustrates a network diagram for a multilayer network. Nodes are used to represent the input, hidden, and output variables, and links between the nodes describe the weight parameters. The

link contribution from additional input and hidden variables,  $X_0$  and  $Z_0$ , represent the bias parameters. The process flow of this network follows the arrows shown in Figure 3-28.

C. M. Bishop (2006) described the derivation of a back-propagation algorithm in his book "Pattern recognition and machine learning".

$$w_{ji}^h(t+1) = w_{ji}^h(t) + \eta \delta_j^h x_i \quad (3-60)$$

where  $\eta$  is learning rate,  $w_{ji}^h$  is weight from unit  $j$  to unit  $i$ ,  $\delta_j^h$  is a error of output from unit  $j$  to,  $x_i$  is input pattern.

Using equation 3-60, we can calculate the weight function using forward propagation as shown in Figure 3-29. The output error is updated. After each forward propagation and used to re-calculate the hidden layer weights using the back-propagation formula. The back-propagation formula can be described by the following equation:

$$\delta_j = h'(a_j) \sum_k w_{kj} \delta_k \quad (3-61)$$

where

$$h'(a_j) = 1e^{-y_h^2} \quad (3-62)$$

Figure 3-29 tells us that the value of  $\delta$  for a particular hidden unit can be obtained by propagating the  $\delta$ 's backwards from units higher up in the network.

The neural network application in this research was designed by the MATLAB PRTools toolbox (Faculty of Applied Physics, Delft University of Technology, The Netherlands). This application used the back-propagation formula of the neural network which shown above.

Appendix B shows a detail MATLAB routine. First, the data files would be loaded. The data files had a reduced texture features about each conditions. Second, the data files were divided between training sets and test sets using the command 'dataset'. The command 'dataset' consists of a set of 'm' objects, each given by 'k' features. A 'm' by 'k' rows represents such a dataset in this MATLAB routine. Third, after importing the training matrix and the testing matrix, the network was trained using the function 'BPXNC'.

The syntax for this function is as follows:

```
[W,HIST] = BPXNC (A,UNITS,ITER,W_INI,T,FID)
```

where

A - Dataset

UNITS- Array indicating number of units in each hidden layer (default: [5])

ITER - Number of iterations to train (default: inf)

W\_INI- Weight initialisation network mapping (default: [], meaning  
initialisation by Matlab's neural network toolbox)

T - Tuning set (default: [], meaning use A)

FID - File descriptor to report progress to (default: 0, no report)

After training, the test data for each class was simulated using the function 'testc'.

### **Support Vector Machine Method**

The support vector machine is very useful approach for classification. In recent years, SVMs method has been developed by many researchers and applied in various fields. In general, the SVM method handles 2 class problems, however multiclass SVMs have been developed by various researchers. The multiclass solution is based on the two-class support vector machine. Gunn (1998) presented the support vector

classification. Figure 3-30 shows numerous possible linear classifiers for separating the data, but only one line maximizes the distance between it and the nearest data point of each class. This linear classifier is called the optimal separating hyper plane or hyper line. He formalized the SVM main problem as, how to find optimal separating hyper plane in n-dimension space. The main formula for the mathematical analysis is equation 3-68.

A separating hyper plane in canonical form must satisfy the following constraints,

$$y_i[(w \cdot x) + b] \geq b_{\text{separating}} \quad (3-63)$$

The distance  $d(w,b;x)$  of point  $x$  from the hyperplane  $(w,b)$  is,

$$d(w, b; x) = \frac{|w \cdot x + b|}{\|w\|} \quad (3-64)$$

The optimal hyperplane is provided by maximizing the margin,  $\rho(w,b)$ , subject to the constraints of equation (3-64).

Where the margin is given by,

$$\rho(w, b) = \min_{\{x_i: y_i=1\}} d(w, b; x_i) + \min_{\{x_j: y_j=-1\}} d(w, b; x_j) \quad (3-65)$$

$$= \min_{\{x_i: y_i=1\}} \frac{|w \cdot x_i + b|}{\|w\|} + \min_{\{x_j: y_j=-1\}} \frac{|w \cdot x_j + b|}{\|w\|} \quad (3-66)$$

$$= \frac{2}{\|w\|} \quad (3-67)$$

Hence, the hyperplane that optimally separates the data is the one that minimizes

$$\phi(w) = \frac{1}{2} \|w\|^2 \quad (3-68)$$

Through this analysis, we can find the maximum margin hyper plane.

### Statistical Classification

After the color co-occurrence matrices (CCM) were generated for input images, each input image has 39 texture features. Because the proper number of texture

features causes the reduced computational requirement such as, time and computer hardware and complexity for classification, it is an important procedure to eliminate redundant texture features. SAS offers this procedure and a discriminant classifier between image classes after training the statistical classification model.

Burks (1997) introduced the procedure for accomplishing above tasks. PROC STEPDISC is a useful function to reduce the number of texture feature by a stepwise selection process. In Stepwise procedure, the main assumption is that all of the classes included in the data set are to be multi-variate normal with a common covariance matrix. The process has two different conditions related with variance. Burks (1997) explained as follows.

"First, the variable within the model which contributes least to the model, as determined by the Wilks' lambda method that does not pass the test to stay is removed from the model. Secondly, the variable outside the model which contributes most to the model and passes the test to be admitted is added. When no more steps can be taken the model is reduced to its final form."

In PROC DISCRIM procedure, the classification accuracy may be determined by a discriminant function established using a measure of the generalized squared distance between the image texture variables and the class texture variable means, and the posterior probability. The pooled covariance matrix of the training set texture variables and the prior probabilities of the classification groups may affect the classification criterion.

## Hyperspectral Image Classification Methods

The Spectral Angle Mapper (SAM) and Spectral Information Divergence (SID) are the two supervised classification methods used in analyzing the spectral characteristics of the peel conditions.

### Spectral Angle Mapper

The Spectral Angle Mapper (SAM) is a physically-based spectral classification that uses an  $n$ -dimensional angle to match pixels to reference spectra. The algorithm determines the spectral similarity between two spectra by calculating the angle between the spectra, treating them as vectors in a space with dimensionality equal to the number of bands. This technique, when used on calibrated reflectance data, is relatively insensitive to illumination and albedo effects. Endmember spectra used by SAM can come from ASCII files, spectral libraries, or can be extracted directly from the image (as ROI average spectra). SAM compares the angle between the endmember spectrum vector and each pixel vector in  $n$ -dimensional space. Smaller angles represent closer matches to the reference spectrum. Pixels further away than the specified maximum angle threshold in radians are not classified.

Figure 3-31 illustrated that the main concept of SAM is to calculate the angle between endmember spectra and target spectra as vectors in a space with dimensionality equal to the number of bands (Shippert, 2003). The formula of spectral angle ( $\theta$ ) is calculated as:

$$\theta = \cos^{-1}\left(\frac{\sum_{\lambda=1}^M \rho_{\lambda} \rho'_{\lambda}}{L_{\rho} L_{\rho'}}\right) \quad (3-69)$$

where  $M$  is the number of spectral bands,  $\rho_\lambda$  is the reflectance of a target spectrum,  $\rho'_\lambda$  is the reflectance of a target spectrum,  $L_p$  is the length of the endmember vector, and  $L_{p'}$  is the length of the target spectrum vector.

The length of the endmember vector and the target spectrum vector are calculated as:

$$L_p = \sqrt{\sum_{\lambda=1}^M \rho_\lambda^2}, \quad L_{p'} = \sqrt{\sum_{\lambda=1}^M \rho_{\lambda'}^2} \quad (3-70)$$

After finding the spectral angle ( $\theta$ ), the spectral angle is compared with threshold value. If the threshold is below than the angle, the target spectrum is determined as classifying to the endmember class (Dennison, 2004).

### Spectral Information Divergence

While SAM is a deterministic method, SID is a probabilistic method that allows for variations in pixel measurements, where probability is measured from zero to a user-defined threshold (Du et al., 2004). Chang (1999) described the derivation of spectral information divergence (SID).

The hyperspectral pixel vector is given by

$$X = (x_1, x_2, x_3, \dots, x_n, \dots, x_N)^T \quad (3-71)$$

Each component  $x$  can be modeled as a random variable by defining an appropriate probability distribution. Due to the nature of reflectance, assume that all component  $x_i$ 's in  $X$  are nonnegative. Thus, the probability measure can be defined as:

$$p_j = \frac{x_j}{\sum_{n=1}^N x_n} \quad (3-72)$$

and the desired probability vector is  $P = \{p_i\}_{i=1}^L$ . For an information theory to capture relationship and correlation between two hyperspectral pixel vectors, another

pixel vector  $Y = (y_1, y_2, y_3, \dots, y_n, \dots, y_N)^T$  with the probability distribution given by  $Q = \{q_l\}_{l=1}^L$  and  $q_j = y_j / \sum_{n=1}^N y_n$ . Thus, spectral information divergence (SID) can be given by:

$$SID(X, Y) = D(X \parallel Y) + D(Y \parallel X) \quad (3-73)$$

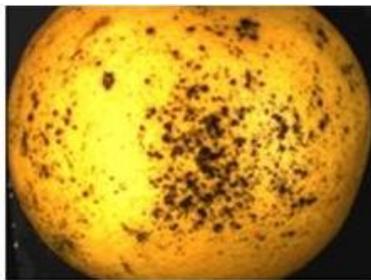
where  $(X \parallel Y)$ , called as the relative entropy of Y with respect to X, is defined by

$$D(X \parallel Y) = \sum_{l=1}^L p_l \log\left(\frac{p_l}{q_l}\right) \text{ and } D(Y \parallel X) = \sum_{l=1}^L q_l \log\left(\frac{q_l}{p_l}\right) \quad (3-74)$$

Therefore, the value of SID shows the use of the relative entropy and the similarity between two spectral pixels.



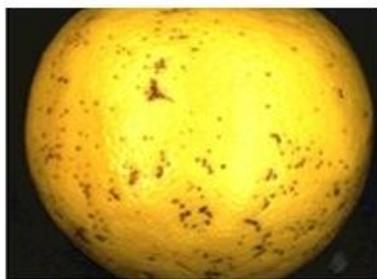
**Canker**



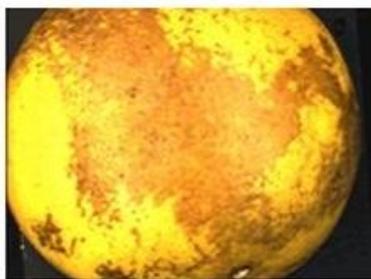
**Copper Burn**



**Greasy Spot**



**Melanose**

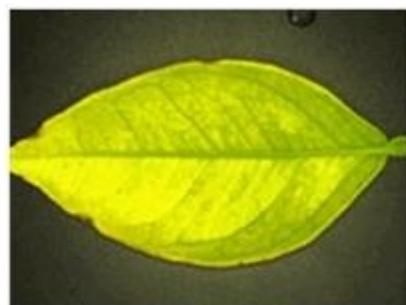


**Wind Scar**



**Market**

Figure 3-1. Images of abnormal peel conditions



**Iron Deficiency**



**Zinc Deficiency**



**Manganese Deficiency**

Figure 3-2. Images of nutritional deficiency



**Blotchy Mottle**

**Greening Islands**

Figure 3-3. Images of citrus greening



**Young Flush**

**Normal Mature**

Figure 3-4. Images of normal conditions



Figure 3-5. 15 images of blotch mottle conditions



Black Spot



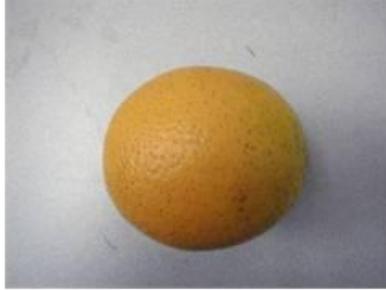
Greasy Spot



Melanose



Wind Scar



Market

Figure 3-6. Representative images for each peel condition

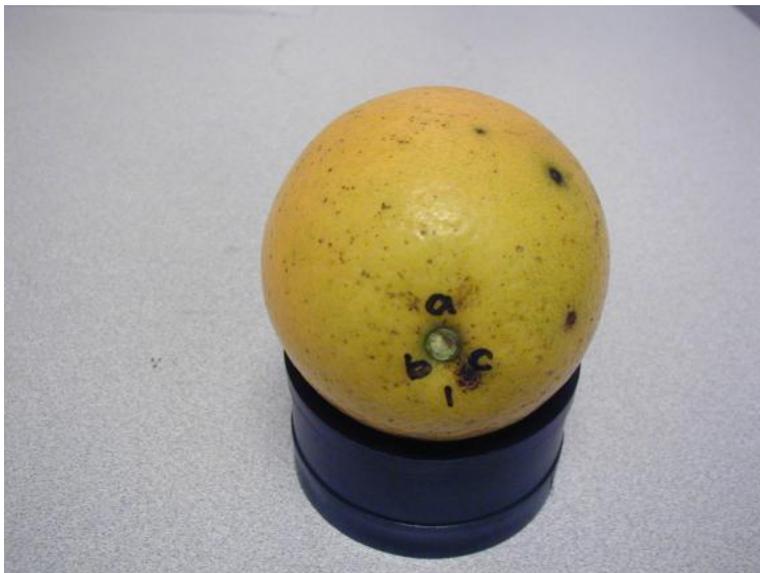


Figure 3-7. Three faces of each fruit with 120° interval

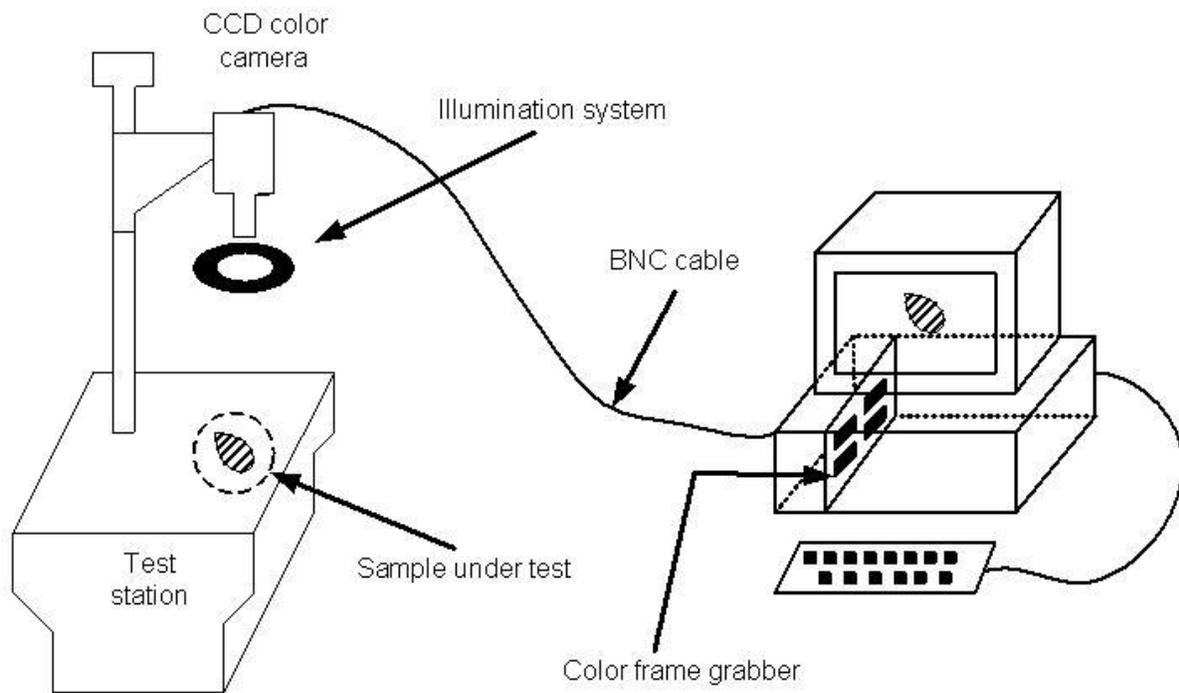


Figure 3-8. Typical image system for acquiring RGB images from citrus samples



Figure 3-9. Digital microscope system for acquiring RGB images from citrus samples

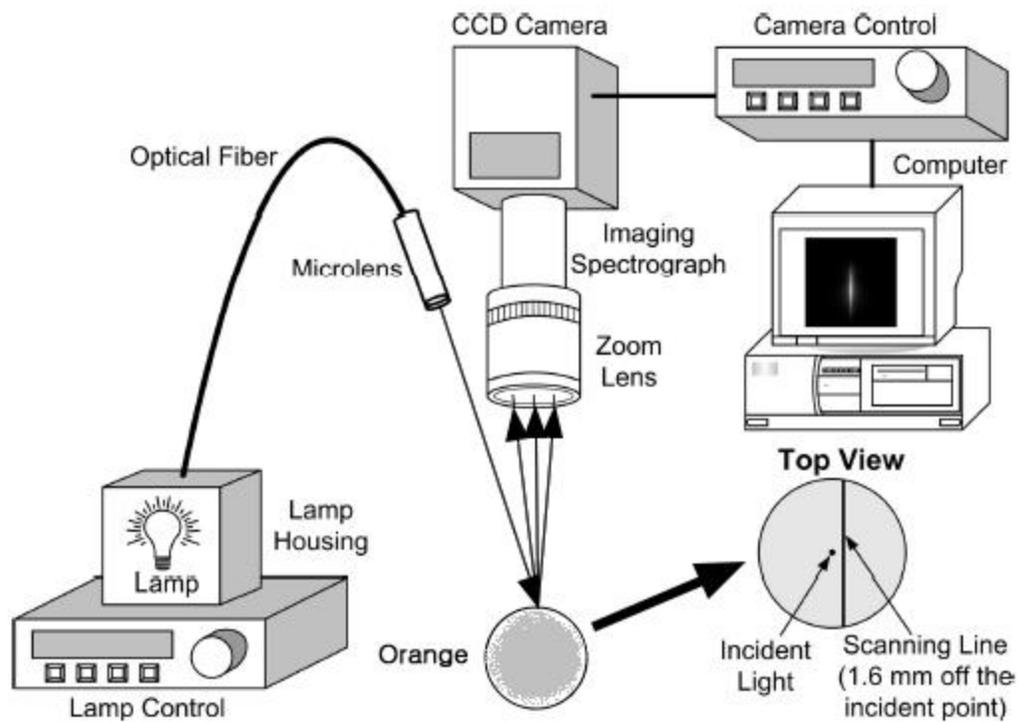


Figure 3-10. Typical hyperspectral line-scan imaging system

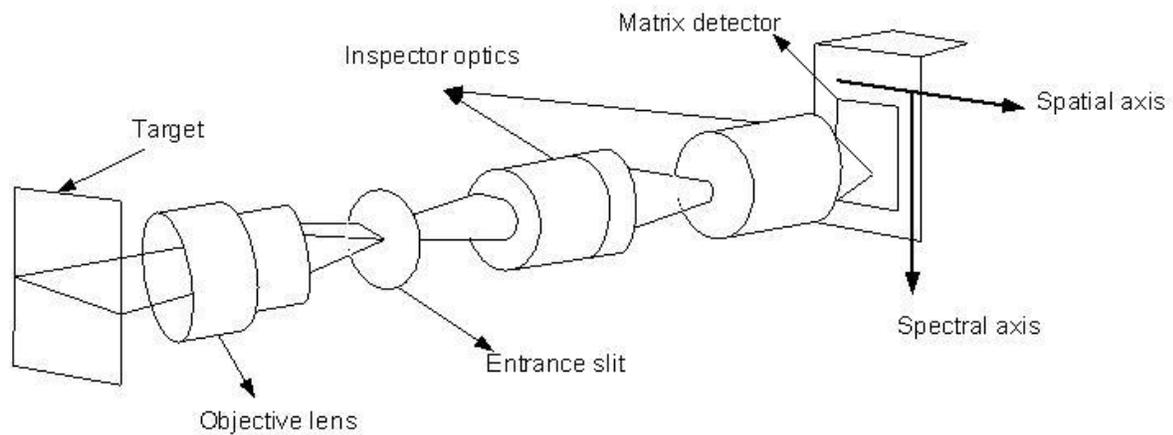


Figure 3-11. Principle of the prism-grating-prism imaging spectrograph for acquiring spatial and spectral information from an object

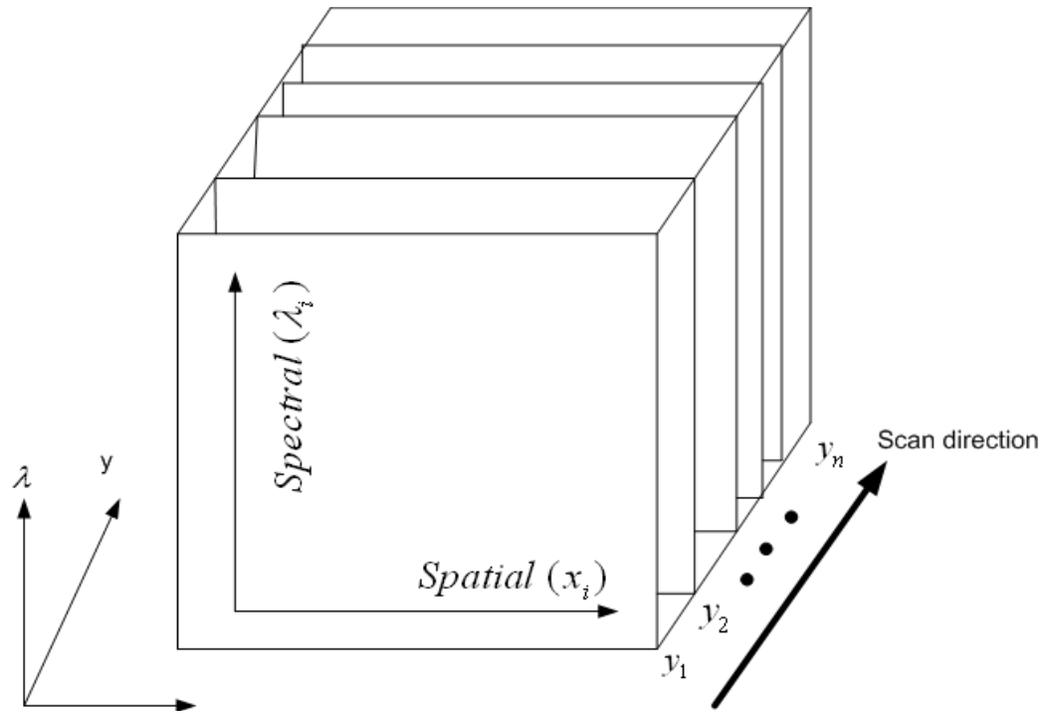


Figure 3-12. Conceptual representation of a volume of hyperspectral image data

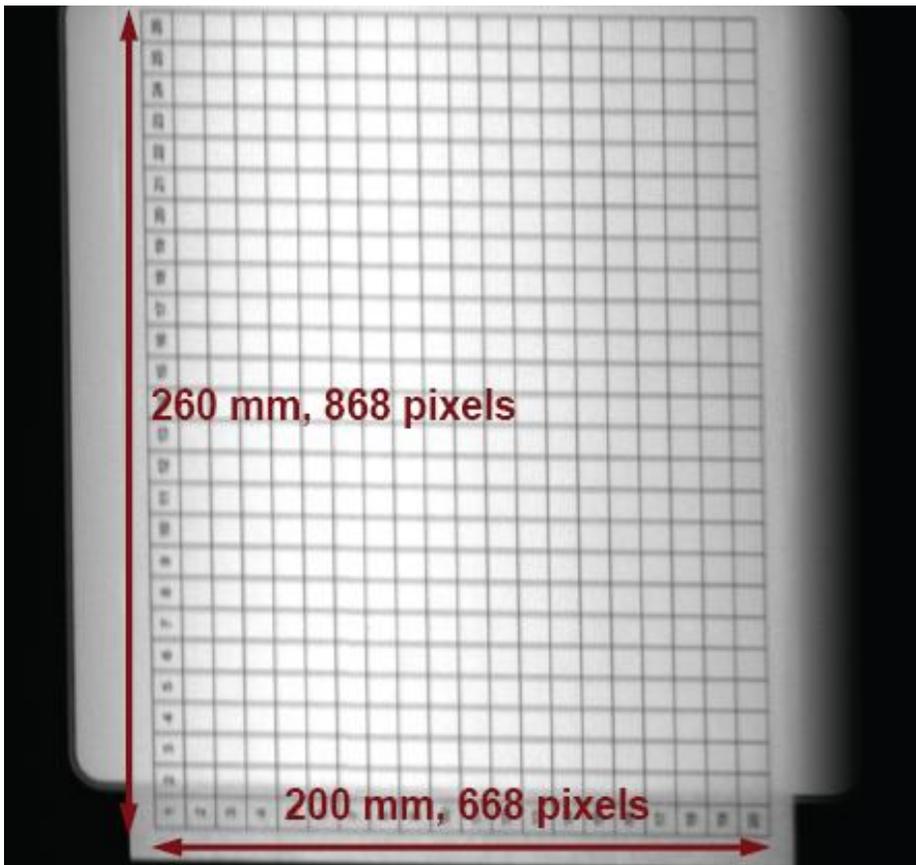


Figure 3-13. White paper printed with thin parallel lines 2 mm apart for geometric calibration

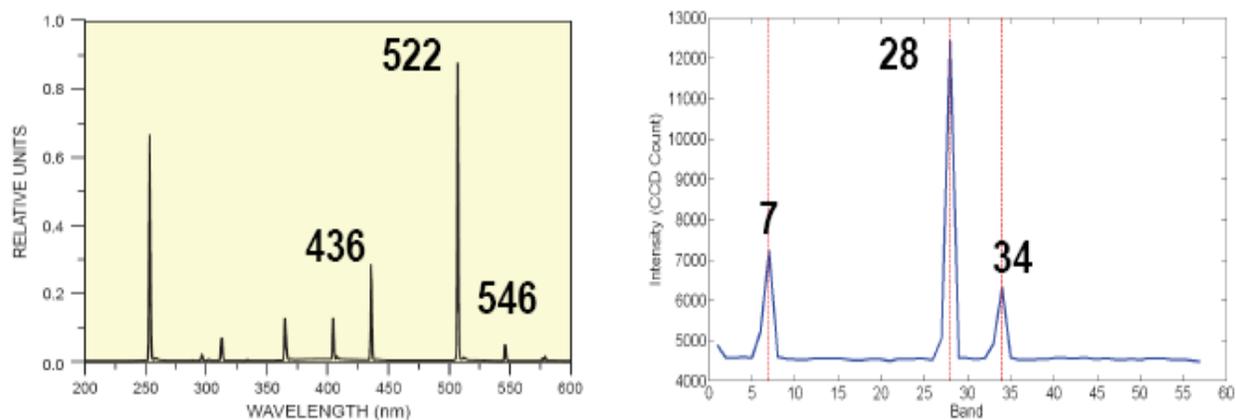


Figure 3-14. Spectral profiles of calibration lamps: xenon lamp and mercury-argon lamp

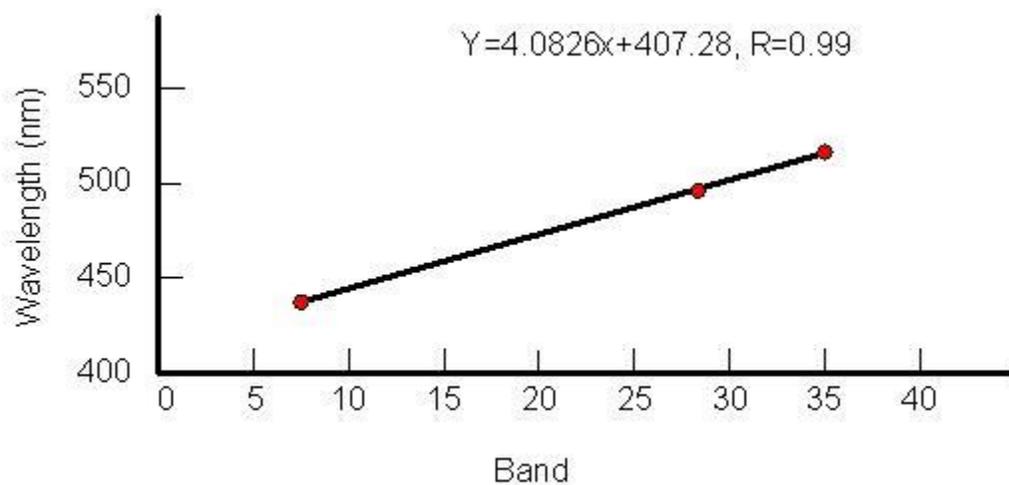


Figure 3-15. The relationship between the vertical band position and wavelength from two calibration lamps

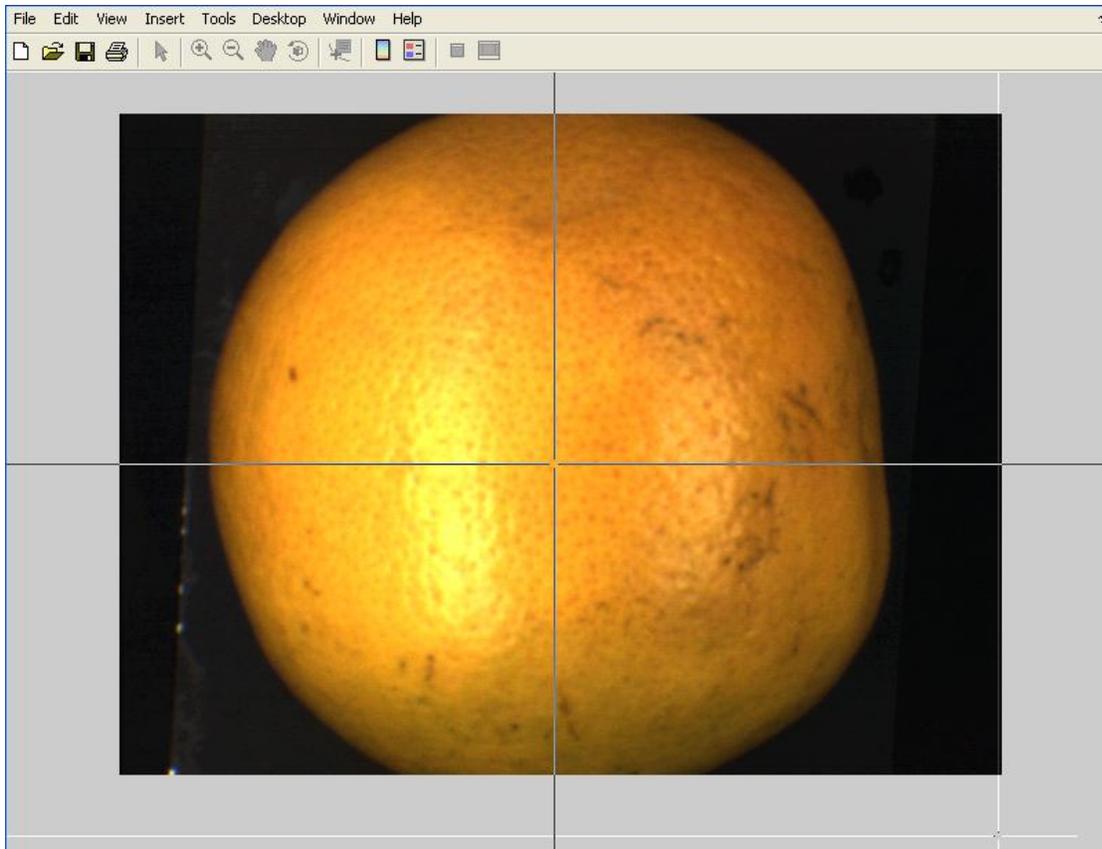


Figure 3-16. ROI selection program developed in MATLAB R2007b

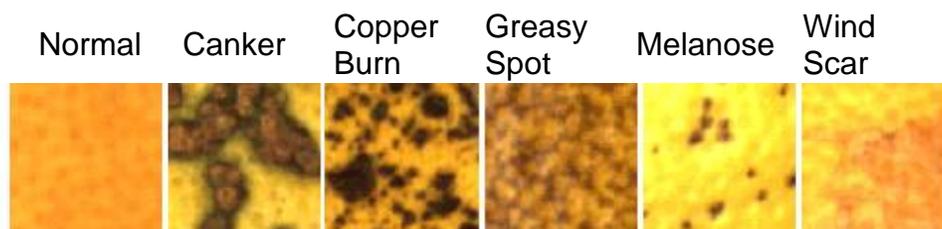


Figure 3-17. Typical ROI images for normal and diseased citrus peel conditions

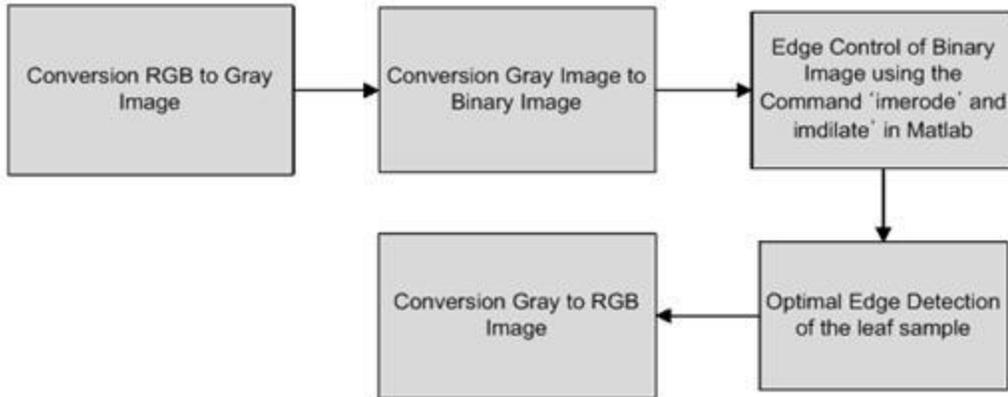


Figure 3-18. Procedures for edge detection

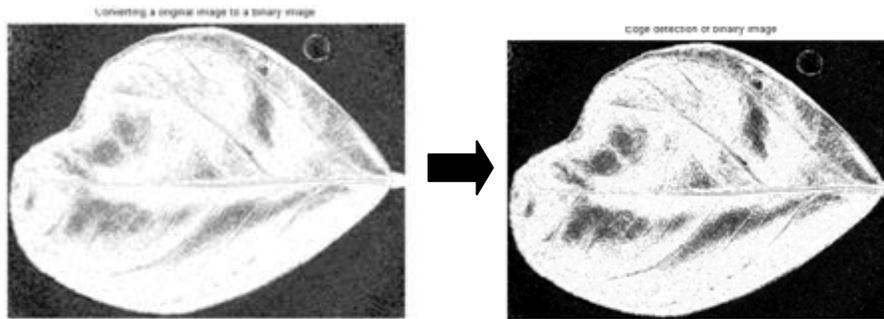


Figure 3-19. Converted images of a leaf samples

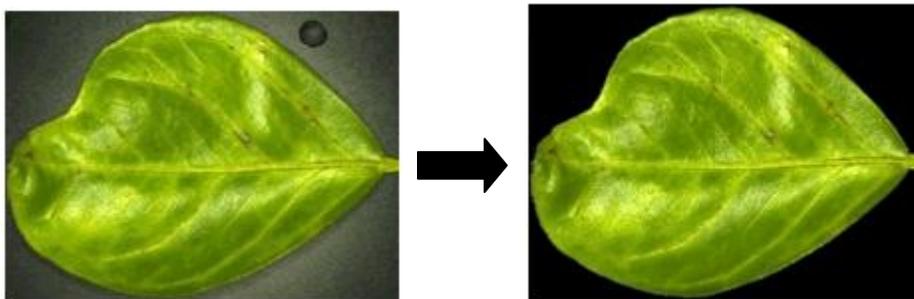


Figure 3-20. Edge detected image of a leaf samples

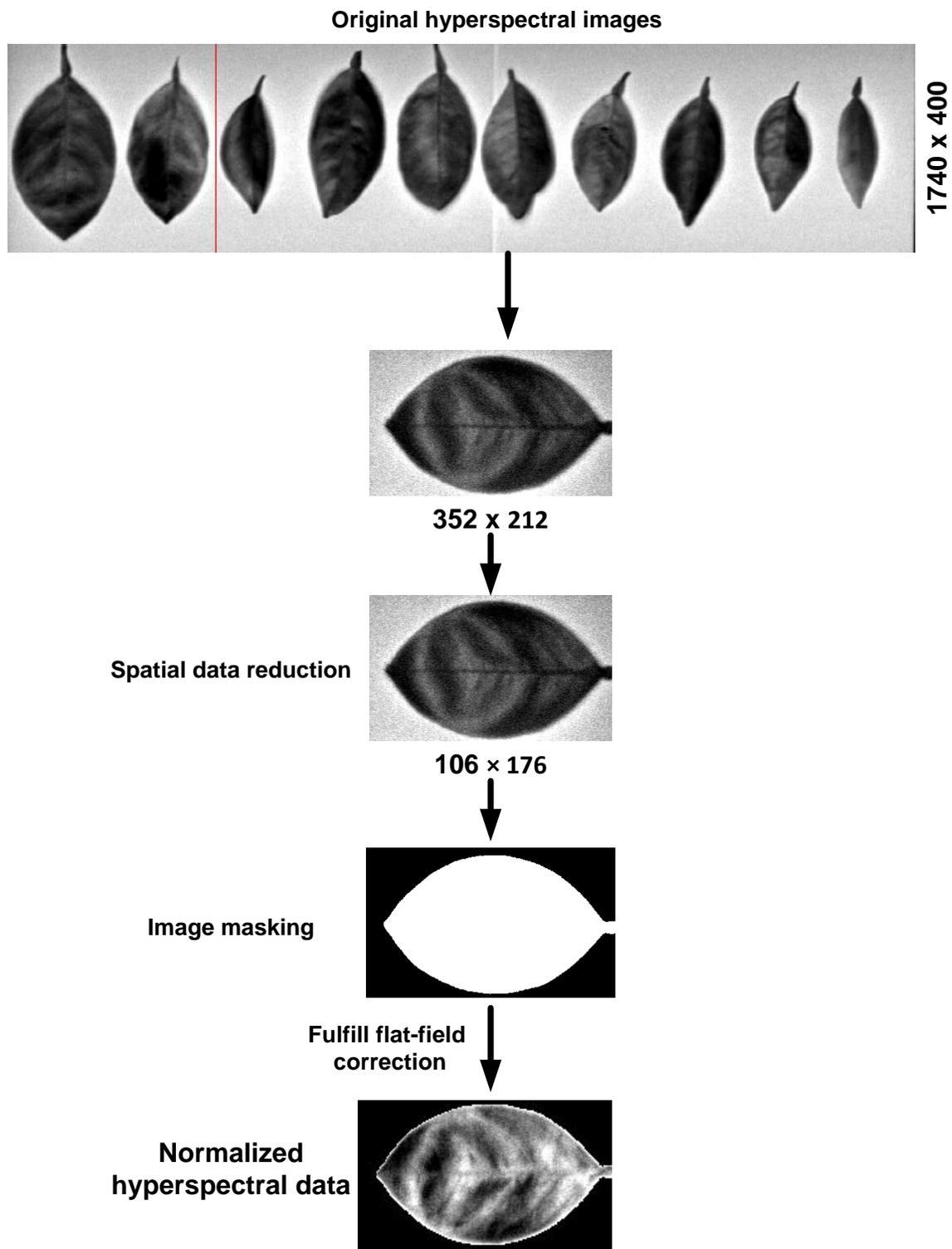


Figure 3-21. Hyperspectral image pre-processing

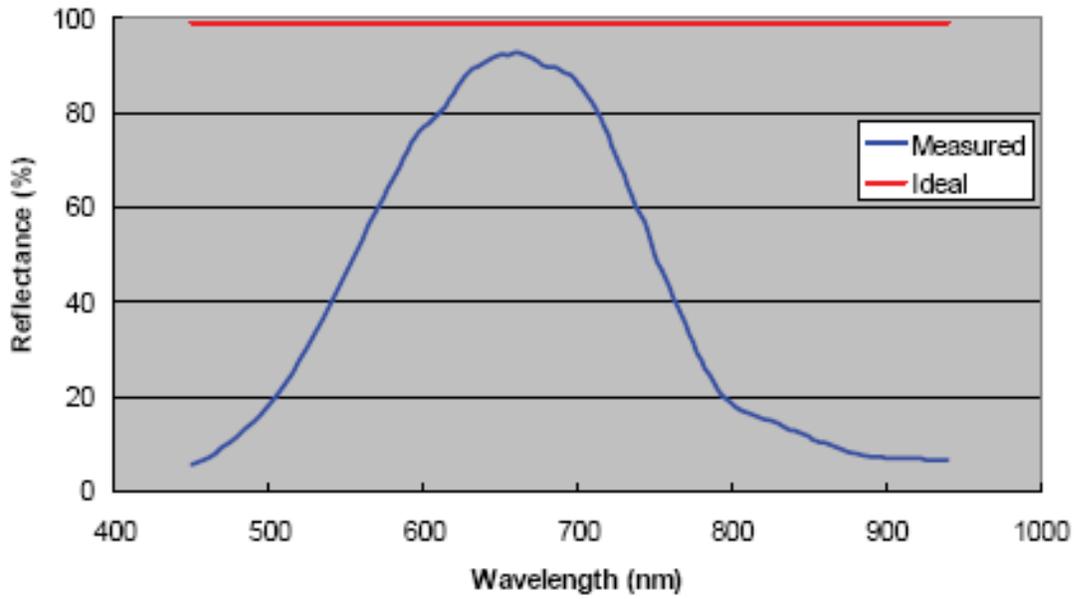


Figure 3-22. The plot of reflectance factor for white panel

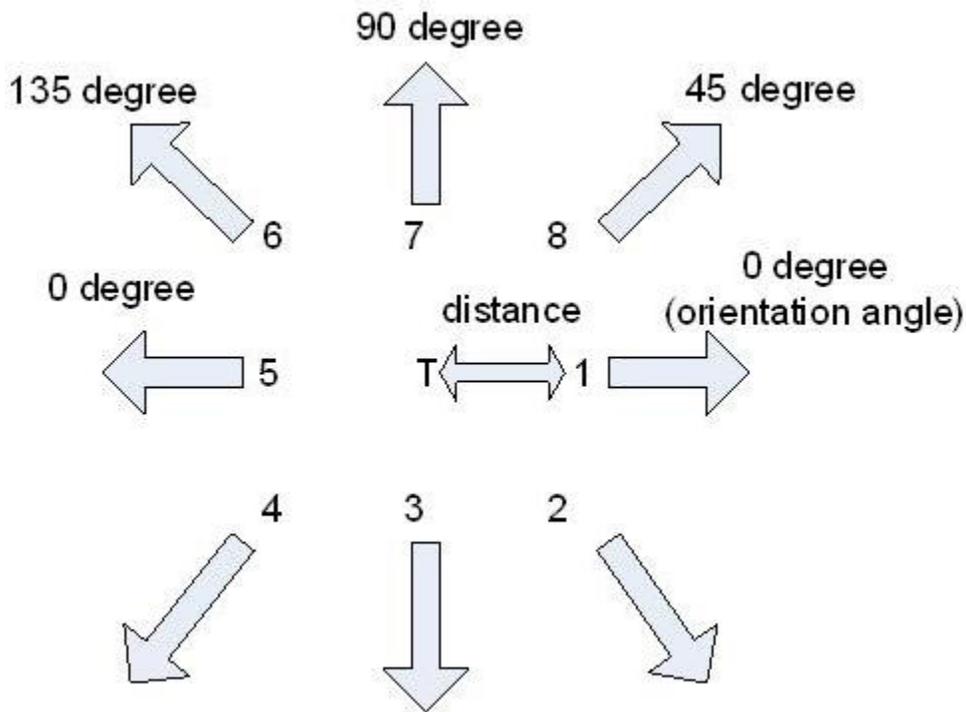


Figure 3-23. Nearest neighbor diagram

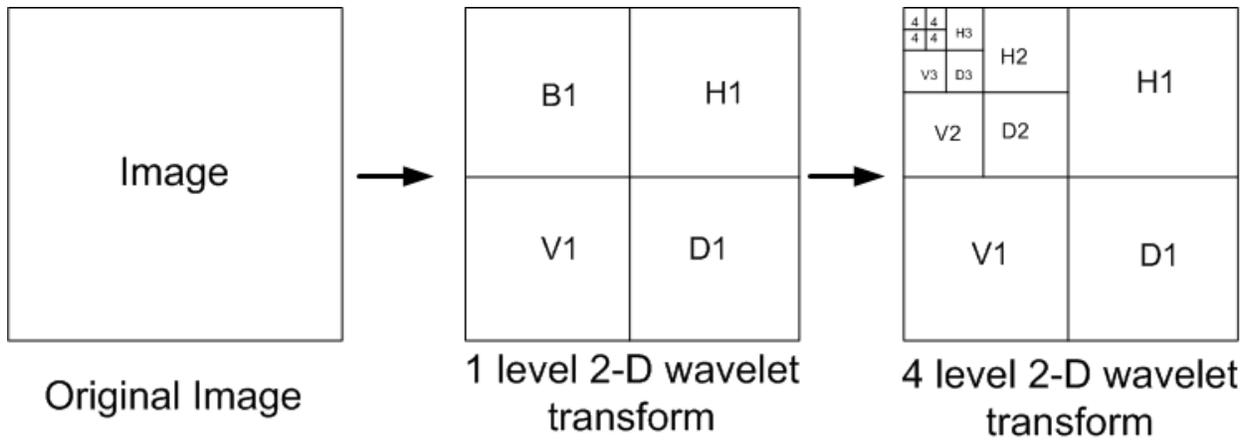


Figure 3-24. Decomposition of an image

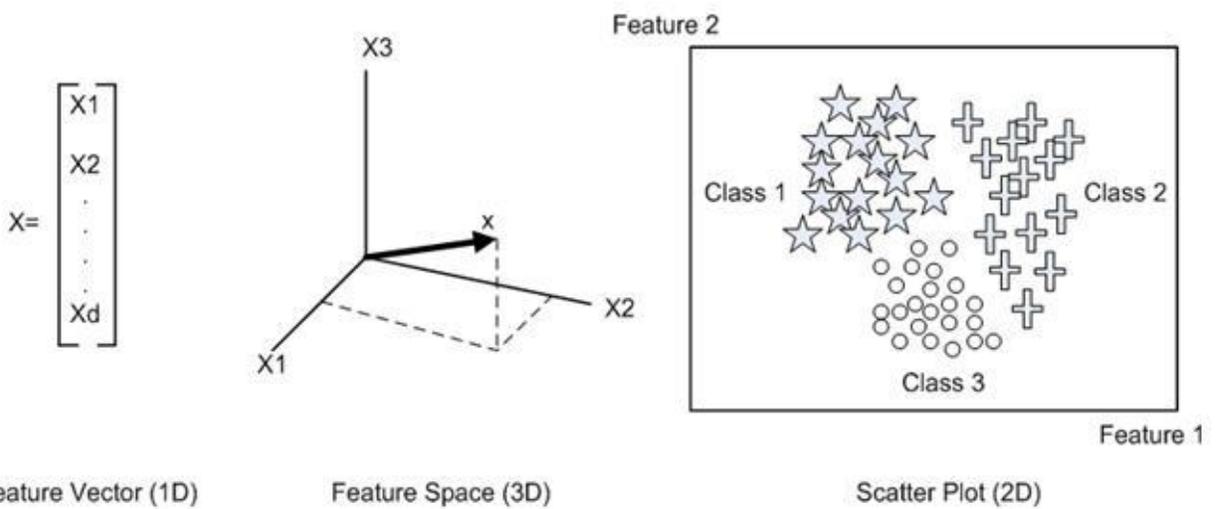
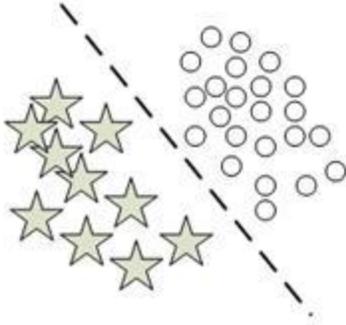


Figure 3-25. Feature vector, feature space and scatter plot

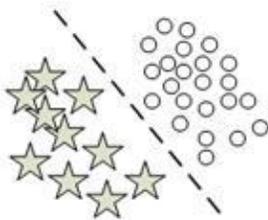


A example of good feature



A example of bad feature

Figure 3-26. Good feature and bad feature



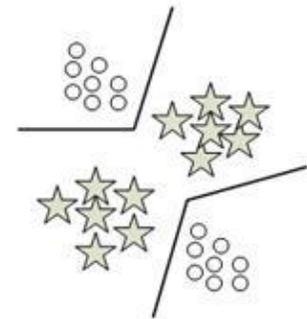
1) Linear distribution type



2) Nonlinear distribution type



3) High correlation distribution type



4) Multi-class distribution type

Figure 3-27. Distribution plots with Pattern types

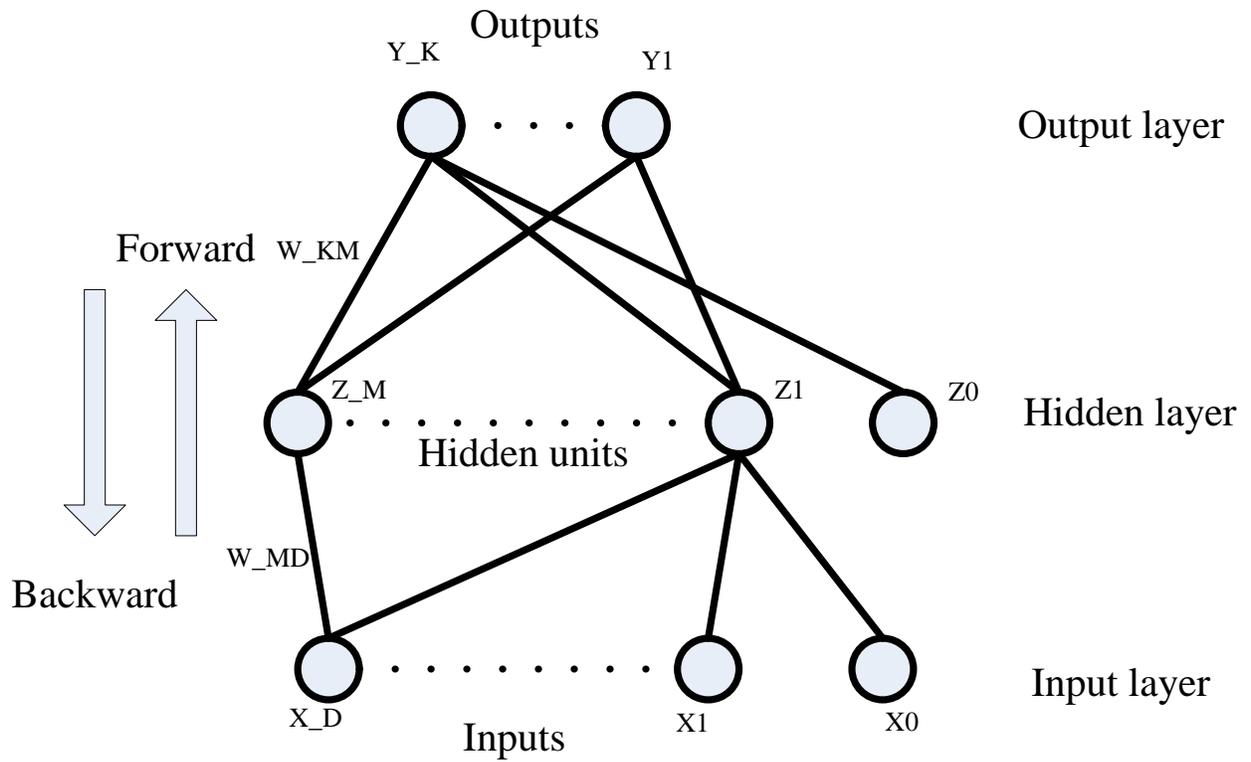


Figure 3-28. The flow of a network

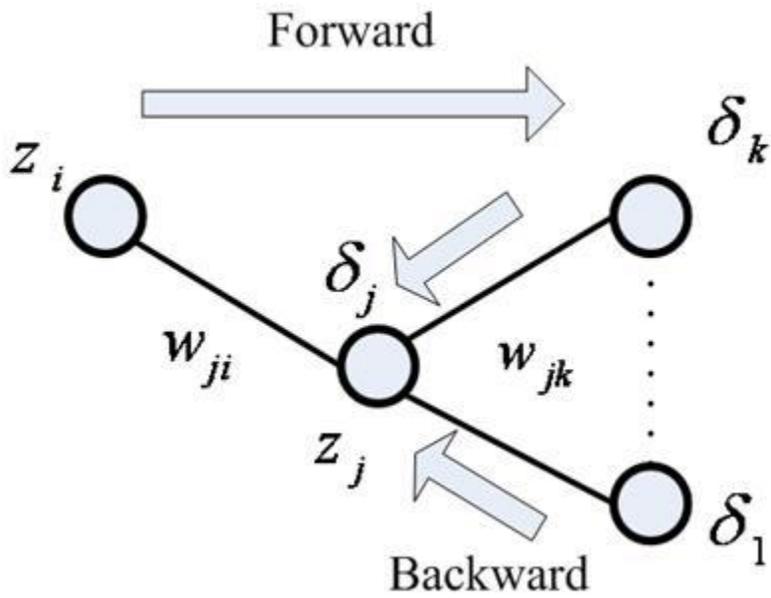


Figure 3-29. Illustration of the calculation

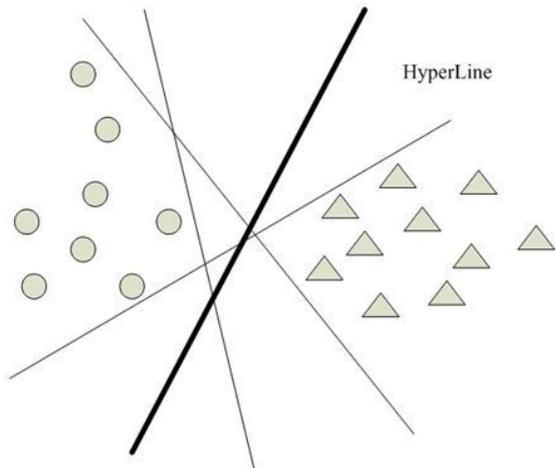


Figure 3-30. Possible linear classifiers for separating the data

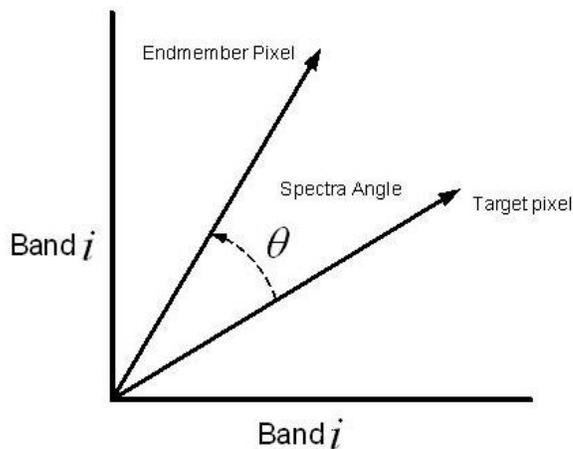


Figure 3-31. The angle between endmember spectra and target spectra as vectors in a space

Table 3-1. The number of black spot and other symptoms

Class	Symptoms	Number	Total
Black spot	Black spot	135	100
No black spot	Greasy spot	90	390
	Market	90	
	Melanose	105	
	Wind scar	105	

## CHAPTER 4 DETECTION OF CITRUS CANKER DISEASE

Citrus trees can exhibit a host of symptoms reflecting various disorders that can adversely impact their health, vigor, and productivity to varying degrees. In some cases, disease control actions or remedial measures can be undertaken if the symptoms are identified early. Additional opportunities for disease control exist when precision agriculture techniques are involved, which could use early detection along with a global positioning system to map diseases in the grove for future control actions.

Environmental pollution is another concern throughout the world. Indiscriminate use of fungicides, pesticides, and herbicides to control pest and diseases has led to problems, such as deteriorating ground water quality, and health hazards for operators and the general public. Increased pressures to reduce chemical applications have led researchers to study new ways for early detection of various diseases on citrus trees with an aim to reduce chemical usage and maintain cost-effective crop production. This study explored machine vision based techniques that can visually differentiate common citrus peel disorders using individual fruit color texture features. Citrus samples were collected in the field and evaluated under laboratory conditions. Future studies will expand the technologies to in-field inspections.

In the past decade, researchers have used image processing and pattern recognition techniques in agricultural applications, such as detection of weeds in the field, and sorting of fruits and vegetables. The underlying approach for all of these techniques is the same. First, images are acquired from the environment using analog, digital, or video cameras. Then, image processing techniques are applied to extract useful features that are necessary for further analysis of the images. Afterwards,

discriminant techniques, such as parametric or non-parametric statistical classifiers and neural networks, are employed to classify the images. The selection of the image processing techniques and the classification strategies are important for the successful implementation of any machine vision system. Object shape matching functions, color-based classifiers, reflectance-based classifiers, and texture-based classifiers are some of the common methods that have been tried in the past.

A number of techniques have been studied to detect defects and diseases related to citrus. Gaffney (1973) obtained reflectance spectra of citrus fruit and some surface defects. Edwards and Sweet (1986) developed a method to assess damages due to citrus blight disease on citrus plants using reflectance spectra of the entire tree. Miller and Drouillard (2001) collected data from Florida grapefruit, orange, and tangerine varieties using a color vision system. They used various neural network classification strategies to detect blemish-related features for the citrus fruit. Aleixos et al. (2002) developed a multispectral camera system that could acquire visible and near infrared images from the same scene, and used it on a real-time system for detecting defects on citrus surface. Blasco et al. (2007) reported the application of near-infrared, ultraviolet and fluorescence computer vision systems to identify the common defects of citrus fruit. They proposed a fruit sorting algorithm that combines the different spectral information to classify fruit according to the type of defect. Their results showed that non-visible information can improve the identification of some defects. Most recently, Qin et al. (2008) developed an approach for citrus canker detection using hyperspectral reflectance imaging and PCA-based image classification method. Their results

demonstrated that hyperspectral imaging technique could be used for discriminating citrus canker from other confounding diseases.

This research was aimed to develop a method to detect citrus peel diseases using color texture features. The use of color texture features in classical gray image texture analysis was first reported by Shearer (1986). Shearer and Holmes (1990) reported a study for classifying different types of nursery stock by the color co-occurrence method (CCM). This method had the ability to discriminate between multiple canopy species and was insensitive to leaf scale and orientation. The use of color features in the visible light spectrum provided additional image characteristic features over traditional gray-scale texture representation. The textural methods employed were statistical-based algorithms that measured image features, such as smoothness, coarseness, graininess, and so on. The CCM method involves three major mathematical processes briefly described in the following. A complete discussion of the color co-occurrence method could be found in Shearer and Holmes (1990).

- Transformation of a red, green, blue (RGB) color representation of an image to an equivalent hue, saturation, and intensity (HSI) color representation;
- Generation of color co-occurrence matrices from the HSI pixel maps. Each HSI matrix is used to generate a spatial gray-level dependence matrix (SGDM) providing three SGDM's.
- Calculation of texture features from the three SGDM's.

Burks et al. (2000) developed a method for weed species classification using color texture features and discriminant analysis. In their study, CCM texture feature data models for six classes of ground cover (giant foxtails, crabgrass, velvet leaf, lambs quarter, ivy leaf morning glory, and soil) were developed and stepwise discriminant analysis techniques were utilized to identify combinations of CCM texture feature

variables, which have the highest classification accuracy with the least number of texture variables. A discriminant classifier was trained to identify weeds using the models generated. Classification tests were conducted with each model to determine their potential for classifying weed species. Pydipati et al. (2006) utilized the color co-occurrence method to extract various textural features from the color RGB images of citrus leaves. The CCM texture statistics were used to identify three diseased conditions and normal citrus leaves using discriminant analysis.

The overall objective of this research was to develop a machine vision based method for detecting various diseases on citrus peel using color texture features under a controlled lighting condition. Specific objectives implemented to accomplish the overall objective were to:

- Use a color imaging system to collect RGB images from grapefruits with normal and five peel conditions (i.e., canker, copper burn, greasy spot, melanose, and wind scar);
- Determine image texture features based on the color co-occurrence method (CCM); and
- Develop algorithms for selecting useful texture features and classifying the citrus peel conditions based on the reduced texture feature sets.

## **Materials and Methods**

### **Image Acquisition System**

A color image acquisition system was assembled for acquiring RGB images from citrus samples, and it is shown in Figure 4-1. The imaging system consisted of two 13 W high frequency sealed fluorescent lights (SL Series, StockerYale, Salem, NH, USA), a zoom lens (Zoom 7000, Navitar, Rochester, NY, USA), a 3-CCD RGB color camera (CV-M90, JAI, San Jose, CA, USA), a 24-bit color frame grabber board with 480×640 pixel resolution (PC-RGB, Coreco Imaging, St. Laurent, Quebec, CA), and a computer

installed with an image capture software. The setup of the lighting system was designed to minimize specular reflectance and shadow, and to maximize the contrast of the images. The height of the camera and its focus were adjusted to contain the image of the whole fruit, with an approximate 100 mm×100 mm field of view. Automatic white balance calibration was conducted using a calibrated white balance card before acquiring images from fruit samples. The digital color images were saved in uncompressed BMP format.

### **Image Analysis**

The data analysis methods for analyzing the color images of the fruit samples based on the color co-occurrence method (CCM) are illustrated in the flow chart shown in Figure 4-2, which involve the procedures for selection of region of interest (ROI), transformation from RGB format to HSI format, generation of spatial gray-level dependence matrices (SGDM's), calculation of texture features, selection of useful texture features, and discriminant analysis for disease classification. All image processing and data analysis procedures were executed using programs developed in MATLAB 7.0 (MathWorks, Natick, MA, USA) and SAS 9.1 (SAS Institute Inc., Cary, NC, USA).

The color co-occurrence texture analysis method was developed through the use of the spatial gray-level dependence matrices (SGDM's). The SGDM's were generated for each color pixel map of the ROI HSI images, one each for hue, saturation and intensity. These matrices measure the probability that a pixel at one particular gray-level will occur at a distinct distance and orientation from any pixel given that pixel has a second particular gray-level (Shearer and Holmes, 1990). The SGDM is represented by the function  $P(i, j, d, \theta)$  where  $i$  represents the gray-level of the pixel at  $(x_1, y_1)$  in the

image, and  $j$  represents the gray-level of the pixel at  $(x_2, y_2)$  located at a distance  $d$  and an orientation angle  $\theta$  from  $(x_1, y_1)$  (Shearer, 1986). The matrix is constructed by counting the number of pixel pairs of  $(x_1, y_1)$  and  $(x_2, y_2)$  with the grey value  $i$  and  $j$  at distance  $d$  and direction  $\theta$ .

The SGDM's generated for hue, saturation and intensity were then used to calculate the texture features. Shearer and Holmes (1990) reported a reduction for the 16 gray scale texture features through elimination of redundant variables, resulting in 11 texture features. Donohue et al. (2001) added two more texture features (i.e., image contrast and modus) to those used by Shearer and Holmes (1990). In this study, the combined 13 texture features proposed by Shearer and Holmes (1990) and Donohue et al. (2001) were used for citrus peel disease classification, and they included (1) uniformity, (2) mean intensity, (3) variance, (4) correlation, (5) product moment, (6) inverse difference, (7) entropy, (8) sum entropy, (9) difference entropy, (10) information correlation #1, (11) information correlation #2, (12) contrast, and (13) modus. The equations for calculating the 13 texture features can be found in Pydipati et al. (2006).

The calculations were performed for each of the three SGDM's, producing 13 texture features for each HSI component and thereby a total of 39 texture statistics. The texture features were identified by a coded variable name where the first letter represents whether it is a hue (H), saturation (S) or intensity (I) feature and the number following represents one of the thirteen texture features described above.

As an example, the feature  $(I_7)$  is a measure of the entropy in the intensity CCM matrix, which represents the amount of order in an image and is calculated by equation (4-1).

$$I_7 = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p(i,j) \ln p(i,j) \quad (4-1)$$

The  $p(i,j)$  matrix represents the normalized intensity co-occurrence matrix and  $N_g$  represents the total number of intensity levels. The equation for normalizing the co-occurrence matrix is given in equation (4-2), where  $P(i,j,1,0)$  is the intensity co-occurrence matrix.

$$p(i,j) = \frac{P(i,j,1,0)}{\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P(i,j,1,0)} \quad (4-2)$$

A physical representation of entropy (uncertainty) may be visualized by comparing a checkerboard-like image to an image where one half is black and the other half is white. The latter image is highly ordered having all pixels of the same intensity segregated into two distinct pixels groups, which gives greater certainty of the pixel value of the adjacent pixels. The checkerboard image has a lower amount of order due to intermixing of black and white squares, which results in a greater level of uncertainty of neighboring pixel values. The lower order image would therefore have more uncertainty and thus a higher entropy measure.

After the texture statistics were obtained for each image, feature selection was conducted to reduce the redundancy in the texture feature set. The SAS procedure STEPDISC can reduce the size of the variable set and find the variables that are important for discriminating samples in different classes, and it was used for the texture feature selection. The stepwise discriminant analysis begins with no variables in the classification model. At each step of the process, the variables within and outside the model are evaluated. The variable within the model, at that particular step, which

contributes least to the model as determined by the Wilks' Lambda method is removed from the model. Likewise, the variable outside the model that contributes most to the model and passes the test to be admitted is added. A test significant level of 0.0001 for the variables of SLS (test for variable to stay) and the SLE (test for variable to enter) in the STEPDISC procedure was chosen for the stepwise discrimination of the variable list (SAS, 2004). When no more steps can be taken, the number of variables in the model is reduced to its final form. Burks et al. (2000) had shown that classification performances were poor if only hue or saturation information was used in the classification models. Thus three color feature combinations including hue, saturation, and intensity (H, S, I), hue and saturation (H, S), and intensity (I) only were used to perform the texture feature selections. These three color combinations have demonstrated high classification accuracies in the applications for other plant discriminations (Burks et al. 2000; Pydipati et al. 2006).

### **Texture Classification**

The classification models were developed using the SAS procedure DISCRIM, which creates a discriminant function based on a measure of the generalized squared distance between a specific test image texture variable input set and the class texture variable means, with an additional criteria being the posterior probability of the classification groups (Rao, 1973). Each sample in the testing set was placed in the class for which it had the smallest generalized square distance between the test observation and the selected class, or the largest posterior probability of being in the selected class. The DISCRIM procedure utilized a likelihood ratio test for homogeneity of the within-group covariance matrices at a 0.1 test significance level.

The 30 samples from each peel condition were divided into two datasets consisting of 20 samples for training and 10 samples for testing. The samples were first arranged in ascending order for the time the images were acquired. The first two samples were selected for training and the third sample for testing. This approach minimizes negative time dependent variability, and reduces potential for data selection bias between the training and test datasets. A training data set and a test data set were created for each of the subsets of the texture features selected by the stepwise discriminant analysis described above. The training sets were used to train the classification models and the testing sets were used to evaluate the accuracies of different classification models.

## **Results and Discussion**

### **Selection of Texture Features**

The texture feature selection results are summarized in Table 4-1. Four classification models were developed using the selected texture feature sets from the three color combinations [(H, S, I), (H, S), and (I)]. The variables listed in the column of “Texture Feature Set” in Table 4-1 were generated by the SAS STEPDISC procedure, and they were arranged in the descending order of the importance for the classification models. The subscript numbers indicate the texture statistics as the following: (1) uniformity, (2) mean intensity, (3) variance, (4) correlation, (5) product moment, (6) inverse difference, (7) entropy, (8) sum entropy, (9) difference entropy, (10) information correlation #1, (11) information correlation #2, (12) contrast, and (13) modus. As an example, H<sub>9</sub> represents the difference entropy of hue, and it is selected as the most important texture feature for the first two classification models developed using two different color combinations [(H, S, I), and (H, S)].

The classification models were named using the color features involved in the texture feature selections followed by the total numbers of the selected texture features. For example, model HSI\_13 consists of a reduced set of hue, saturation and intensity texture features, and there are 13 texture features in total that were used to construct the model. As shown in Table 4-1, significant eliminations of redundant texture features were accomplished through the stepwise discriminant analysis. Nine and eleven texture features were selected for model HS\_9 and model I\_11, respectively. The simplification of the texture features largely reduces the computation burden due to the redundant data, and it also helps improve the performance of classification models. In addition to the three models described above, a classification model that used all 39 HSI texture features was developed for the purpose of comparisons with other models. Thus there are four classification models that were used to differentiate the citrus peel diseases, and they were independently evaluated for classification performance.

### **Classification of Citrus Peel Conditions**

The SAS procedure DISCRIM was used to test the accuracies of the classification models. Table 4-2 summarizes the classification results for differentiating different citrus peel conditions using model HSI\_13 listed in Table 4-1. As shown in Table 4-2, four peel conditions (normal, canker, copper burn, and wind scar) among the total six conditions tested in this study were perfectly classified into the appropriate categories. For the other two conditions (greasy spot and melanose), there was one misclassified sample for each case. One greasy spot sample was misclassified as copper burn, and one melanose sample was misclassified as wind scar. The classification accuracies for greasy spot and melanose were 90.0%. In general, there were only two samples that

were misclassified in the 60 samples in the testing set, and the overall classification accuracy for the model HSI\_13 was 96.7%.

Same procedures were applied for the other three classification models listed in Table 4-1, and the classification results, along with those from the model HSI\_13, are summarized in Table 4-3. Using nine selected hue and saturation texture features, model HS\_9 provided the classification accuracies of 90.0% for normal, canker, copper burn, greasy spot, and wind scar, and 70.0% accuracy for melanose. The average accuracy of the model HS\_9 was 86.7%. Model I\_11 used 11 selected intensity texture features alone. Although it achieved two perfect classification results (100.0%) for copper burn and greasy spot, the performances for the other four conditions were poor, especially for melanose (70.0%) and wind scar (50.0%). The overall accuracy of the model I\_11 was 81.70%, which is the worst among the four models tested in this study. When all 39 texture features were used by model HSI\_39, the classification accuracy was achieved as 88.3%, which was higher than those of the models HS\_9 and I\_11, but lower than that of the model HSI\_13 (96.7%).

Based on the results shown in Table 4-3, we could find that classification model using intensity texture features only (model I\_11) gave the worst performance when compared to the other models. It is likely that the poor performance of the intensity texture features only is due to the variations of the light intensity during the image acquisition. On the other hand, classification model that used hue and saturation texture features (model HS\_9) outperforms the model that used intensity texture features only (model I\_11). When intensity texture features were added to hue and saturation features, the performance of the classification models was further improved. Texture

feature selection is necessary for obtaining better classification accuracy, and this is confirmed by the fact that the model using 13 selected hue, saturation and intensity texture features (model HSI\_13) achieved better accuracy than the one that using all 39 HSI texture features (model HSI\_39). The model HSI\_13 emerged as the best one among the four classification models tested in this study, suggesting that it would be best to use a reduced hue, saturation and intensity texture feature set to differentiate different citrus peel conditions.

### **Stability Test of the Classification Model**

It is important to test high classification accuracy using various texture features because this classification results presented in above section is established statistically. Moreover, if such stability can be demonstrated by this test, this image analysis method and procedures will be more useful for detecting purpose in predicting citrus disease. Image samples had a fixed order (i.e., one from every thirty samples arranged in ascending order for the time the images were acquired). Using the order, training samples and testing samples were separated. To test the stability of the classification model, 20 training samples and 10 testing samples were randomly chosen from the 30 samples for each peel condition, and they were used to train and test the model HSI\_14, which gave the best classification performance, following the same procedures described earlier. Ten runs were repeated for the training and testing. The average value and standard deviation in Table 4-4 were 96.0% and 2.3%, respectively. These results indicate that the classification model developed using the 14 selected hue, saturation and intensity texture features is robust in performance, and thus is able to classify the new fruit samples according to their peel conditions.

## Summary

Color imaging coupled with texture feature analysis based on color co-occurrence method provides a useful means for identifying common diseases on citrus fruit. A color imaging system was assembled to acquire RGB images from grapefruits with normal and five peel disorders including canker, copper burn, greasy spot, melanose, and wind scar. Small images covering the interested areas on the fruit surface were extracted from the original RGB images, and they were then transformed to hue, saturation, and intensity color representation. Spatial gray-level dependence matrices were generated from the hue, saturation, and intensity images. A total of 39 image texture features were determined from the SGDM's for each fruit sample. Algorithms for selecting useful texture features were developed based on a stepwise discriminant analysis for three color combinations including hue, saturation, and intensity (HSI), hue and saturation (HS), and intensity (I). Classification models were constructed using the reduced texture feature sets through a discriminant function based on a measure of the generalized squared distance.

Significant eliminations of redundant texture features were accomplished through the stepwise discriminant analysis. 13, 9, and 11 texture features were selected for the color combinations of HSI, HS, and I, respectively. The simplification of the texture features largely reduces the computation burden, and it also helps improve the performance of classification models. The classification model using intensity texture features only gave the worst accuracy (81.7%), and the model using 13 selected HSI texture features achieved the best classification accuracy (96.7%) among four classification models including the one using all 39 HSI texture features. The results suggested that it would be best to use a reduced hue, saturation and intensity texture

feature set to differentiate different citrus peel conditions. A stability test for the classification model with the best performance was accomplished by 10 runs using randomly selected training and testing samples. Average classification accuracy and standard deviation in Table 4-4 were 96.0% and 2.3%, respectively, indicating that the classification model is robust for classifying new fruit samples according to their peel conditions.

This research demonstrated that color imaging and texture feature analysis could be used for differentiating citrus peel diseases under the controlled laboratory lighting conditions. Future studies will explore the utility of these algorithms in outdoor conditions, and develop pattern recognition methods such as self-organizing map (SOM) or support vector machines (SVM) for real-time application. The most significant challenge will be created by the inherent variability of color under natural lighting conditions. By eliminating intensity-based texture features, this variability can be significantly reduced. However, hue and saturation can be somewhat influenced by low lighting conditions. This may point to the need to use cameras with light availability color compensation, supplemental lighting, or night time applications where lighting levels can be controlled.

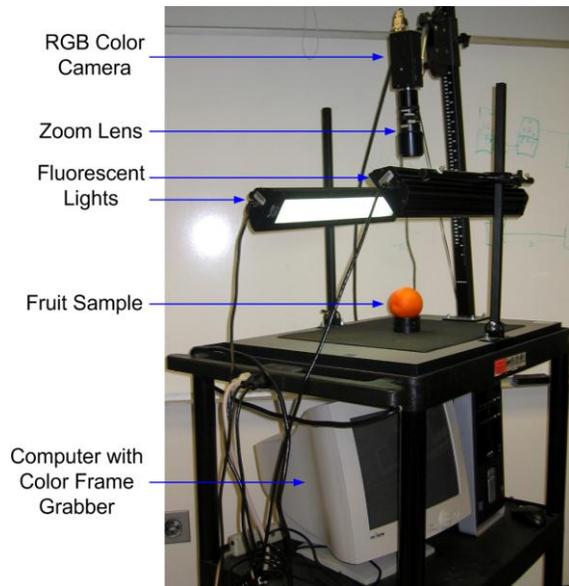


Figure 4-1. Color image system for acquiring RGB images from citrus samples

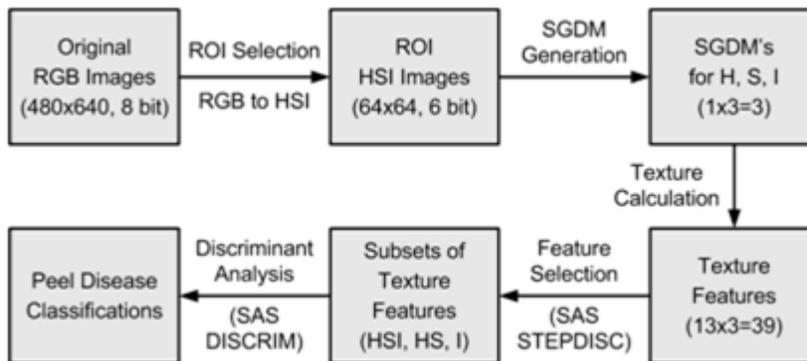


Figure 4-2. Procedures for color image analysis.

Table 4-1. Texture features selected by stepwise discriminant analysis

Classification Model	Color Feature	Texture Feature Set
HSI_13	H, S, I	H <sub>9</sub> , H <sub>10</sub> , I <sub>12</sub> , S <sub>7</sub> , I <sub>3</sub> , I <sub>2</sub> , S <sub>12</sub> , I <sub>11</sub> , I <sub>1</sub> , I <sub>8</sub> , S <sub>1</sub> , H <sub>2</sub> , H <sub>5</sub>
HS_9	H, S	H <sub>9</sub> , H <sub>10</sub> , S <sub>7</sub> , H <sub>5</sub> , H <sub>11</sub> , S <sub>12</sub> , S <sub>11</sub> , H <sub>7</sub> , H <sub>13</sub>
I_11	I	I <sub>2</sub> , I <sub>3</sub> , I <sub>5</sub> , I <sub>10</sub> , I <sub>6</sub> , I <sub>13</sub> , I <sub>8</sub> , I <sub>9</sub> , I <sub>1</sub> , I <sub>11</sub> , I <sub>7</sub>
HSI_39	H, S, I	All 39 texture features (H <sub>1</sub> - H <sub>13</sub> , S <sub>1</sub> - S <sub>13</sub> , I <sub>1</sub> - I <sub>13</sub> )

Table 4-2. Classification results using model HSI\_13 in Table 4-1

Actual Peel Condition	Classified Peel Condition						Accuracy (%)
	Normal	Canker	Copper Burn	Greasy Spot	Melanose	Wind Scar	
Normal	10	0	0	0	0	0	100.00
Canker	0	10	0	0	0	0	100.00
Copper Burn	0	0	10	0	0	0	100.00
Greasy Spot	0	0	1	9	0	0	90.00
Melanose	0	0	0	0	9	1	90.00
Wind Scar	0	0	0	0	0	10	100.00
Total	10	10	11	9	9	11	96.70

Table 4-3. Classification results in percent correct for all models in Table 4-1.

Peel Condition	Classification Model			
	HSI_13	HS_9	I_11	HSI_39
Normal	100.00	90.00	80.00	80.00
Canker	100.00	90.00	90.00	100.00
Copper Burn	100.00	90.00	100.00	90.00
Greasy Spot	90.00	90.00	100.00	90.00
Melanose	90.00	70.00	70.00	70.00
Wind Scar	100.00	90.00	50.00	100.00
Overall Accuracy (%)	96.67	86.70	81.70	88.30

Table 4-4. Classification results for shuffle data models in percent correct

Number of random data	Canker (%)	Copper (%)	Greasy Spot (%)	Normal (%)	Melanose (%)	Windscar (%)	Total (%)
1	100.00	90.00	100.00	100.00	100.00	100.00	98.33
2	100.00	100.00	100.00	80.00	90.00	100.00	95.00
3	100.00	100.00	100.00	100.00	90.00	100.00	98.33
4	80.00	100.00	100.00	90.00	90.00	100.00	93.33
5	100.00	100.00	90.00	100.00	90.00	90.00	95.00
6	100.00	100.00	100.00	100.00	100.00	100.00	100.00
7	100.00	90.00	100.00	100.00	100.00	90.00	96.67
8	100.00	100.00	100.00	90.00	100.00	80.00	95.00
9	100.00	90.00	100.00	90.00	90.00	100.00	95.00
10	80.00	100.00	100.00	100.00	90.00	90.00	93.33
Average Accuracy (%)							96.50

## CHAPTER 5 DETECTION OF CITRUS GREENING DISEASE ON ORANGE LEAVES

Huanglongbing (HLB), commonly known as citrus greening, is one of the most dangerous diseases that affect citrus production, and citrus greening has threatened to destroy an estimated 60 million trees in Africa and Asia (Ruangwong et al., 2006). Citrus greening was found in Miami-Dade County, Florida in August 2005. Florida citrus growers are fighting this disease which has the potential to destroy the state's \$9 billion commercial citrus industry (The American Phytopathological Society, 2008).

Citrus greening is a bacterial disease that affects the phloem system of citrus trees and causes the leaves of infected trees to become yellow, the trees to become unproductive, decline and possibly die within a few years. The bacterium is spread by an insect, the citrus psyllid. Citrus greening infects all types of citrus species, cultivars, and hybrids and some citrus relatives. The symptoms of citrus greening usually include blotchy, chlorotic mottling of leaves, yellow shoots, misshapen or lopsided small fruit that fail to color properly and stay green, hence the name of the disease "greening". The name huanglongbing means "yellow shoot", which is descriptive of the yellow sectors of infected trees (Gottwald et al., 2007). Currently, there is no cure for citrus greening, but early detection of the disease and appropriate management of the insect vector should alleviate the severity of the greening disease and minimize its spread. To reach this goal, many image processing and computer vision technologies have been developed to achieve the automatic identification of disease symptoms. The design and implementation of these technologies will greatly aid in scouting for the disease, selective chemical application, reducing costs and thus leading to improved productivity and fruit quality.

The identification of various plants and crops using image processing techniques has been attempted by several researchers. Haralick et al. (1973) used gray level co-occurrence features to analyze remotely sensed images. They computed gray level co-occurrence matrices for a pixel offset equal to one and with four directions ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ). For a seven-class classification problem, they obtained approximately 80% classification accuracy using texture features. Tang et al. (1999) developed a texture-based weed classification method using Gabor wavelets and neural networks for real-time selective herbicide application. The method comprised a low-level Gabor wavelet-based feature extraction algorithm and a high-level neural network-based pattern recognition algorithm. The model was specifically developed to classify images into broadleaf and grass categories for real-time herbicide application. Their analyses showed that the method is capable of performing texture based broadleaf and grass classification accurately with 100% classification accuracy. Burks et al. (2000) developed a method for classification of weed species using color texture features and discriminate analysis. The image analysis technique used for this method was the color-co-occurrence (CCM) method. The method had the ability to discriminate between multiple canopy species and was insensitive to leaf scale and orientation. The use of color features in the visible light spectrum provided additional image characteristic features over traditional gray-scale representation. The CCM method involved three major mathematical processes:

- Transformations of an RGB color representation of an image to an equivalent HSI color representation.
- Generation of color co-occurrence matrices from the HSI pixels

- Generation of texture features from the CCM matrices.

CCM texture feature data models for six classes of ground cover (giant foxtails, crabgrass, velvet leaf, lambs quarter, ivy leaf morning glory, and soil) were developed and stepwise discriminant analysis techniques were utilized to identify combinations of CCM texture feature variables, which have the highest classification accuracy with the least number of texture variables. A discriminant classifier was trained to identify weeds using the models generated. Classification tests were conducted with each model to determine their potential for classifying weed species. Overall classification accuracies above 93% were achieved when using hue and saturation features alone. A complete discussion of the CCM approach is found in Shearer and Holmes (1990). Pydipati et al. (2006) analyzed detection in citrus leaves using machine vision. The image data of the leaves selected for disease monitoring was collected. Then, algorithms based on image processing techniques for feature extraction and classification were designed. Manual feeding of datasets, in the form of digitized RGB color photographs was conducted for feature extraction and training the SAS statistical classifier. After training the SAS classifier, the test data sets were used to analyze the performance of accurate classification.

The overall objective of this research was to develop a machine vision based method for detecting citrus greening on leaves. This approach would use color texture features under controlled lighting in order to discriminate between greening and leaf conditions that are commonly confused with greening. This preliminary approach used low level magnification to enhance features. As a result, this would be conducted in a laboratory setting. Future studies would use field based detection. Specific objectives implemented to accomplish the overall objective were to:

- Use a digital color microscope system to collect RGB images from orange leaves with eight conditions (i.e., young flush, normal mature, blotchy mottle, green islands, zinc deficiency, iron deficiency, manganese deficiency and dead).
- Determine image texture features based on the color co-occurrence method (CCM).
- Create a set of reduced feature data models through a stepwise elimination process and classify different citrus leaf conditions.
- Compare the classification accuracies.

## **Materials and Methods**

### **Image Acquisition System**

A Digital Microscope system (VHX-600K, Keyence, JAPAN) was used for acquiring RGB images from citrus leaf samples, as shown in Figure 5-1. The imaging system consisted of a halogen lamp (12V, 100W), a zoom lens (C-mount lens, OP-51479), a 2.11-million-pixel CCD image sensor (1/1.8-inch), a 15-inch Color LCD monitor (TFT, 1600x1200, UXGA), and a computer installed with an image capture function and a hard disk drive unit (image format: JPEG and TIFF, Storage capacity: 700MB). The setup of the light source was designed to minimize specular reflectance and shadow, and to maximize the contrast of the images. The height of the camera and its focus were adjusted to contain the whole leaf, centered on the main leaf vein. Automatic white balance calibration was conducted using a calibrated white balance function in this system before acquiring images from leaf samples. The digital color images were saved in uncompressed JPEG format (1200x1600, 8bit).

### **Image Processing**

The image analysis technique selected for this study was the CCM method. The use of color image features in the visible light spectrum provides additional image characteristic features over the traditional gray-scale representation. The CCM

procedure consists of three primary mathematical processes. First the RGB images of leaves are converted to a hue, saturation, and intensity (HSI) color space representation. Intensity is calculated using the mean value of the three RGB values. The hue and saturation values are determined using a geometrical transformation of the International Commission on Illumination's (CIE) chromaticity diagram (Ohta, 1985). In this process, the CIE chromaticity diagram represents a two-dimensional hue and saturation space (Wyszecki et al., 1992). The pixel RGB values determine the chromaticity coordinates on the hue and saturation space, which are then used to geometrically calculate the value of hue and saturation. This process has been documented by Shearer (1986). Each pixel map was used to generate a color co-occurrence matrix after the HSI image was completed, resulting in three CCM matrices. That is, one CCM matrix for each of the HSI pixel maps. Through the use of spatial gray-level dependence matrices (SGDM's), the color co-occurrence texture analysis method was developed. The gray level co-occurrence methodology is a statistical method to describe shape by statistically sampling the way certain gray-levels occur in relation to other gray-levels. Shear and Homes (1990) explained that these matrices measure the probability that a pixel at one particular gray level will occur at a distinct distance and orientation from any pixel given that pixel has a second particular gray level. For a position operator  $p$ , we can define a matrix ' $P_{ij}$ ' that counts the number of times a pixel with grey-level  $i$  occurs at position  $p$  from a pixel with grey-level  $j$ . For example, if we have four distinct grey-levels 0,1,2 and 3, then Figure 5-2a, where  $i$  is the row indicator and  $j$  is the possible column indicator in the SGDM matrix  $P(i,j,d,\theta)$ . If we

normalize the matrix  $P$  by the total number of pixels so that each is between 0 and 1, we get a gray-level co-occurrence matrix.

The SGDMs are represented by the function  $P(i,j,d,\theta)$  where  $i$  represents the gray-level of location  $(x,y)$  in the image  $I(x,y)$ , and  $j$  represents the gray-level of the pixel at a distance  $d$  and an orientation angle of  $\theta$  from location  $(x,y)$ . An example image matrix  $I(x,y)$ , with a gray scale ranges from zero to three is shown in Figure 5-2b.

The hue, saturation and intensity CCM matrices are then used to generate the texture features described by Haralick and Shanmugam (1974). Shearer and Holmes (1990) reported a reduction in the 16 gray scale texture features through elimination of redundant variables. The resulting 11 texture feature equations are defined by Shearer and Holmes (1990). Donohue et al (1985) added image contrast and modulus texture features to those used by Ohta (1985), for a total of thirteen features when classifying cancer tissue. The same equations are used for each of the three CCM matrices, producing 13 texture features for each HSI component and thereby a total of 39 CCM texture statistics. The texture features are identified by a coded variable name where the first letter represents whether it is a hue (H), saturation (S) or intensity (I) feature and the number following represents one of the thirteen texture features described in Shearer (1990). As an example, the feature  $(I_7)$  is a measure of the entropy in the intensity CCM matrix, which represents the amount of order in an image and is calculated by equation 5-1.

$$I_7 = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p(i,j) \ln p(i,j) \quad (5-1)$$

The  $p(i,j)$  matrix represents the normalized intensity co-occurrence matrix and  $N_g$  represents the total number of intensity levels. The equation for normalizing the co-

occurrence matrix is given in equation 5-2, where  $P(i,j,1,0)$  is the intensity co-occurrence matrix.

$$p(i, j) = \frac{P(i, j, 1, 0)}{\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P(i, j, 1, 0)} \quad (5-2)$$

A physical representation of entropy (uncertainty) may be visualized by comparing a checkerboard-like image to an image where one half is black and the other half is white. The latter image is highly ordered having all pixels of the same intensity segregated into two distinct pixels groups, which gives greater certainty of the pixel value of the adjacent pixels. The checkerboard image has a lower amount of order due to intermixing of black and white squares, which results in a greater level of uncertainty of neighboring pixel values. The lower order image would therefore have more uncertainty and thus a higher entropy measure.

### **Features extraction**

Sixty images were taken of the top surface for each leaf class and centered on the mid. Digital images were stored in uncompressed JPEG format. The three classification models discussed previously were treated as separate classification problems. The 60 images from each class were divided into two datasets consisting of 30 samples for training and 30 samples for testing. The samples were first arranged in ascending order for the time the images were acquired. This approach minimizes negative time dependant variability, and reduces potential for data selection bias between the training and test datasets.

A detailed illustration of the image acquisition and classification process is given in Figure 5-3. Algorithms for image segmentation and texture feature generation were

developed in MATLAB. In the initial step, the RGB images of all leaf samples were obtained. For reducing the computational burden with minimal loss of texture feature quality, the image resolution was reduced from 1600x1200 pixels to 800x600 pixels and the reduced images were then converted from eight bit to six bit per channel RGB format. The subsequent steps were repeated for each image in the dataset. After the images were reduced, edge detection of the leaf was completed on each image of the leaf sample using the MATLAB program. Figure 5-4 exhibits a detailed edge detection process. First, each RGB image was converted to a gray image and then a binary image. Next, the edge of a binary image was detected by the command 'imerode' and 'imdilate' in Matlab. Once the edge detection was finished, the image was scanned from left to right for each row in the pixel map, and the area outside the leaf was zeroed to remove any background noise. In the next step, the images were then converted from RGB format to HSI format.

The Spatial Gray-Level Dependency Matrices (SGDMs) were then generated for each color pixel map of the image, one each for hue, saturation and intensity. It was decided during preliminary testing that the experiment would use the 0° CCM orientation angle and one offset, where the smaller the offset, the finer is the texture measured. Thus, a one pixel offset is the finest texture measure. From the SGDM matrices, the 39 CCM texture statistics described earlier were generated for each image using the three color feature co-occurrence matrices, as each SGDM matrix provided 13 texture features. A more complete description of this technique can be found in Haralick and Shanmugam (1974), and Shearer and Holmes (1990).

## **Classification algorithms**

Once the texture statistics were generated for each image, SAS statistical analyses were conducted using procedure STEPDISC to reduce redundancy in the texture feature set. The training image dataset was used for the variable reduction analysis. SAS offers procedures for reducing variable set size and for discriminating between classes (SAS, 1985). PROC STEPDISC is used to reduce the number of texture features by a stepwise selection process. The "stepwise" selection procedure begins with no variables in the classification model (SAS, 1985). At each step of the process, the variable within and outside the model are evaluated. The variable within the model, at that particular step, which contributes least to the model as determined by the Wilks` Lambda method is removed from the model. Likewise, the variable outside the model that contributes most to the model is added. When no more steps can be taken, the number of variables in the model is reduced to its final form. Based on these analyses, several data models were created, which are shown in Table 5-1. Model HSI\_16 consisted of all conditions, HSI\_11 model consisted of all condition except normal young and the HSI\_9 model consisted of all conditions except blotchy mottle and normal young leaves.

When using PROC DISCRIM procedure, a discriminant function is established using a measure of the generalized squared distance between the image texture variables and the class texture variable means. The posterior probability determines the classification accuracy. The classification criterion may be affected by the pooled covariance matrix of the training set texture variables and the prior probabilities of the classification groups.

## Results and Discussion

### Classifications of Citrus Disease Conditions

The texture feature dataset was generated by containing 39 texture features for each image. The dataset had 420 rows each, representing 60 samples from each of the seven classes of leaves. Each row had 39 columns representing the 39 texture features extracted for a particular sample image. Each row had a unique number (1, 2, 3, 4, 5, 6, or 7) representing which class the particular row of data belonged. '1' represented blotchy mottle leaves, '2' represented green island leaves, '3' represented leaves deficient in iron, '4' represented leaves deficient in manganese, '5' represented young flush leaves, '6' represented normal mature leaves and '7' represented leaves deficient in zinc.

To compare classification accuracies under various disease conditions, three models were created which are shown in Table 5-1. These models represent the compilation of three different leaf conditions sets, which isolate leaf conditions that are difficult to discriminate. Table 5-2 shows four different models which have all leaf conditions except young flush, but have various combinations of color texture features. This set of models was selected to isolate crucial color texture features which can lead to more efficient feature generation. The training and testing sets for each model mentioned in Tables 5-1 and 5-2 were obtained by selecting either, intensity, hue and saturation or all three HSI features from the total 39 texture features in the original data files.

Once several data models were formed, SAS procedure STEPDISC was used to reduce the number of texture features included in the models. As can be seen in Table 5-2, significant elimination of redundant variables was accomplished. For instance,

HSI\_18 model had 39 texture features in the unreduced form, and was reduced to 18 features through the stepwise linear reduction process. The simplification of the data model serves several important purposes: 1) it reduces the computational burden of the redundant features, 2) it tends to improve the performance of classification algorithms, and 3) it reduces memory and storage demands. The most significant variable reduction was found in I\_8 model, which were reduced from 39 to 8 texture features after using STEPDISC.

SAS procedure DISCRIM was used to test the various data model classification accuracies. Each of the models was trained and tested using the appropriate image data set. The classification results were recorded on an individual disease category basis using the SAS procedure output listing. The results shown in Table 5-3 are the classification summary from the HSI\_18, HSI\_15 and HSI\_14 given in Table 5-1. As previously indicated, the test data consisted of 30 images from each category. The overall performance of HSI\_18 model was 86.67%, which is the lowest accuracy among the three models shown in Table 5-1. HSI\_15 and HSI\_14 models had high classification accuracies (95.56% and 97.33%). Based on the results shown in Table 5-4, the classification model using only intensity texture features presented the worst performance at 81.11% for the I\_11 model. When compared with other models, HS\_10 model had 87.78%, HSI\_15 model had 95.56% and HSI\_39 model had 95.60%. Therefore, other models provided better performance than the model that used only intensity texture features. In Table 5-4, the highest overall performance was 95.60% for HSI\_39. This demonstrates that significant classification improvement occurs when intensity features are used, and there is no loss in accuracy when using the reduced

HSI data set or the unreduced data set. As shown in Table 5-5, most images were correctly classified into the appropriate category; however, young flush leaves had a very low classification at 23%. The negative influence of young flush leaves was further demonstrated in the results from Table 5-3 where the classification accuracy is 86.67% while other leaf condition models that exclude young flush leaves have accuracy above 95%. Table 5-6 show improved classification accuracy of 95.56% and thus proved that young flush leaves affected overall performance result. Table 5-7 demonstrates that HSI\_14 model was the best accuracy (97.33%) in three leaf condition models in Table 5-1. However, HSI\_14 model excluded citrus greening blotchy mottle, and thus ignored the most important greening identifier. Moreover, there was no significant difference in the classification results between HSI\_15 and HSI\_14 models. These effects can be seen in confusion matrices, where a model exhibits the classification between positive vs. negative greening symptoms.

### **The Confusion Matrix for Greening Positive vs. Negative**

A consistent analysis of classifier behavior can be provided by the semi-global performance matrix, known as Confusion Matrix. This matrix provides a quantitative performance representation for each classifier in terms of class recognition. One benefit of a confusion matrix is that it is easy to see if the system is confusing two classes, citrus greening symptom and non-greening symptoms. The classification results for the confusion matrix obtained under the positive vs. negative greening model are shown in Tables 5-8, 5-9 and 5-10. In general the HSI\_18 model had the lowest classification accuracy among all symptom models. However, in the confusion matrix shown in Table 5-8, the discrimination of citrus greening symptoms had high success rate (96.7%). On the other hand, the accuracy for greening was only 82.67%, giving an overall accuracy

of 86.67%. In the HSI\_15 model, the young flush leaves were removed and a 95.6% overall classification accuracy was achieved. This model also has good classification accuracies between positive (91.67%) and negative (97.50%) as shown in Table 5-9. Model HSI\_14 excluded young flush leaves and blotchy mottle, and achieved an increase in classification performance when compared with the HSI\_18 model (97.3% versus 86.7%). In the confusion matrix shown in Table 5-10, HSI\_14 used all disease conditions except young flush and blotchy mottle and had the same positive greening accuracy as HSI\_18 model which used all disease condition. However, the overall accuracy was much higher 97.3%. When comparing each model, it is likely that the similarity between young flush leaves and other conditions affected the detection accuracy of citrus greening disease.

### **Stability Test of the Greening Classification Model**

From the results stated above, leaf condition models were evaluated to determine which scenario would perform the best in distinguishing greening symptoms. It was also important to evaluate various texture feature combinations to determine which would provide high classification accuracy and demonstrate model stability, under varying training and testing conditions. The classification results presented above were obtained using test samples selected in a fixed order. In order to test the stability of the classification model, 30 training samples and 30 testing samples were randomly chosen from the 60 samples for each condition. They were used to train and test a selected model using the same procedures described earlier. Ten runs were repeated for training and testing. In this research, stability tests were provided for the HSI\_15 model, since this model had demonstrated good performance on the leaf conditions of most significant interest. The average value shown in Table 5-11 was 94.06%. These results

demonstrated that the classification model, excluding young flush leaves, was robust under varying leaf sample conditions, and therefore should provide a viable classification of greening conditions.

### **Summary**

Data analysis based on the color co-occurrence method is useful for detection of citrus greening disease. A color imaging system was selected to obtain RGB images from citrus leaves consisting of two normal leaf conditions, young flush and mature. In addition, five leaf conditions including greening blotchy mottle, green islands, manganese deficiency, iron deficiency, and zinc deficiency were collected. Images of the leaf surface were extracted from the original RGB images, and then converted into hue, saturation, and intensity (HSI) color space representation. Each HSI image was used to generate spatial gray-level dependence matrices. Once SGDMs were generated, a total of 39 image texture features were obtained from each citrus leaf sample. Algorithms for selecting useful texture features were developed based on a stepwise discriminant analysis for three disease combinations including all conditions, all conditions excluding young flush, and conditions excluding blotchy mottle and young flush. Through a discriminant function based on a measure of the generalized squared distance, classification models were constructed using the reduced texture feature sets.

Beneficial elimination of redundant texture features were accomplished through the stepwise discriminant analysis. Various texture features models were selected from the color combinations of HSI. The elimination of redundant texture features significantly reduces the computation burden, and it also helps improve the performance of classification models. The classification model excluding blotchy mottle and young flush (HSI\_14) gave the best accuracy (97.33%), while HSI\_18 model achieved the worst

classification accuracy (86.67%). When excluding only young flush condition, the classification had high accuracy of 95.60%. The results suggested that young flush samples collected in fall created confusion between normal mature leaves. This fact also can be seen in the confusion matrix accuracies in Tables 5-8, 5-9 and 5-10. HSI\_18 model had the lowest classification accuracy, but the success rate of positive for greening disease was 96.67%. It was the same or higher than others. A stability test for the classification model with the best performance was accomplished by 10 runs using randomly selected training and testing samples. Average classification accuracy was 94.06%, indicating that the classification model is robust for classifying new citrus leaf samples according to their conditions. For further study of the influence of young flush, it is suggested that a new model consisting of a mixed data set of young flush and mature normal leaves should be evaluated to see how it compares to the model, which excluded young flush leaves.

This research demonstrated that color imaging and texture feature analysis could be used at low magnification for differentiating citrus greening symptoms from other leaf conditions. Future studies will explore the utility of these algorithms in outdoor conditions.

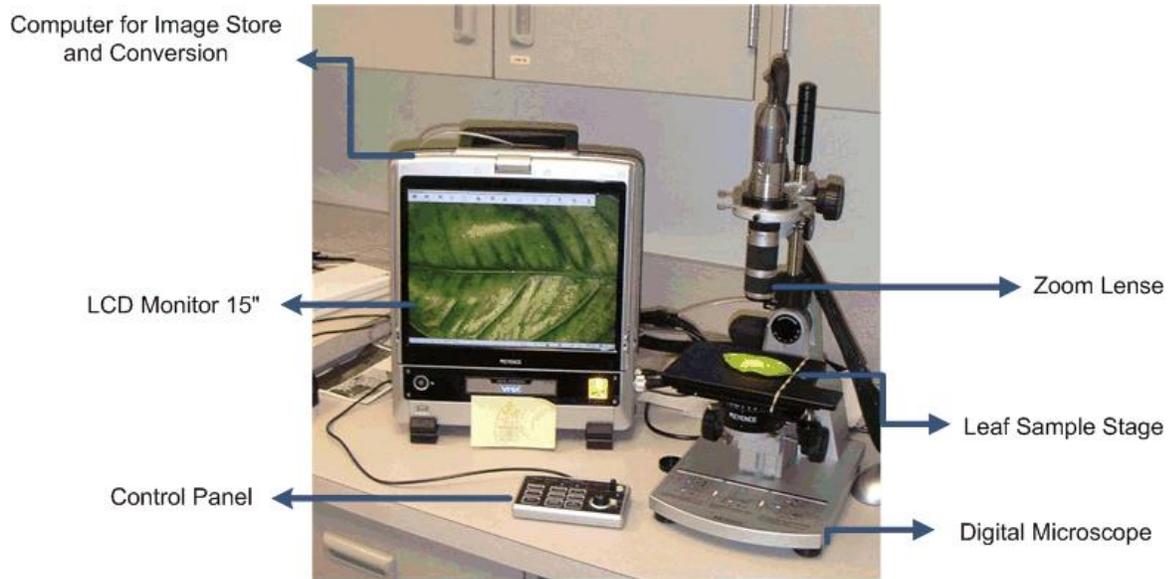


Figure 5-1. Digital microscope system for acquiring RGB images from citrus leaf samples

$$(a) P(i, j, 1, 0^\circ) = \begin{bmatrix} 2 & 1 & 2 & 2 \\ 1 & 0 & 3 & 2 \\ 2 & 3 & 0 & 1 \\ 2 & 2 & 1 & 0 \end{bmatrix} \quad (b) I(x, y) = \begin{bmatrix} 0 & 0 & 3 & 1 \\ 2 & 1 & 0 & 0 \\ 3 & 2 & 0 & 3 \\ 1 & 2 & 1 & 3 \end{bmatrix}$$

Figure 5-2. Gray-level dependence example: (a) SGDM for different orientations, (b) gray-level image

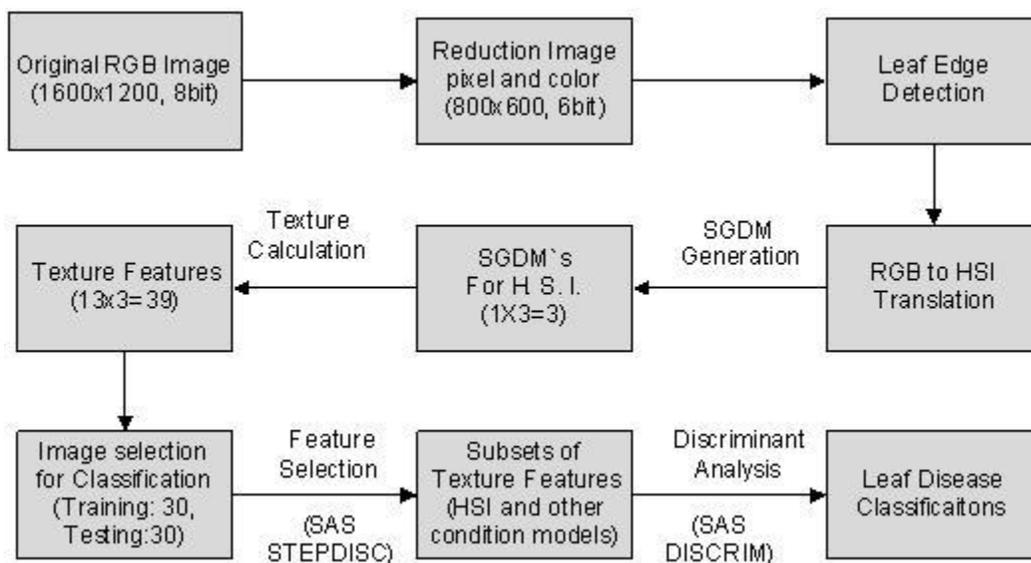


Figure 5-3. Procedures for color image analysis

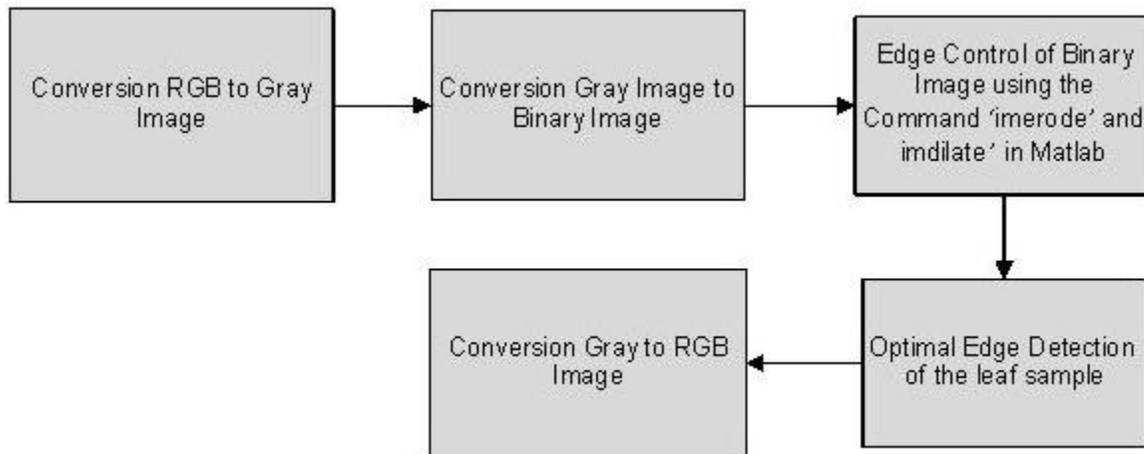


Figure 5-4. Procedures for leaf edge detection

Table 5-1. Texture feature models selected by stepwise discriminant analysis for fall season

Classification Condition	Classification Model <sup>1</sup>	Color Feature <sup>2</sup>	Texture Feature Set <sup>3</sup>
All disease condition	HSI_18	H, S, I	S <sub>4</sub> , I <sub>2</sub> , H <sub>7</sub> , S <sub>13</sub> , H <sub>2</sub> , H <sub>9</sub> , S <sub>5</sub> , I <sub>7</sub> , S <sub>7</sub> , I <sub>9</sub> , S <sub>8</sub> , I <sub>1</sub> , I <sub>10</sub> , H <sub>4</sub> , I <sub>6</sub> , S <sub>6</sub> , H <sub>8</sub> , I <sub>13</sub>
All conditions except young flush	HSI_15	H, S, I	S <sub>5</sub> , I <sub>2</sub> , H <sub>7</sub> , H <sub>2</sub> , S <sub>6</sub> , S <sub>4</sub> , H <sub>9</sub> , S <sub>8</sub> , I <sub>6</sub> , S <sub>13</sub> , H <sub>4</sub> , I <sub>4</sub> , I <sub>13</sub> , S <sub>7</sub> , I <sub>7</sub>
All conditions except blotch mottle and young flush	HSI_14	H, S, I	S <sub>5</sub> , I <sub>2</sub> , H <sub>7</sub> , H <sub>2</sub> , S <sub>4</sub> , H <sub>9</sub> , S <sub>13</sub> , S <sub>7</sub> , I <sub>7</sub> , I <sub>1</sub> , I <sub>9</sub> , S <sub>8</sub> , I <sub>10</sub> , I <sub>6</sub>

1. Classification model designation based color features in model and the total number of variable selected by STEPDISC.

2. Color texture features included in initial data set prior to reduction. 13 variables for color texture feature set.

3. Find texture features selected, given in order of discriminant power.

Table 5-2. Texture feature models to all conditions except young flush for fall season

Classification Model <sup>1</sup>	Color Feature <sup>2</sup>	Texture Feature Set <sup>3</sup>
HSI_15	H, S, I	S <sub>5</sub> , I <sub>2</sub> , H <sub>7</sub> , H <sub>2</sub> , S <sub>6</sub> , S <sub>4</sub> , H <sub>9</sub> , S <sub>8</sub> , I <sub>6</sub> , S <sub>13</sub> , H <sub>4</sub> , I <sub>4</sub> , I <sub>13</sub> , S <sub>7</sub> , I <sub>7</sub>
HS_10	H, S	S <sub>5</sub> , H <sub>7</sub> , H <sub>5</sub> , H <sub>12</sub> , S <sub>4</sub> , S <sub>7</sub> , H <sub>8</sub> , S <sub>8</sub> , H <sub>3</sub> , S <sub>11</sub>
I_8	I	I <sub>2</sub> , I <sub>8</sub> , I <sub>9</sub> , I <sub>6</sub> , I <sub>5</sub> , I <sub>7</sub> , I <sub>10</sub> , I <sub>1</sub>
HSI_39	H, S, I	All 39 texture features (H <sub>1</sub> - H <sub>13</sub> , S <sub>1</sub> - S <sub>13</sub> , I <sub>1</sub> - I <sub>13</sub> )

1. Classification model designation based color features in model and the total number of variable selected by STEPDISC.

2. Color texture features included in initial data set prior to reduction. 13 variables for color texture feature set.

3. Find texture features selected, given in order of discriminant power.

Table 5-3. Classification summary in percent correct for all models in Table 5-1

Disease Condition	Classification Model		
	HSI_18	HSI_15	HSI_14
Blotchy mottle	96.67	90.00	-
Islands	96.67	93.33	96.67
Iron deficiency	90.00	100.00	90.00
MN deficiency	100.00	96.67	100.00
Zinc deficiency	100.00	100.00	100.00
Normal	100.00	93.33	100.00
Young flush	23.33	-	-
Overall Accuracy (%)	86.67	95.56	97.33

Table 5-4. Classification summary in percent correct for all models in Table 5-2

Disease Condition	Classification Model			
	HSI_15	HS_10	I_11	HSI_39
Blotchy mottle	90.00	70.00	70.00	96.67
Islands	93.33	76.67	73.33	93.33
Iron deficiency	100.00	100.00	86.67	93.33
MN deficiency	96.67	93.33	93.33	96.67
Zinc deficiency	100.00	100.00	83.33	96.67
Normal	93.33	86.67	80.00	96.67
Overall Accuracy (%)	95.56	87.78	81.11	95.60

Table 5-5. Classification result in percent correct for HSI\_18 model in Table 5-1

Actual Leaf Condition	Classified leaf Condition							Accuracy (%)
	Blotchy mottle	Islands	Zinc deficiency	Iron deficiency	MN deficiency	Normal	Young flush	
Blotchy mottle	29	0	0	0	0	1	0	96.67
Islands	0	29	1	0	0	0	0	96.67
Zinc deficiency	0	0	30	0	0	0	0	100.00
Iron deficiency	0	0	1	27	0	0	2	90.00
MN deficiency	0	0	0	0	30	0	0	100.00
Normal	0	0	0	0	0	30	0	100.00
Young flush	4	0	0	3	1	15	7	23.33
Total	33	29	32	27	31	46	9	86.67

Table 5-6. Classification result in percent correct for HSI\_15 model in Table 5-1

Actual Leaf Condition	Classified leaf Condition						Accuracy (%)
	Blotchy mottle	Islands	Zinc deficiency	Iron deficiency	MN deficiency	Normal	
Blotchy mottle	27	0	0	0	0	3	90.00
Islands	0	28	1	1	0	0	93.33
Zinc deficiency	0	0	30	0	0	0	100.00
Iron deficiency	0	0	0	30	0	0	100.00
MN deficiency	1	0	0	0	29	0	96.67
Normal	2	0	0	0	0	28	93.33
Total	30	28	31	31	29	31	95.56

Table 5-7. Classification result in percent correct for HSI\_14 model in Table 5-1

Actual Leaf Condition	Classified leaf Condition					Accuracy (%)
	Islands	Zinc deficiency	Iron deficiency	MN deficiency	Normal Mature	
Islands	29	1	0	0	0	96.67
Zinc deficiency	0	30	0	0	0	100.00
Iron deficiency	0	3	27	0	0	90.00
MN deficiency	0	0	0	30	0	100.00
Normal Mature	0	0	0	0	30	100.00
Total	29	34	27	30	30	97.33

Table 5-8. Confusion matrix in percent correct for HSI\_18 model in Table 5-1

Prediction outcome		Positive for greening (Blotch mottle, Islands)	Negative for greening (Young flush, Normal, MN, IR, ZN)	Total
Actual value	True (success rate)	58/60 (96.67%)	124/150 (82.67%)	182/210 (86.67%)
	False (fail rate)	26/150 (17.33%)	2/60 (3.33%)	28/180 (13.33%)

Table 5-9. Confusion matrix in percent correct for HSI\_15 model in Table 5-1

Prediction outcome		Positive for greening (Blotch mottle, Islands)	Negative for greening (Normal, MN, IR, ZN)	Total
Actual value	True (success rate)	55/60 (91.67%)	117/120 (97.50%)	172/180 (95.56%)
	False (fail rate)	3/120 (2.50%)	5/60 (8.33%)	8/180 (4.44%)

Table 5-10. Confusion matrix in percent correct for HSI\_14 model in Table 5-1

Prediction outcome		Positive for greening (Islands)	Negative for greening (Normal, MN, IR, ZN)	Total
Actual value	True (success rate)	29/30 (96.67%)	117/120 (97.5%)	146/150 (97.33%)
	False (fail rate)	3/120 (2.50%)	1/30 (3.33%)	4/150 (2.67%)

Table 5-11. Classification results for shuffle data about HSI\_15 model in percent correct

Number of random data	Blotchy mottle (%)	Islands (%)	Normal (%)	MN (%)	Zinc (%)	Iron (%)	Total (%)
1	83.33	83.33	93.33	100.00	86.67	100.00	91.11
2	90.00	96.67	96.67	93.33	96.67	100.00	95.56
3	73.33	100.00	96.67	96.67	90.00	100.00	92.78
4	83.33	90.00	96.67	86.67	96.67	90.00	90.56
5	83.33	100.00	96.67	100.00	100.00	100.00	96.67
6	83.33	96.67	96.67	100.00	100.00	90.00	94.44
7	93.33	96.67	96.67	96.67	90.00	96.67	95.00
8	80.00	83.33	100.00	96.67	90.00	93.33	90.56
9	90.00	100.00	83.33	100.00	93.33	100.00	94.44
10	90.67	100.00	96.67	96.67	96.67	100.00	97.78
Average Accuracy (%)	86.67	94.67	95.34	96.67	94.00	97.00	94.06

## CHAPTER 6 DETECTION OF THE DISEASE USING PATTERN RECOGNITION METHODS

Early detection of citrus diseases is important for citrus fruit production and quality. In Florida, the citrus groves are under attack from two major diseases, namely citrus canker and greening. Citrus greening diseases affect fruit quality. The negative impacts reduce the profitability of the citrus industry, and threaten the agricultural economy of the state of Florida. Hence, early detection of the disease can reduce its spread and minimize losses for citrus growers.

For reducing labor cost and improving detection accuracy, pattern recognition method have demonstrated success in a variety of research areas such as aerospace, defense, medical, neurobiology, and linguistics etc. Artificial intelligence is an approach that mimics humans intelligent to build learning ability, reasoning, and perception. The research about artificial intelligence has developed into a discipline known as intelligent system.

In agricultural engineering, many researchers have applied pattern recognition methods to agricultural management. Park et al. (2007) detected fecal contamination in the visceral cavity of broiler carcasses using a pattern recognition method. The method comprised fisher linear discriminant analysis. Images of poultry carcasses were collected using hyperspectral imaging processing. Their analysis showed that the method is capable of detecting fecal contamination on the surface of broiler carcasses with 98.9% classification accuracy.

Cheng et al. (2003) developed an approach for fruit and vegetable defect inspection. The hyperspectral imaging classification techniques used for this study were the principal component analysis (PCA) and fisher's linear discriminant (FLD) method.

These methods have the ability to maximize the representation and classification effects on the extracted new feature bands. The use of hyperspectral image features provided high dimension feature space and reflection properties. PCA were employed for reducing the hyperspectral dimension. FLD techniques were utilized to classify wholesome and unwholesome objects. Overall classification accuracy using only FLD solution was 90% but the combined PCA-FLD solution had the accuracy of 93.3%. When the PCA and FLD methods were integrated, classification accuracy was better.

H. Zhang et al. (2007) analyzed fungal infected wheat kernels using a support vector machine (SVM). The image data of wheat kernels was collected by a near-infrared reflectance hyper spectral imaging system. Then, after four features were extracted from input images, algorithms based on principal component analysis (PCA) technique for reducing the dimensionality of pattern vectors were designed. The NIR hyperspectral image datasets was used for the SVM classifier. After classifying the datasets, the overall classification accuracy was 94.8%, with 531 kernels correctly classified and 29 kernels not.

Pydipati et al. (2005) developed disease detection in citrus leaves using statistical and neural network classifiers. He used co-occurrence matrices to extract texture features from HSI images, and the SAS classifier was used to train HSI feature dataset. After reducing image datasets, the classification results using neural network method and SAS classifier was compared. SAS classifier achieved accuracy above 95% for all classes, while neural network algorithms achieved accuracy above 90% for all classes.

There are various approaches in pattern recognition. In this study, three pattern recognition approaches were used for finding an optimal pattern recognition approach.

- Linear models for classification: Fisher's linear discriminant analysis method
- Neural Networks for classification: Back propagation based on neural network method
- Nonlinear Kernel for classification: Support Vector Machine (SVM)

The objective of this study is to find a pattern recognition method for detection of citrus diseases. For preliminary study, image analysis techniques based on color co-occurrence method will be developed and three pattern recognition methods will be compared for best disease image classification. Images were acquired under controlled lighting conditions and low level magnification to enhance features.

Specific objectives were to:

- Collect two image data sets of citrus canker and citrus greening diseases.
- Evaluate the color co-occurrence method for disease detection.
- Develop various pattern recognition algorithms for classification of the citrus disease conditions based on the features obtained from the color co-occurrence method.
- Compare the classification accuracies of the various pattern recognition classifiers.

## **Materials and Methods**

### **Image Acquisition System**

There are differences between the hardware systems used for Canker and Greening disease detection acquiring best RGB images. First, canker images were acquired by the color image acquisition system in Figure 6-1. The imaging system consisted of two 13 W high frequency sealed fluorescent lights (SL Series, StockerYale, Salem, NH, USA), a zoom lens (Zoom 7000, Navitar, Rochester, NY, USA), a 3-CCD RGB color camera (CV-M90, JAI, San Jose, CA, USA), a 24-bit color frame grabber board with 480x640 pixel resolution (PC-RGB, Coreco Imaging, St. Laurent, Quebec,

CA), and a computer installed with an image capture software. The setup of the lighting system was designed to minimize specular reflectance and shadow, and to maximize the contrast of the images. The height of the camera and its focus were adjusted to contain the image of the whole fruit, with an approximate 100 mm×100 mm field of view. Automatic white balance calibration was conducted using a calibrated white balance card before acquiring images from fruit samples. The digital color images were saved in uncompressed BMP format.

Second, A Digital Microscope system (VHX-600K, Keyence, JAPAN) was used for acquiring RGB images from greening samples, as shown in Figure 6-2. The imaging system consisted of Halogen lamp (12V, 100W), a zoom lens (C-mount lens, OP-51479), a 2.11-million-pixel CCD image sensor (1/1.8-inch), a 15-inch Color LCD monitor (TFT, 1600x1200, UXGA), and a computer installed with an image capture function and a hard disk drive unit (Image format: JPEG and TIFF, Storage capacity: 700MB). The setup of the light source was designed to minimize specular reflectance and shadow, and to maximize the contrast of the images. The height of the camera and its focus were adjusted to contain the whole samples. Automatic white balance calibration was conducted using a calibrated white balance function in this system before acquiring images from samples. The digital color images were saved in uncompressed JPEG format (1200x1600, 8bit and 480x640, 8bit).

### **Image Analysis**

The methodology employed for citrus canker and greening classification is very similar. Figure 6-3 and Figure 6-4 illustrated citrus canker and greening classification procedures. However, there are differences in pre-processing since the citrus canker

approach focuses on specific sub region, while the citrus greening approach examines the whole leaf surface.

From the original RGB color images (480x640 pixel), citrus canker region of interest (ROI) images were extracted. This ROI was focused on peel conditions of interest. The ROI selection was started manually by establishing the center point of a 64x64 pixel window on the original image. Once extracted, each image was converted from RGB (red, green, blue) to HSI (hue, saturation, intensity) color format. This approach obtains the useful image data and significantly reduces the computational burden for the following data analysis procedures.

Citrus greening analysis has additional pre-processing steps. For reducing the computational burden with minimal loss of texture feature quality, the image resolution was reduced from 1600x1200 pixels to 800x600 pixels and the reduced images were then converted from eight bit to six bit per channel RGB format. The subsequent steps were repeated for each image in the dataset. After the images were reduced, edge detection of the leaf boundary was completed on each the leaf sample image. Once the edge detection was finished, the image was scanned from left to right for each row in the pixel map, and the area outside the leaf was zeroed to remove any background noise. Then, the images were converted from RGB format to HSI format.

For extracting features from digitized HSI images, Color Co-occurrence Method (CCM) is used. Shear and Holmes (1990) described 39 CCM texture statistics using the Spatial Gray-level Dependence Matrices (SGDM). CCM texture statistics were generated from the SGDM of each HSI color feature. Each of the three matrices is evaluated by thirteen texture statistic measures resulting in 39 texture features per

image. The SGDM is a measure of the probability that a given pixel at one particular gray-level will occur at a distinct distance and orientation angle from another pixel, given that pixel has a second particular gray-level. The SGDM represented by the function,  $P(i, j, d, \theta)$ , when  $i$  and  $j$  represent gray values,  $d$  is the offset distance and  $\theta$  is orientation angle. Haralick and Shanmugam (1974) and Shearer and Holmes (1990) made detailed paper about this technique.

After obtaining 39 texture features for each image, feature selection was used to eliminate the redundancy in the texture feature set. The SAS procedure STEPDISC can reduce the size of the variable set and find the most significant variables for discriminating samples into different classes. The stepwise discriminant analysis begins with no variables in the classification model. At each step of the process, the variables within and outside the model are evaluated. The variable within the model, which contributes least to the model as determined by the Wilks' Lambda method is deleted from the model, while the variable outside the model that contributes most to the model and passes the test to be admitted is included. When more steps cannot be taken, the number of variables in the model is reduced to its final form. Burks et al. (2000) had shown that when only hue or saturation information was used in the classification models, classification accuracies were increased. Thus three color feature combinations including 1) hue, saturation, and intensity (H, S, I), 2) hue and saturation (H, S), and 3) intensity (I) only were used to perform the texture feature selections. These three color combinations have demonstrated high classification accuracies in the applications for other plant discriminations (Burks et al. 2000; Pydipati et al. 2006).

## **Classification Using Pattern Recognition Methods**

Once reduced texture features were obtained from STEPDISC procedure, two datasets were created, one for training and one for testing. Table 6-1 and Table 6-5 represents two datasets for citrus canker disease and citrus greening disease, respectively. For canker disease classification, the rows of training datasets consisted of 20 samples from each of the six classes of peel condition as discussed earlier. The columns represented the reduced texture features. For each image sample, test datasets had 10 samples and reduced texture features. Datasets for citrus greening had seven classes with 30 samples for training and 30 samples for testing.

These datasets were analyzed using a linear fisher discriminant analysis (FDA), a neural network based on back-propagation, and support vector classification (SVC). All analysis was done using MATLAB R2007b (The Mathworks, Inc., Natick, Mass.) and the MATLAB PRTools toolbox (Faculty of Applied Physics, Delft University of Technology, The Netherlands). PRTools is a toolkit for MATLAB released for the developed and evaluation of pattern recognition algorithms.

### **Classification Results**

#### **Canker Disease Classification Based on Pattern Recognition Algorithms.**

The texture feature selection results are summarized in Table 6-1. Four classification models were developed using the selected texture feature sets from the three color combinations [(H, S, I), (H, S), and (I)]. From the SAS STEPDISC procedure, the variables listed in Table 6-1 were selected for the first three models with variables arranged in the descending order of importance for the classification models.

The classification models were named using the color features involved in the texture feature selections, followed by the total numbers of the selected texture

features. For example, model HSI\_13 consists of a reduced set of 13 hue, saturation and intensity texture features. As shown in Table 6-1, significant eliminations of redundant texture features were accomplished through the stepwise discriminant analysis. Nine and eleven texture features were selected for model HS\_9 and model I\_11, respectively. The simplification of the texture features reduces the computation burden due to the redundant data, and it also helps improve the performance of classification models. In addition to the three models described above, a classification model that used all 39 HSI texture features was used for the purpose of comparisons with other models. Thus four classification models were used to differentiate citrus peel diseases, each model was independently evaluated for classification performance.

The results shown in Table 6-2 were obtained using a fisher discriminant analysis (FDA) classifier. In particular, better overall classification rates were achieved by models HSI\_13 and HSI\_39 which had overall accuracy of 91.67% and 93.33%. The results using neural network based on back propagation algorithm in Table 6-3 also had good overall accuracy. Model HSI\_13 achieved an overall accuracy of 95.00% and model HSI\_39 an accuracy of 93.33%. Table 6-4 showed classification results using support vector method (SVM) algorithm. This approach showed lower overall accuracy than other methods, where only model HSI\_13 achieved good accuracy at 95.00%.

These results proved that model HSI\_13 was best model for all classifier of canker disease. Model HSI\_39 also obtained good classification accuracy in FDA and the neural network methods, but it is not useful in real world applications, since it would require large calculation time due to the number of texture features. The best overall

performance was the SVM using model HSI\_13, which had overall classification accuracies of 95.00%, and no individual class below 80%.

### **Citrus Greening Disease Classification Based on Pattern Recognition Algorithms.**

The dataset for citrus greening consisted of 39 texture features for each image. The dataset had 360 rows each, representing 60 samples from each of the seven classes of leaves. Each row had 39 columns representing the 39 texture features extracted for a particular sample image. Each row had a unique number (1, 2, 3, 4, 5, or 6) representing which class the particular row of data belonged. '1' represented greening blotchy mottle leaves, '2' represented greening island leaves, '3' represented leaves deficient in iron, '4' represented leaves deficient in manganese, '5' represented normal mature leaves and '6' represented leaves deficient in zinc.

Table 6-5 shows the classification models which are named using the total numbers of selected texture features similar to the citrus canker models. A classification models using all 39 HSI texture features was selected for comparison with the reduced data models.

The results shown in Tables 6-6, 6-7 and 6-8 were obtained using a fisher discriminant analysis (FDA) classifier, neural network based on back propagation (BP) algorithm, and support vector machine (SVM). In Tables 6-6 and 6-7, models HSI\_15 and HSI\_39 achieved excellent overall classification rates. Models HSI\_15 and HSI\_39 in Table 6-7 had an overall accuracy of 95.55% and 93.89%, while Models HSI\_15 and HSI\_39 in Table 6-8 had an overall accuracy of 93.33% and 93.89%.

The results using SVM method in Table 6-8 showed lower overall accuracy than other results. In Table 6-8 using SVM method, model HSI\_15 and HS\_9 only had a accuracy of 80%. Model HS\_9 had the best overall accuracy at 84.44%, yet was lower

than other classification approaches. Therefore, SVM method was not as good for detecting citrus greening disease, since blotchy mottle confused the SVM classifier and significantly reduced the accuracy. When blotchy mottle and other disease classes have a very similar pattern, it resulted in low accuracy.

Hence, the FDA classifier using model HSI\_15 was best model for detection of citrus greening disease. Although model HSI\_39 also performed well on both cases, model HSI\_39 may increase computation time for training and classification. As with the canker disease classifier above, the reduced HSI texture model was best for greening detection. However, in this case, the FDA classifier outperformed the SVM.

### **Summary**

Pattern recognition and color co-occurrence texture method is a useful approach for detection of citrus disease. Digitized RGB images from various citrus disease conditions were obtained using an image acquisition system. Texture features containing useful information for diseases classification were extracted from the pre-processed RGB images that had been converted into hue, saturation, and intensity (HSI) color space representation. For each HSI image, three spatial gray-level dependence matrices (SGDMs) were generated, and a total of 39 image texture features were obtained from each image sample. A stepwise discriminant analysis was used finding useful texture features from three color combinations including 1) hue, saturation, and intensity (HSI), 2) hue and saturation (HS), and 3) intensity (I). Classification models were constructed using the reduced texture feature sets through a discriminant function based on a measure of the generalized squared distance. Various texture features models were selected from the color combinations of HSI. The elimination of redundant texture features significantly reduces the computation burden,

and it also improved the performance of classification models. For canker detection, the reduced classification model (HSI\_13) gave best accuracy (95.00%) for back-propagation based neural network method and support vector machine. In citrus greening, the reduced models (HSI\_15) showed an overall accuracy above 93% in back-propagation based on neural network method and the fisher discriminant analysis method. The HSI\_15 model using support vector machine had low accuracy of 82.78%. This result suggested that reduced HSI models were useful for building citrus diseases detection system.

In general, back-propagation based on neural network method has good performance above 93% for both diseases conditions. The support vector machine had good classification results of 95% in citrus canker, but poor accuracy of 82.78% in citrus greening. Back-propagation method can be a good application for disease detection. This method showed good performance and the algorithms also was simple to implement for a detection system. For future study, these methods will be evaluated on other citrus disease condition as well as under outdoor lighting conditions. In addition, the stability of these algorithms will be tested. In conclusion, this research demonstrated that color imaging texture feature analysis, and pattern recognition could be used in the laboratory to classify citrus disease conditions.

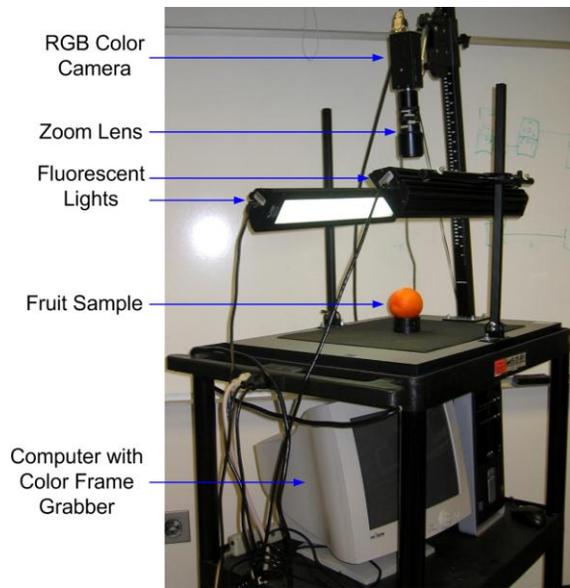


Figure 6-1. Color image system for acquiring RGB images from citrus samples

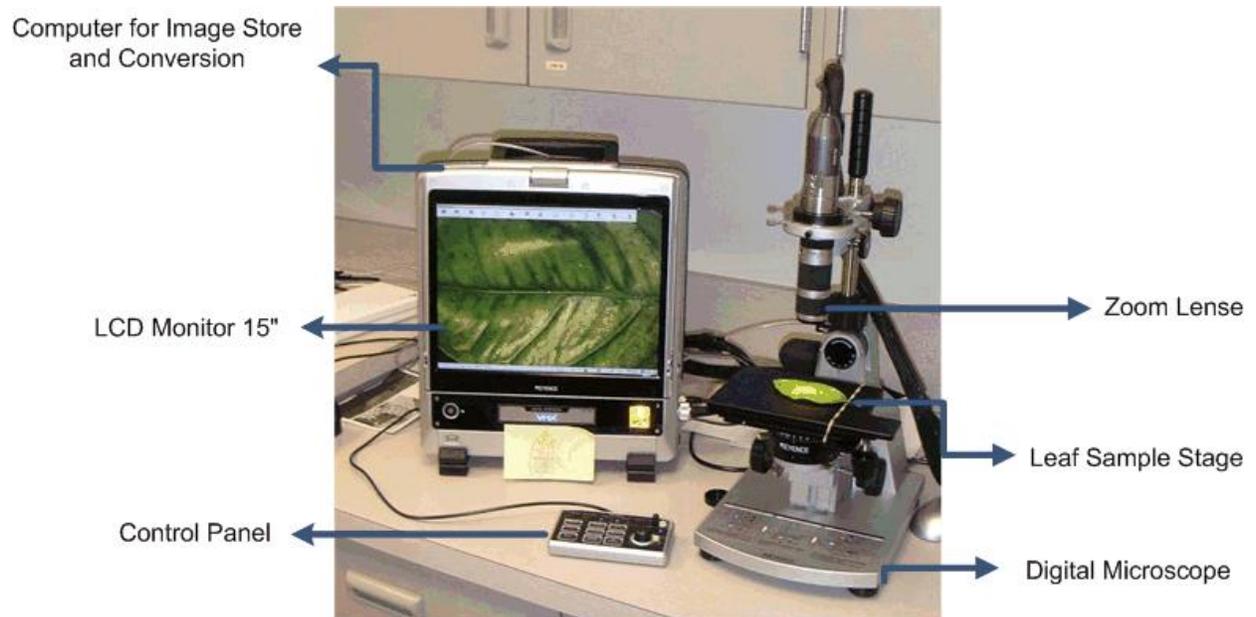


Figure 6-2. Digital microscope system for acquiring RGB images from citrus disease samples

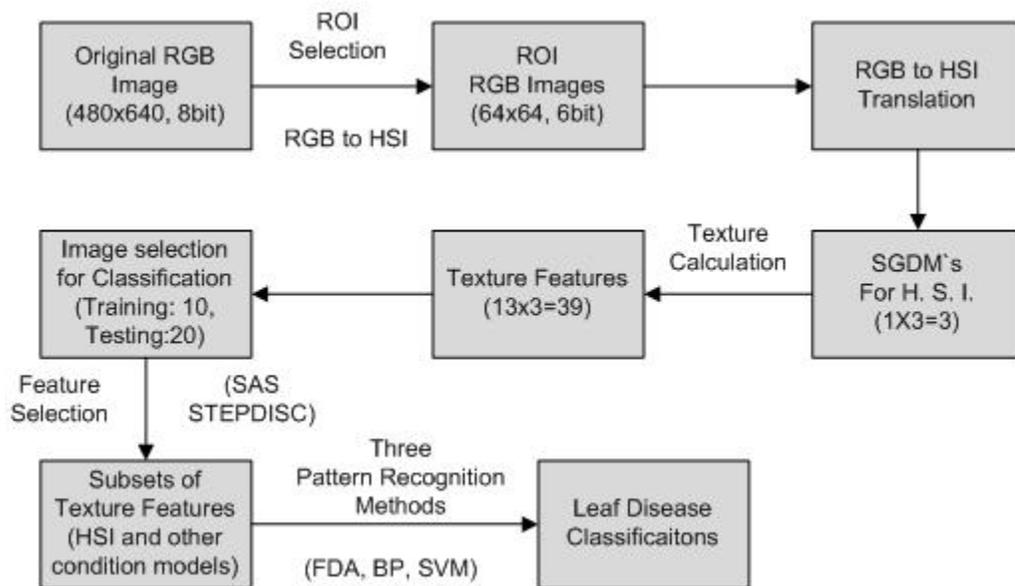


Figure 6-3. Procedures for color image analysis for citrus canker

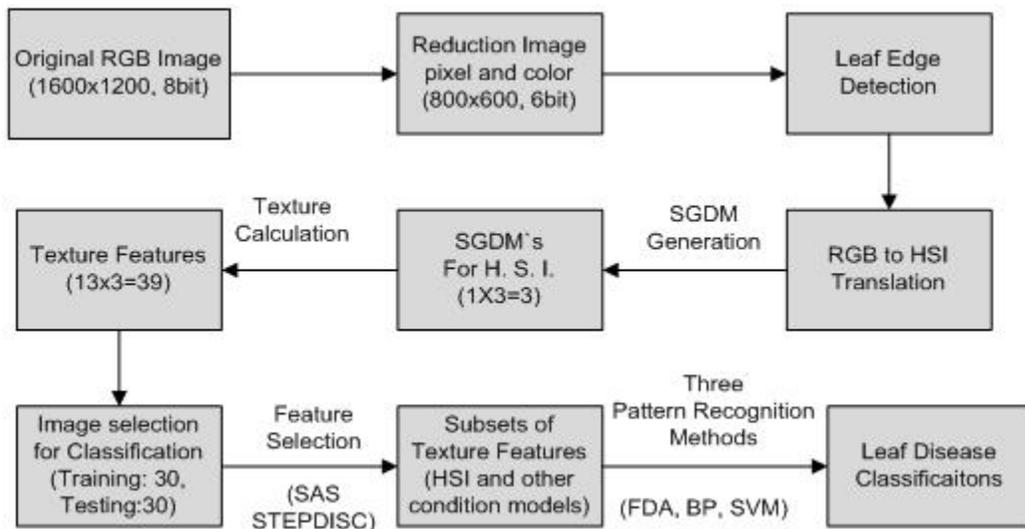


Figure 6-4. Procedures for color image analysis for citrus greening

Table 6-1. Texture feature models selected by SAS stepwise analysis for citrus canker

Classification Model	Color Feature	Texture Feature Set
HSI_13	H, S, I	H <sub>9</sub> , H <sub>10</sub> , I <sub>12</sub> , S <sub>7</sub> , I <sub>3</sub> , I <sub>2</sub> , S <sub>12</sub> , I <sub>11</sub> , I <sub>1</sub> , I <sub>8</sub> , S <sub>1</sub> , H <sub>2</sub> , H <sub>5</sub>
HS_9	H, S	H <sub>9</sub> , H <sub>10</sub> , S <sub>7</sub> , H <sub>5</sub> , H <sub>11</sub> , S <sub>12</sub> , S <sub>11</sub> , H <sub>7</sub> , H <sub>13</sub>
I_11	I	I <sub>2</sub> , I <sub>3</sub> , I <sub>5</sub> , I <sub>10</sub> , I <sub>6</sub> , I <sub>13</sub> , I <sub>8</sub> , I <sub>9</sub> , I <sub>1</sub> , I <sub>11</sub> , I <sub>7</sub>
HSI_39	H, S, I	All 39 texture features (H <sub>1</sub> - H <sub>13</sub> , S <sub>1</sub> - S <sub>13</sub> , I <sub>1</sub> - I <sub>13</sub> )

Table 6-2. Canker disease classification results in percent correct for all models using FDA.

Peel Condition	Classification Model			
	HSI_13	HS_9	I_11	HSI_39
Normal	90.00	90.00	80.00	80.00
Canker	90.00	90.00	90.00	100.00
Copper Burn	100.00	60.00	100.00	100.00
Greasy Spot	90.00	90.00	100.00	90.00
Melanose	80.00	70.00	70.00	90.00
Wind Scar	100.00	90.00	10.00	100.00
Overall Accuracy (%)	91.67	81.67	75.00	93.33

Table 6-3. Canker disease classification results in percent correct for all models using BP neural network

Peel Condition	Classification Model			
	HSI_13	HS_9	I_11	HSI_39
Normal	100.00	90.00	80.00	100.00
Canker	100.00	70.00	90.00	90.00
Copper Burn	100.00	70.00	100.00	100.00
Greasy Spot	100.00	90.00	100.00	90.00
Melanose	70.00	60.00	70.00	90.00
Wind Scar	100.00	100.00	20.00	90.00
Overall Accuracy (%)	95.00	80.00	76.67	93.33

Table 6-4. Canker disease classification results in percent correct for all models using SVMs

Peel Condition	Classification Model			
	HSI_13	HS_9	I_11	HSI_39
Normal	100.00	90.00	40.00	90.00
Canker	100.00	90.00	90.00	50.00
Copper Burn	100.00	80.00	100.00	30.00
Greasy Spot	90.00	70.00	100.00	80.00
Melanose	80.00	50.00	50.00	90.00
Wind Scar	100.00	10.00	30.00	100.00
Overall Accuracy (%)	95.00	65.00	68.33	73.33

Table 6-5. Texture feature models selected by SAS stepwise analysis for citrus greening

Classification Model	Color Feature	Texture Feature Set
HSI_15	H, S, I	S <sub>5</sub> , I <sub>2</sub> , H <sub>7</sub> , H <sub>2</sub> , S <sub>6</sub> , S <sub>4</sub> , H <sub>9</sub> , S <sub>8</sub> , I <sub>6</sub> , S <sub>13</sub> , H <sub>4</sub> , I <sub>4</sub> , I <sub>13</sub> , S <sub>7</sub> , I <sub>7</sub>
HS_10	H, S	S <sub>5</sub> , H <sub>7</sub> , H <sub>5</sub> , H <sub>12</sub> , S <sub>4</sub> , S <sub>7</sub> , H <sub>8</sub> , S <sub>8</sub> , H <sub>3</sub> , S <sub>11</sub>
I_8	I	I <sub>2</sub> , I <sub>8</sub> , I <sub>9</sub> , I <sub>6</sub> , I <sub>5</sub> , I <sub>7</sub> , I <sub>10</sub> , I <sub>1</sub>
HSI_39	H, S, I	All 39 texture features (H <sub>1</sub> - H <sub>13</sub> , S <sub>1</sub> - S <sub>13</sub> , I <sub>1</sub> - I <sub>13</sub> )

Table 6-6. Citrus greening classification results in percent correct for all models using FDA

Peel Condition	Classification Model			
	HSI_15	HS_9	I_11	HSI_39
Blotchy mottle	83.33	63.33	53.33	93.33
Islands	93.33	80.00	76.67	90.00
Iron deficiency	100.00	100.00	86.67	93.33
MN deficiency	100.00	83.33	86.67	96.67
Zinc deficiency	100.00	100.00	86.67	96.67
Normal	96.67	93.33	73.33	93.33
Overall Accuracy (%)	95.55	86.67	77.22	93.89

Table 6-7. Citrus greening classification results in percent correct for all models using BP

Peel Condition	Classification Model			
	HSI_15	HS_9	I_11	HSI_39
Blotchy mottle	83.33	73.33	66.67	80.00
Islands	86.67	70.00	70.00	100.00
Iron deficiency	100.00	96.67	90.00	100.00
MN deficiency	96.67	86.67	86.67	96.67
Zinc deficiency	100.00	96.67	83.33	93.33
Normal	93.33	83.33	73.33	93.33
Overall Accuracy (%)	93.33	84.45	78.33	93.89

Table 6-8. Citrus greening classification results in percent correct for all models using SVM

Peel Condition	Classification Model			
	HSI_15	HS_9	I_11	HSI_39
Blotchy mottle	60.00	63.33	46.67	60.00
Islands	90.00	80.00	66.67	36.67
Iron deficiency	100.00	100.00	86.67	66.67
MN deficiency	80.00	83.33	66.67	43.33
Zinc deficiency	70.00	93.33	50.00	76.67
Normal	96.67	86.67	80.00	36.67
Overall Accuracy (%)	82.78	84.44	66.11	53.34

## CHAPTER 7 CLASSIFICATION METHODS OF CITRUS GREENING BY HYPERSPECTRAL IMAGING

Citrus greening, also known as Huanglongbing (HLB), is one of the most destructive diseases in citrus. This disease can cause tree decline, death, yield loss and lost marketability. Once citrus trees are infected, the fruit yield, and quality are greatly reduced. The trees also become susceptible to other diseases and health problems. Furthermore, the fruits from the infected trees are unmarketable and not suitable for juice processing due to acidity and bitter taste (Polek et al., 2007). During the recent past, citrus greening has become a serious threat in the citrus-growing regions of Florida. In addition, farmers are concerned about huge costs from tree loss, scouting, and chemicals used in an attempt to control the disease. Florida citrus growers are desperately fighting this plant disease which has the potential to destroy the state's \$9 billion commercial citrus industry (The American Phytopathological Society, 2008).

Citrus greening is a bacterial disease that affects the phloem system of citrus trees and causes the leaves of infected trees to become yellow, the trees to become unproductive, decline and possibly die within a few years. The bacterium is spread by an insect, the citrus psyllid. Citrus greening infects all types of citrus species, cultivars, and hybrids and some citrus relatives. The symptoms of citrus greening usually include blotchy, chlorotic mottling of leaves, yellow shoots, misshapen or lopsided small fruit that fail to color properly and stay green, hence the name of the disease "greening". The name Huanglongbing means "yellow shoot", which is descriptive of the yellow sectors of infected trees (Gottwald et al., 2007). Unfortunately, there is no treatment and prevention for citrus greening in the infected trees. Nevertheless, early detection of the

disease and appropriate management of the insect vector should alleviate the severity of the greening disease and minimize its spread.

Crops become stressed when any infections, such as viral infections, or physiological factors, such as air pollution, adversely affect growth, development, and yield. These stresses are expressed in various ways. For example, problems in the water balance control can slow down photosynthesis, reduce evapotranspiration, and raise leaf surface temperature (Nilsson, 1995). Other symptoms include morphological changes such as leaf curling, change in leaf angle, wilting, chlorosis, discoloration of leaves and fruits and premature drop of fruits. Because of these observable symptoms, humans could easily assess the conditions of their crops. However, the large size of current farms and the decrease of farm labor make it impossible to assess the whole field. Furthermore, the symptoms of some diseases do not show up in the visible range (400 -700 nm), which is the sensitivity of the human eyes. But with the advancement of optical sensing technologies (such as machine vision and spectroscopy), several researchers have shown the potential of these technologies to identify some of these diseases and conditions. Specifically, these technologies include visible imaging (400-700 nm), near infrared imaging (greater than 700 nm), multispectral imaging (a combination of wavelengths in the visible and near infrared), and hyperspectral imaging (400-1000 nm). Camargo et al. (2009) developed a visible imaging system and color image processing to identify banana leaves that were infected with Black Sigatoka. Color images of infected banana leaves were used to develop the image processing which includes color transformation, histogram multi-thresholding, and segmentation. They demonstrated that the algorithm was able to identify the diseased regions in most

of the images tested. A machine vision system in the visible region was used by Blasco et al. (2003) to detect external blemishes in apples and peaches. They achieved 93% accuracy in blemish detection. Visible imaging systems are sufficient if the symptoms occur in the visible region, but most of the diseases and conditions could be identified more effectively by exploiting regions beyond the visible. Zarco-Tejada et al. (2005) used a compact airborne spectrographic imager and hyperspectral sensors to monitor the temporal and spatial chlorosis condition of *Vitis vinifera* L. Their results showed that they could estimate chlorophyll content in the narrow band hyperspectral indices calculated in the 700-750nm. A ground-based hyperspectral imaging system for characterizing vegetation spectral features was developed by Ye et al. (2008). They used a hyperspectral line sensor with a wavelength range of 360-1010 nm to demonstrate the potential of the system to monitor vegetation variability in crop systems. Changes in spectral signatures due to deficiencies of nutrients and damage by pests and environmental factors have been reported (Sepulcre-Canto et al., 2007, Muhammed, 2005, and Gowen et al., 2007). In addition, the spectral signatures is influenced by the amount of pigments, leaf angle, leaf surface texture, diseases and stress, plant growth stage, and measurement condition.

While researchers have demonstrated the potential of imaging technologies for disease detection, not much work can be found on the automatic detection citrus greening. Mishra et al. (2007) investigated the spectral characteristics of leaves infected with HLB and normal leaves using a spectroradiometer (350-2500 nm). The spectral bands of green to red wavelength and the near infrared band were found to have the potential to discriminate HLB-infected leaves from normal leaves by using

discriminability, spectral derivative analysis, and spectral ratio analysis. These wavelengths include 530-564 nm, 710-715 nm, 1041 nm, and 2014 nm. In addition, they also developed a four-band active optic sensor to identify young leaf flushes to control the spot spraying of citrus trees. It is known that the Asian citrus psyllid, vector of HLB disease feeds on young leaves. Although they have characterized the spectral reflectance property of HLB infected leaves and discriminate it from normal leaves, other conditions such as nutrient deficiency of leaves could look similar to HLB infected leaves.

In this study, citrus leaves infected with citrus greening (i.e., blotch mottle and greening islands), leaves exhibiting nutrient deficiencies symptoms (i.e., iron, magnesium, and zinc deficiency), and healthy leaves (i.e., normal young and normal mature) were collected in the field. Hyperspectral images in the spectral region between 400 and 900nm of the samples were acquired. The main objective of this research was to develop image processing classification algorithms based on principal component analysis (PCA) and wavelet transform for the detection of citrus greening in the presence of confounding leaf symptoms.

## **Materials and Methods**

### **Citrus Leaf Samples**

Table 7-1 demonstrated the sample number of the citrus greening and no citrus greening conditions for this study. Fifty samples of each of the citrus greening conditions and forty samples of each of the other conditions were collected. Therefore, a total of 300 samples were selected in this study. The leaf samples were clipped with petioles intact and then sealed in Ziploc bags to maintain the moisture level of the leaves.

## Hyperspectral Image Acquisition

A hyperspectral line-scan imaging system, as shown in Figure 7-1, was used for acquiring spectral images from citrus leaf samples. The imaging system consisted of an electron-multiplying charge-coupled-device (EMCCD) camera (Luca, Andor Technology Inc., CT, USA) with imaging spectrograph (ImSpector V10E, Spectral Imaging Ltd., Oulu, Finland) and a C-mount lens (Rainbow CCTV S6X11, International Space Optics, S.A., Irvine, CA, USA), a pair of halogen line lamp (21V, 150W) powered with a DC voltage regulated power supply (Dolan-Jenner Industries, Inc., Lawrence, MA, USA), and a programmable motorized positioning table (XN10-Xslide, Velmex Inc., Bloomfield, NY, USA). This equipment was placed inside in a dark box to eliminate stray external light.

The EMMCCD has 1004x1002 pixels and a double-stage Peltier device to cool to  $-20^{\circ}\text{C}$ . An imaging spectrograph is based on prism-grating- prism principle. It has a slit 2.8 nm long and 30  $\mu\text{m}$  wide. Through the slit, light from the scanned line is dispersed by a prism-grating- prism device and projected onto the pixels of EMCCD detector. A two-dimensional image is generated with the spatial dimension along the horizontal axis and the spectral dimension along the vertical axis of the EMCCD. When the citrus sample is moved perpendicularly to the scanning direction by the motorized positioning table, one thousand seven hundred and forty line scans were performed for each leaf sample, and four hundred pixels covering the scene of the sample at each scan were saved. Therefore, a three-dimensional image cube (1740x400 for each band) was created.

The hyperspectral imaging software to transfer data and parameterization was developed using the Andor Software Development Kit (SDK, Luca, Andor Technology Inc., CT, USA) for the hyperspectral line scan imaging system. An Hg-Ne spectral calibration lamp (Oriel Instruments, Stratford, CT, USA) was used to investigate spectral calibration of the system. Because of low light output in the visible region less than 450 nm, and low quantum efficiency of the EMCCD in the NIR region beyond 930 nm, the wavelength range between 451.67 nm and 927.71 nm was used (totaling 92 bands with a spectral resolution of 5.2 nm).

## **Hyperspectral Image Analysis**

### **Hyperspectral image transformation using PCA**

Since the input images consist of 3-D image data ( $106 \times 176 \times 92$  bands), it is necessary to reduce the spectral dimension of the hyperspectral images. The data compression technique selected for this study was principal component analysis (PCA). PCA is a useful tool to reduce dimensionality of a data and extract information from original data, and maintain the useful information of the original data. The central idea of PCA method is to rotate the data on the principal axes using orthogonal transformation that best presents given data in least-square sense and to reduce the dimensionality of a given data. The axes can be obtained from the calculation of the eigenvectors of the covariance matrix from an original data (Duda et al., 2001).

For the hyperspectral images, the covariance matrix of the image is assembled, and the eigenvalues and eigenvectors are determined from the covariance matrix. Using the eigenvectors of covariance matrix, the principal components are formed. The principal components can be decomposed into two matrices; loading matrix and scores matrix. The loading matrix is the weight coefficients for each original variable when

calculating the principal component. The scores matrix includes the principal component data rotated from the original data. Therefore, the 3-D hyperspectral data with 92 single-band images is transformed into the 2-D principal component score image in this study. PCA procedures described above were done using Matlab 7.0 (The Mathworks, Inc., Natick, Mass.).

### **Citrus greening classification algorithm using PCA**

Fifty samples for each greening condition and forty samples for each non-greening condition were selected, and therefore a total of 300 citrus leaves were tested in this study. Greening islands and blotch mottle are included in the greening disease condition and Manganese deficiency, Iron deficiency, Zinc deficiency, normal young and normal mature are contained in the non-greening disease condition.

In the PCA transformation, only the first five principal components (PC) were used since the higher PCs did not demonstrate meaningful features and consisted mostly of noise. It was found through visual observation of the PC score images that PC-3 enhanced the citrus greening leaves.

A detailed illustration of the image processing and classification is given in Figure 7-2. After generating the masked image obtained from a single band image at around 800 nm, the images were transformed into PC-3 score images by performing PCA. To separate the citrus greening spots from the leaf surfaces, a thresholding-based segmentation was applied to PC-3. In bi-level thresholding, the pixel scale less than threshold value are set as "1" (greening disease condition) while others are set as "0"(non-greening disease condition and back ground). The threshold value can be found from image histograms in Figure 7-3. The value of 0.79 provided the best overall classification accuracy. After image thresholding, some defects might be present in the

images. As shown figure 7-4, the image-opening morphological filtering can remove defects on the segmented images. This post-processing might reduce the misclassification of citrus greening diseases.

### **Classification algorithm using wavelet transform**

The algorithms shown in Figure 7-5 illustrated procedures of hyperspectral image processing and classification. After generating PC score image using PCA, The image texture features were extracted from each decomposed image. Figure 7-6 showed four-level decomposed images in each condition. The texture feature dataset was generated by containing 13 texture features for each image. The dataset had 300 rows each, representing 50 samples from each of the two conditions of citrus greening symptom, and 40 samples from each of the five conditions of no citrus greening symptom. Each row had 13 columns representing the 13 texture features extracted for a particular sample image. Each row had a unique number (1, 2) representing which class the particular row of data belonged. '1' represented a citrus greening symptom, '2' represented a no citrus greening symptom. To compare classification accuracies under citrus greening symptoms with non-citrus greening symptoms, two classes were created which are shown in Table 7-1. These classes represent the compilation of seven different leaf conditions sets, which isolate leaf conditions that are difficult to discriminate. This dataset were classified with fisher linear discriminant analysis.

All analysis was done using MATLAB R2007b (The Mathworks, Inc., Natick, Mass.) and the MTLAB PRTools toolbox (Faculty of Applied Physics, Delft University of Technology, The Netherlands). PRTools is a toolkit for the MATLAB. This package is released for the developed and evaluation of pattern recognition algorithms.

## **Results and Discussion**

### **The Plot of Relative Reflectance Spectra**

The reflectance mean spectra and corresponding standard deviation of the samples with 7 different symptoms over the wavelength range between 451.67 nm and 927.71 nm are represented in Figures 7-7 and 7-8. The reflectance spectra were calculated using the mean of 10 spectra from 10 randomly selected samples. Each plot in Figure 7-7 is generated by the average ROI window covering 3 x 3 pixels extracted from each hyperspectral image.

In mean spectra obtained from the hyperspectral images of citrus greening with other conditions, the chlorophyll absorption peaks of wavelength from 400 nm (blue) to 600 nm (green) appear to be higher and more prominent. The relative reflectance increase steadily over the wavelengths 720 - 900 nm, but the reflectance rate is low (10 -30%). Reflectance spectra from hyperspectral images was extracted using ENVI 4.3 (ITT Visual Information Solutions, Boulder, CO, USA)

### **Classification of Citrus Greening using PCA**

Once the space between citrus greening and other surface condition in all regions was narrowed and the difference between conditions is not evident, the specified region in plot would not help to identify citrus greening. Moreover, If PCA was performed in the entire spectra region, a more uniform surface background in the score image was created and the possibility of converting the normal areas on the leaf surface to the citrus greening conditions in the binary images was reduced. Therefore, the entire regions in reflectance spectra would give a better classification and the performance of PCA.

As observed in Figure 7-9, the PC-3 score images with disease conditions have brighter spots and higher contrast on the leaf surface. Leaf samples with normal mature and young conditions do not have the contrasting regions because of the absence of disease. When the two citrus greening conditions (i.e., greening islands and blotch mottle) were compared, the leaf surface including the blotch mottle spots appear dark in the PC-2 score image while the greening islands spots. The other score images, PC-1, PC-4 and PC-5 have low contrast between the diseased and the non-diseased surface. Thus, due to distinctive patterns and high contrast, the PC-3 score images were used for the classification of greening islands and blotch mottle, and they were independently evaluated for classification between citrus greening and non citrus greening.

The results shown in Table 7-2 is the classification summary for differentiating citrus greening from other nutrient deficiency and normal conditions using PC-3 score image. As previously indicated, the test data consisted of 50 images from each citrus greening and 40 images from others. The overall performance was 94.33% for all the samples. The classification accuracy for the 'Greening Disease' class was 96%, and for the 'No Greening Disease' class was 93.5%, respectively. Based on the results shown in Table 7-2, four samples were misclassified in the 'Greening Disease' classes. Each 'Blotch Mottle' and 'Greening Islands' symptom had two misclassified samples. There were 17 samples that were misclassified in the 'No Greening Disease' class which included four 'Manganese Deficiency' symptoms, two 'Zinc Deficiency' symptoms, three 'Normal Young' symptoms, and four 'Normal Mature' symptoms. This misclassification demonstrated the confusion in classifying citrus greening between manganese deficiency, zinc deficiency, normal young, and normal mature. In particular, manganese

deficiency and normal mature samples had the lowest classification (90%) On the other hand, all 'Iron Deficiency' samples were correctly classified. The plot of reflectance spectra in Figure 7-8 showed that iron deficiency was distinct to citrus greening, while the other conditions were very close and overlap each other. This misclassification may have been caused by the similarity of the reflectance spectra between citrus greening and other conditions. However, the high classification rate of PCA classification have shown that transforming the reflectance spectra increased the disparity between citrus greening and other confounding conditions.

### **Classification of Citrus Greening using Wavelet Transform**

The results using Wavelet Transform algorithms had a good overall accuracy. The overall performance shown in Table 7-3 has accuracy of 93%. For the other two classes (citrus greening and no citrus greening), there were two misclassified sample for 'Citrus Greening' class, and eight misclassified samples for 'No Citrus Greening' class. The classification accuracies for 'Citrus Greening' class, and 'No Citrus Greening' class were 95% and 91%, respectively. In citrus greening class, the classification accuracies for 'greening islands' and 'blotch mottle' symptoms were 95%. In general, there was only one sample that was misclassified in each 30 test set. For the other class (i.e., non-citrus greening), there were nine misclassified samples for this class. Two manganese deficiency samples, six zinc deficiency samples, and one normal young sample were misclassified as greening disease class. Iron deficiency and normal mature were correctly classified as non-citrus greening disease class. Although it achieved classification accuracies of above 90% and two perfect classification results (100%), the performances for zinc deficiency condition was poor (70%). It is means that texture features from zinc deficiency was similar to citrus greening class in wavelet transform.

## Summary

The early detection of citrus greening disease could be very useful in managing and controlling the infection. The challenge of identifying citrus greening is the classification of nutrient deficient symptoms as citrus greening. In this study, a hyperspectral imaging system was developed to recognize leaves exhibiting symptoms of citrus greening along with leaves with nutrient deficiency and normal leaves. Seven leaf classes were evaluated; 1) normal young, 2) normal mature, 3) blotchy mottle (i.e., citrus greening), 4) green islands (i.e., citrus greening), 5) manganese deficiency, 6) iron deficiency, and 7) zinc deficiency. The leaf samples were collected in the field and hyperspectral images of the leaf samples were taken in the laboratory with the spectral range of 451.67 nm to 927.71 nm.

In the image processing algorithm, two classification approaches were developed. First, Principal Component Analysis (PCA) was used to transform the hyperspectral images to principal component score image. It was found that the third principal component images enhanced the two citrus greening symptoms from the other classes. By using a simple thresholding method to the principal score image, the citrus greening samples were classified. The second approach used the wavelet transform to extract 13 texture features from the principal component score image. A linear discriminant classifier based on the texture features was computed to classify the citrus greening samples.

Results showed that the PCA-thresholding approach had a total classification rate of 94.33%, while 93% total classification rate was obtained for the Wavelet Transform classification. This indicates that although the spectral reflectance profile of citrus greening and nutrient deficiency are near to each other, utilizing PCA to increase the

variance proved to be useful. Furthermore, the texture features obtained from the wavelet transform showed that the morphological properties of citrus greening can be used to differentiate from nutrient deficient leaf samples.

This research demonstrated that hyperspectral imaging combined with an appropriate image processing algorithm such as PCA and wavelet transform could be used for differentiating citrus greening symptoms from other leaf conditions, specifically nutrient deficiency. Future studies will explore the utility of these algorithms in outdoor conditions.

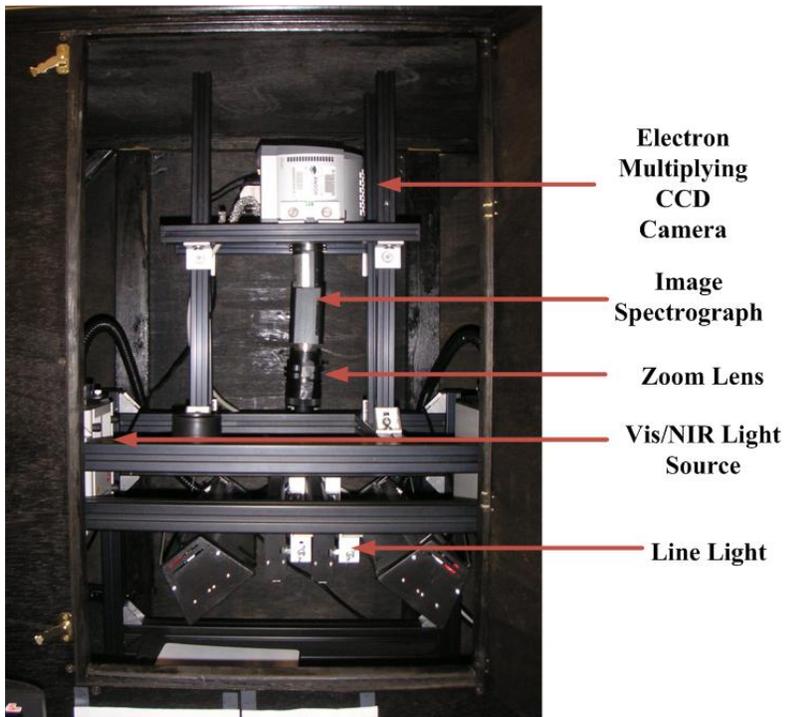


Figure 7-1. Hyperspectral line-scan imaging system

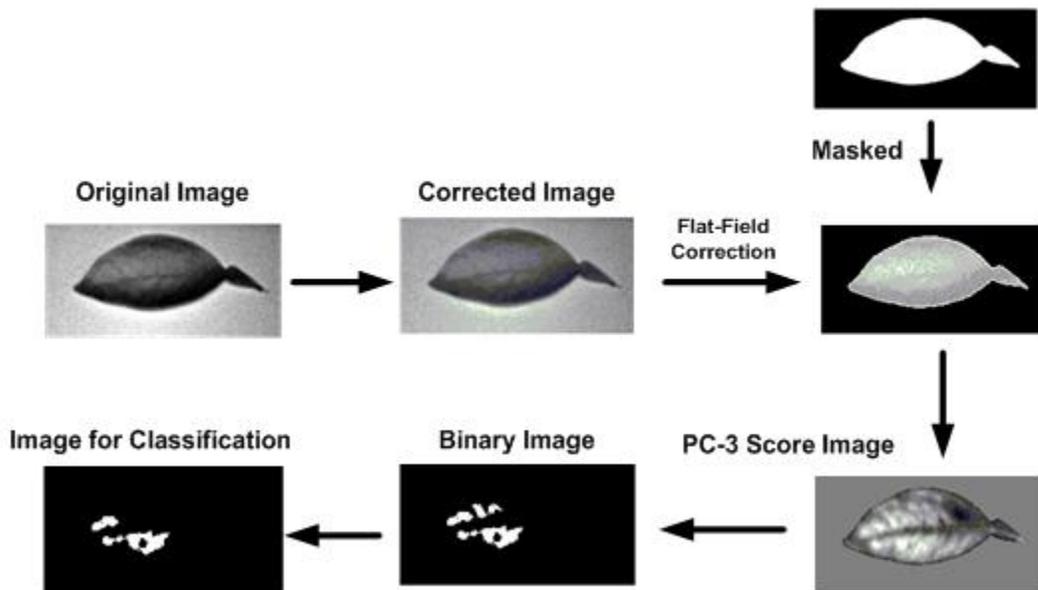


Figure 7-2. A detailed illustration of Threshold based on classification algorithms

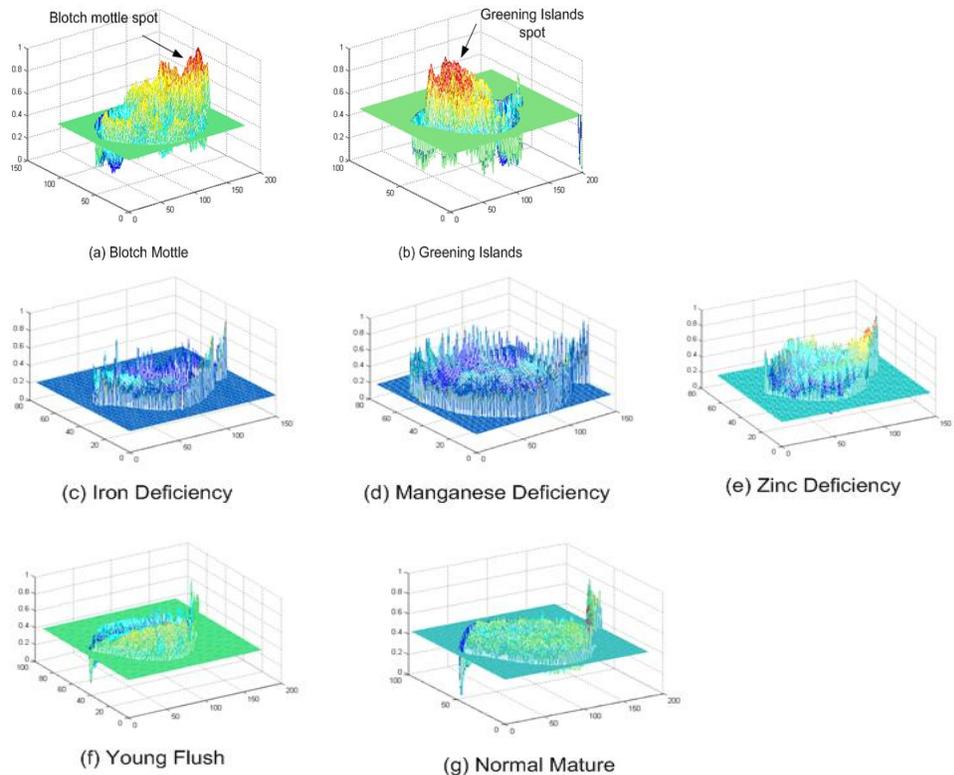


Figure 7-3. Comparison of the threshold value distribution between citrus greening and other conditions

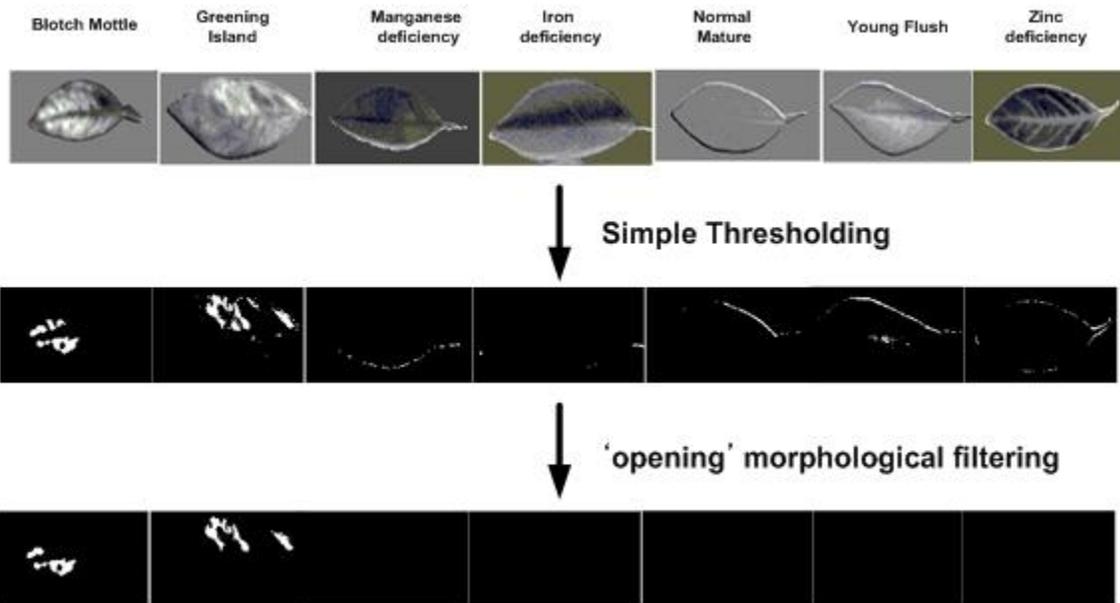


Figure 7-4. The image-opening morphological filtering

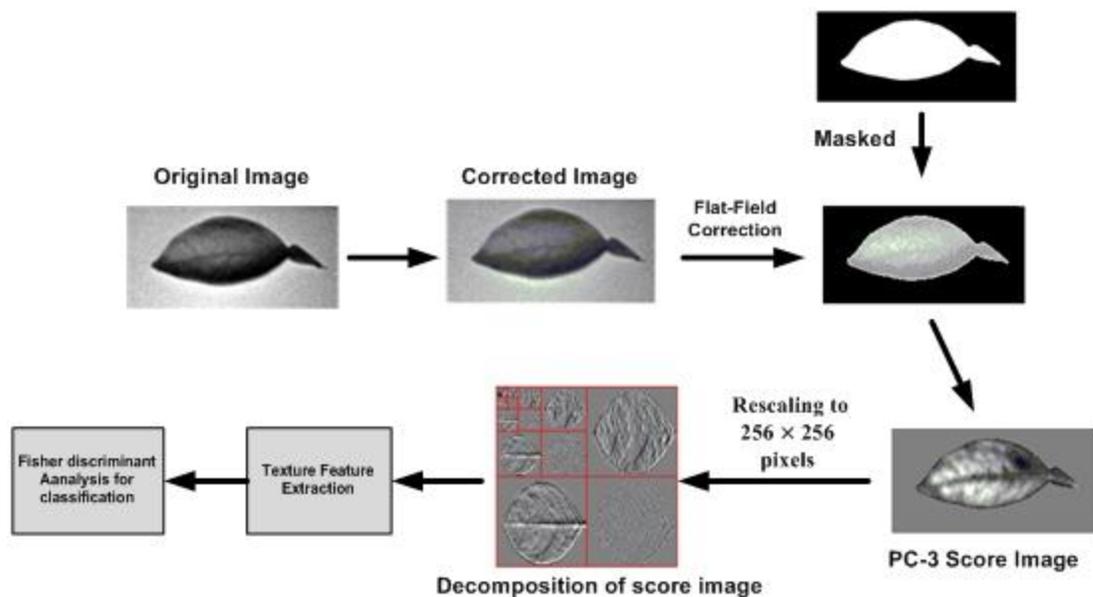


Figure 7-5. A detailed illustration of the feature extraction using Wavelet Transform

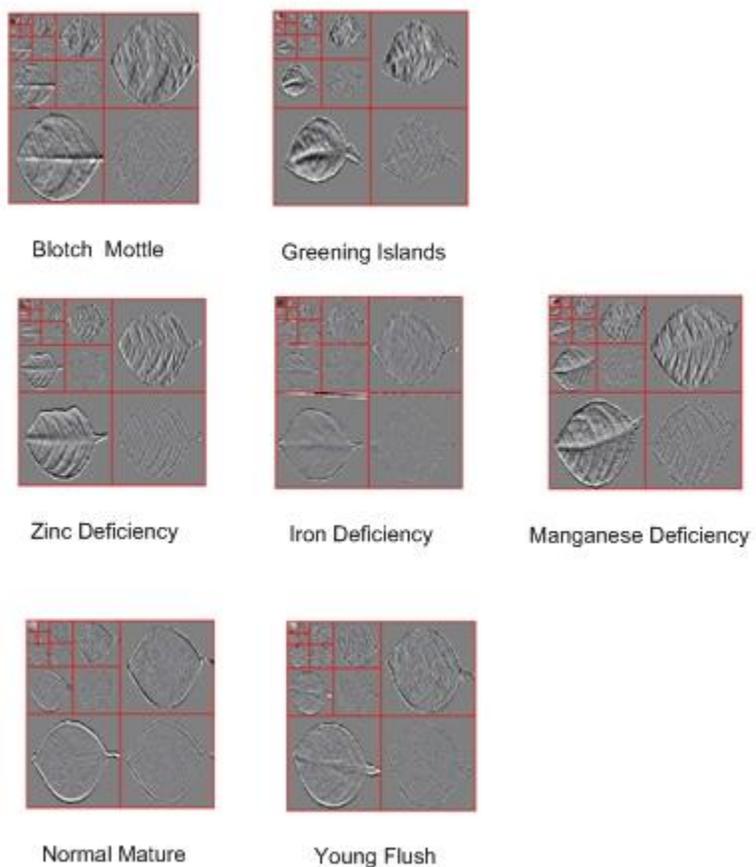


Figure 7-6. Four-level decomposed images in each condition.

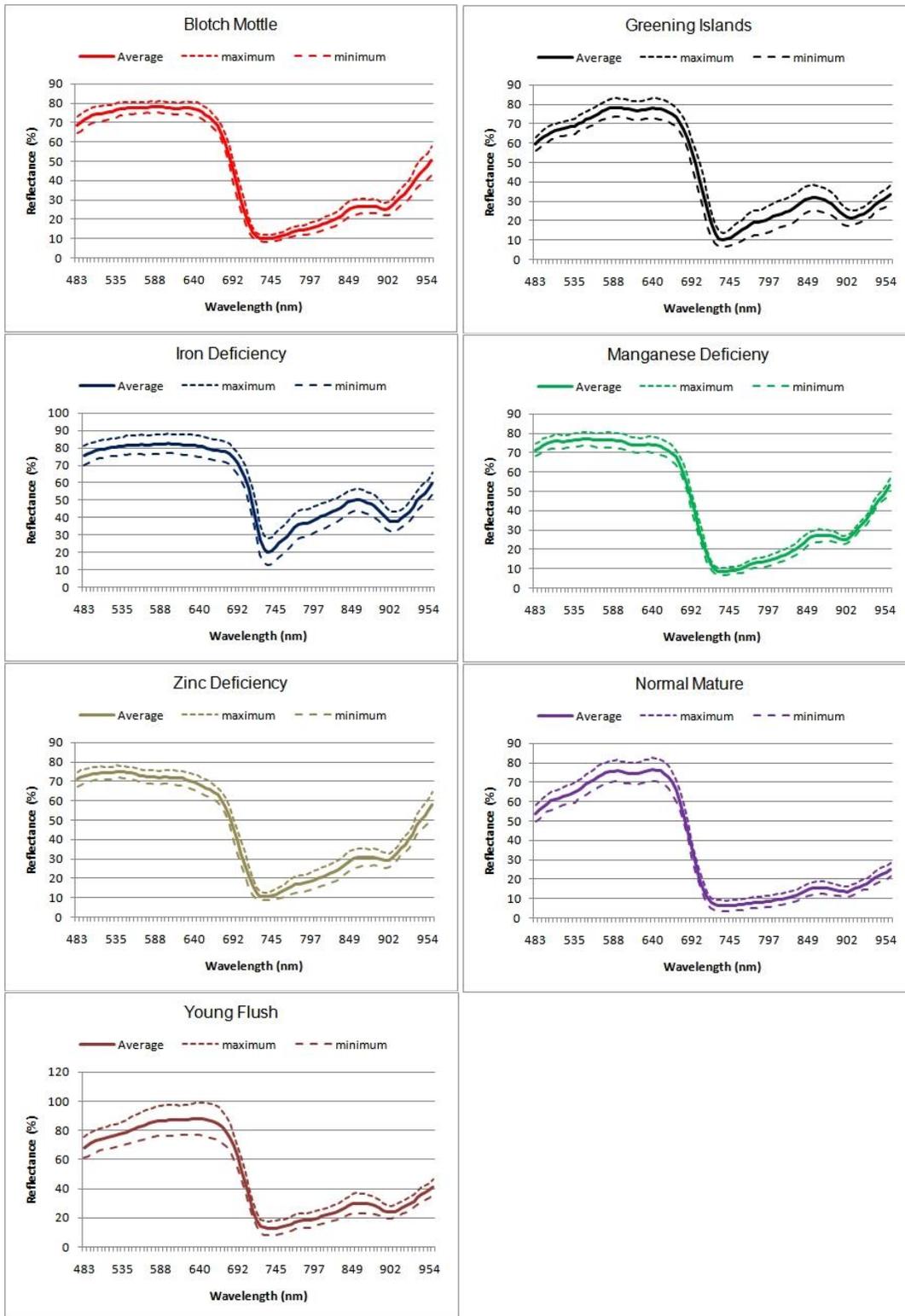


Figure 7-7. Reflectance mean spectra and corresponding standard deviation of each ROI from citrus greening and other symptoms

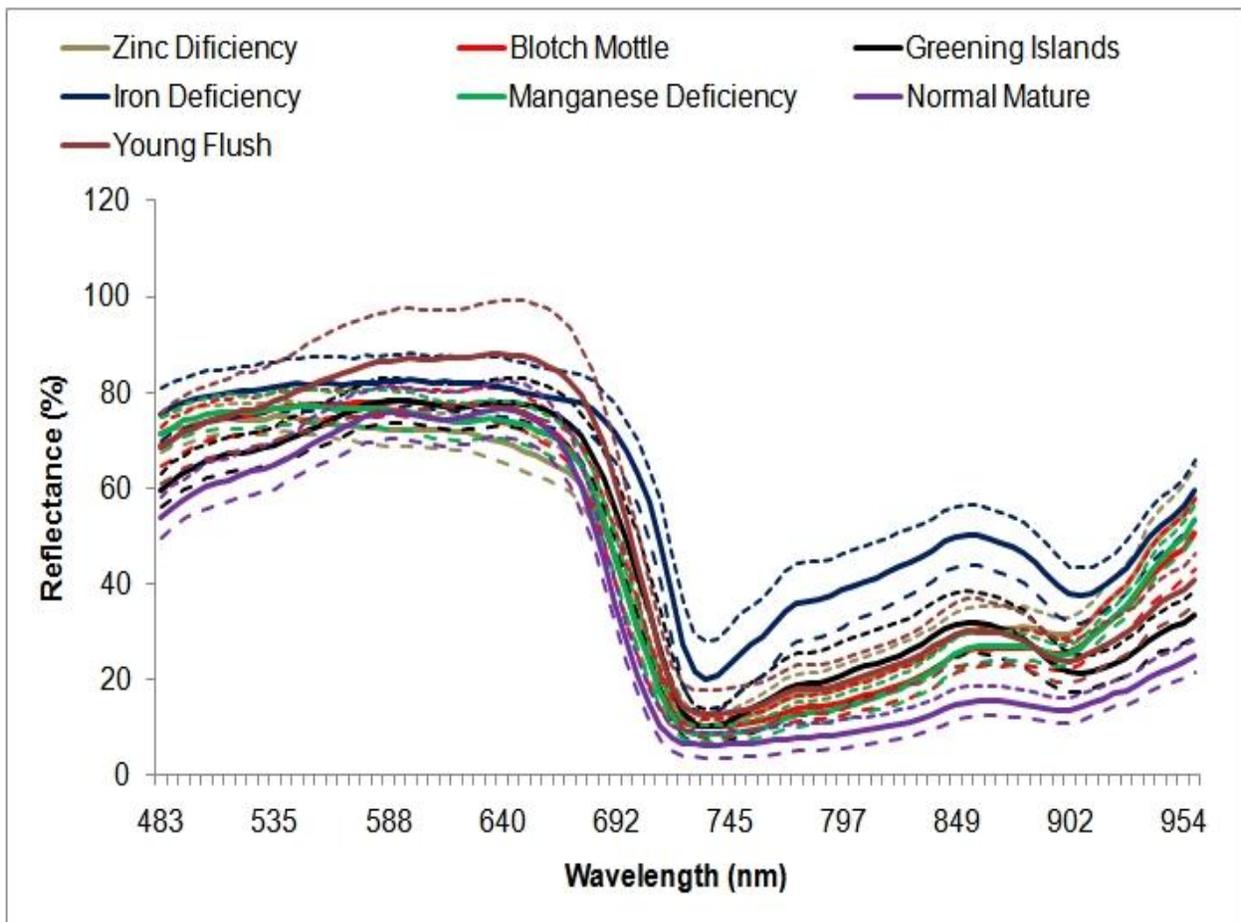


Figure 7-8. Reflectance of spectra of citrus leaf samples with citrus greening and other conditions

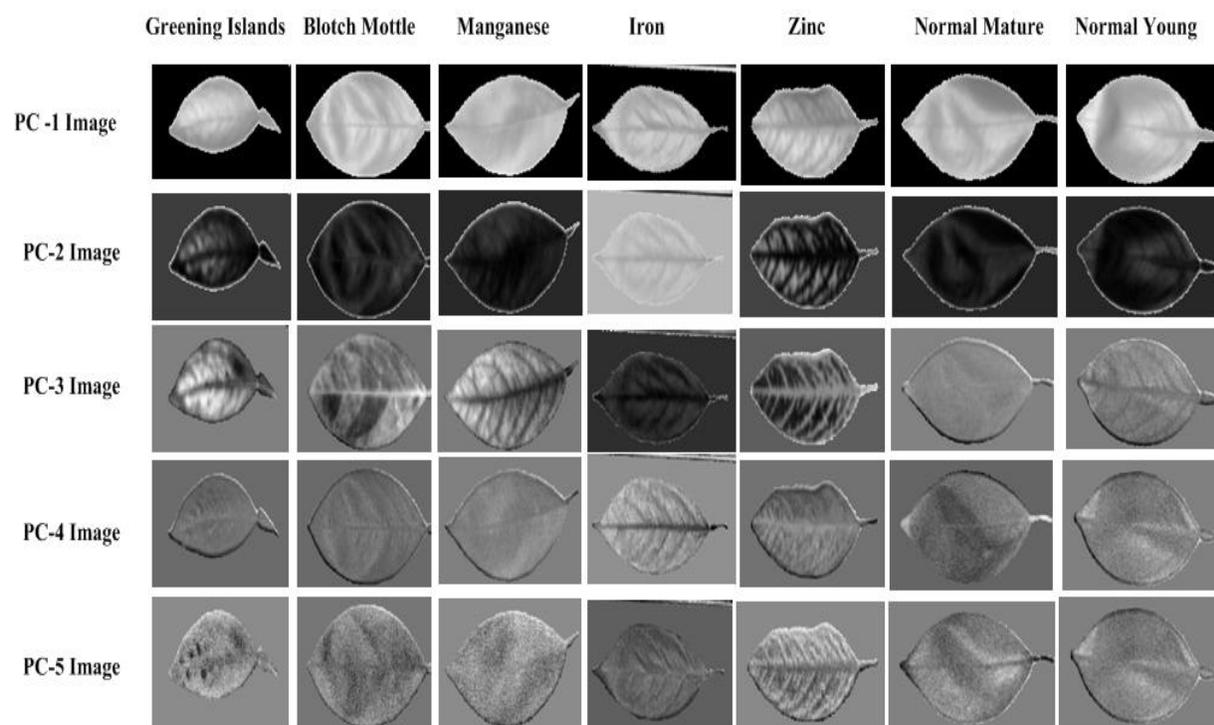


Figure 7-9. Five score images obtained from principal component analysis using full spectral regions

Table 7-1. The number of citrus greening and other conditions

Class	Symptoms	Number	Total
Greening Disease	Blotch Mottle	50	100
	Greening Islands	50	
No Greening Disease	Iron Deficiency	40	200
	Manganese Deficiency	40	
	Zinc Deficiency	40	
	Young Flush	40	
	Normal Mature	40	

Table 7-2. Classification result in percent correct using PCA based on method

Class	Symptoms	Misclassified	Accuracy (%)	
Greening Disease	Blotch Mottle	2	96.0	96.0%
	Greening Islands	2	96.0	
No Greening Disease	Iron Deficiency	0	100.0	93.5%
	Manganese Deficiency	4	90.0	
	Zinc Deficiency	2	95.0	
	Young Flush	3	92.5	
	Normal Mature	4	90.0	
Total		17	94.3%	

Table 7-3. Classification result in percent correct using WT based on method

Class	Symptoms	Misclassified	Accuracy (%)	
Greening Disease	Blotch Mottle	1	95.0	95.0%
	Greening Islands	1	95.0	
No Greening Disease	Iron Deficiency	0	100.0	91.0%
	Manganese Deficiency	2	90.0	
	Zinc Deficiency	6	70.0	
	Young Flush	1	95.0	
	Normal Mature	0	100.0	
Total		11	93.0%	

## CHAPTER 8 CITRUS BLACK SPOT DETECTION USING HYPERSPECTRAL IMAGING

Citrus black spot is the latest exotic disease to be discovered in Florida. This fungal disease poses a threat to Florida groves similar to citrus canker and greening because it is one of the most well-known fungal diseases worldwide. Once citrus trees are infected with citrus black spots, the fruit yield, and quality are greatly reduced. Furthermore, fruits infected with citrus black spots are not acceptable for fresh market (Chung et al., 2009). Therefore it is important to control this disease to achieve profitable production.

Citrus black spot can be identified by fruit symptoms which cause cosmetic lesions on the peel of fruit that are the most conspicuous symptom of infection (Dewdney, 2010). Fruit symptoms can be quite variable. Black spot lesions begin as small orange or red spots with black margins and enlarge to become necrotic lesions. Other symptoms of black spots include hard spot lesions, virulent spot, cracked spot, and false melanose, appear on the fruit surface (The Institute of Food and Agricultural Sciences Extension at University of Florida, 2010). Detecting fruits infected with black spots can help in controlling the spread of this disease specifically in areas that are black spot free. The design and implementation of technologies that can efficiently detect black spot disease will greatly aid in the control effort.

The identification of various crops and plant parts using machine vision and image processing techniques has been studied by several researchers. Jimenez et al. (2000) surveyed several computer vision approaches for locating fruit in trees. Regunathan and Lee (2005) identified fruit count and size using machine vision and an ultrasonic sensor. Burks et al. (2000) developed a method for classification of weed species using

color texture features and discriminate analysis. Tang et al. (1999) developed a texture-based weed classification method using Gabor wavelets and neural networks for real-time selective herbicide application. Pydipati et al (2006) identified citrus disease using the co-occurrence matrix method, (CCM) texture feature method, and a discriminant analysis. Du et al. (2006) described five different texture feature methods, including the common first-order gray-level statistics (FGLS), run length matrix (RLM), gray-level co-occurrence matrix (GLCM), fractal dimension (FD), and wavelet transform (WT) based method. After that, both simple correlation analysis and partial least squares regression (PLSR), to investigate tenderness of cooled pork ham.

In recent years, optical techniques have been used widely in the food processing and inspection application. In particular, hyperspectral imaging technologies have a growing interest in quality evolution and inspection of food and agricultural product (Sun, 2008). Previous research has showed hyperspectral imaging technologies and applications for agricultural products. Jiang et al. (2007) used hyper-spectral fluorescence imaging to analyze the differences between walnut shells and meat. Hyper-spectral fluorescence imaging system scanned samples at 79 different wavelengths ranging from 425nm to 775nm with 4.5 nm increments, and data redundancy was reduced by principal component analysis (PCA). Zhang et al. (2005) suggested a creative classification approach for distinguishing healthy and fungal infected wheat kernels during storage. The research showed the potential use of NIR hyper-spectral imaging in grain quality assessment. The research used NIR hyper-spectral imaging and support vector machine (SVM) for identifying the fungi that caused the infection. Kim et al. (2002) researched a method for using hyper-spectral data to

identify wavebands to be used in multispectral detection systems, and evaluated spatial and spectral responses of hyper-spectral reflectance images of fecal-contaminated apples. Lee et al (2005) used the hyper-spectral imaging technique to detect defects on apple peel after harvest. They developed a proper wavelength selection method for detecting the defects.

In hyperspectral image classification methods, a spectral angle mapper (SAM) and spectral information divergence (SID) classification that measures the spectral similarity between two spectra was applied to various agricultural products and systems. Park et al (2007) used SAM algorithms to detect fecal and ingesta contaminants on the surface of poultry carcasses. Qin et al (2009) introduced the detection of citrus canker using SID classification methods. Yang et al (2006) studied the SAM method to airborne hyperspectral imagery for mapping yield variability.

The overall objective of this research was to develop two spectral classification methods, spectral angle mapper (SAM) and spectral information divergence (SID), for detection of black spot by hyperspectral image.

## **Methodology**

### **Samples**

Table 8-1 demonstrated the sample number of the black spot and no black spot conditions for this study. Thus, a total of 525 samples were selected in this study. 135 samples of the black spot conditions, 90 samples of the greasy spot and market, and 105 samples of the melanose and windscar were collected.

### **Hyperspectral Image Acquisition**

A hyperspectral line-scan imaging system, as shown in Figure 8-1, was used for acquiring hyperspectral images of the fruit samples. The imaging system consisted of

an electron-multiplying charge-coupled-device (EMCCD) camera (Luca, Andor Technology Inc., CT, USA) with imaging spectrograph (ImSpector V10E, Spectral Imaging Ltd., Oulu, Finland) and a C-mount lens (Rainbow CCTV S6X11, International Space Optics, S.A., Irvine, CA, USA), a pair of halogen line lamp (21V, 150W) powered with a DC voltage regulated power supply (Dolan-Jenner Industries, Inc., Lawrence, MA, USA), and a programmable motorized positioning table (XN10-Xslide, Velmex Inc., Bloomfield, NY, USA). This equipment was placed inside in a dark box to eliminate stray external light.

The EMMCCD has 1004x1002 pixels and a double-stage Peltier device to cool to  $-20^{\circ}\text{C}$ . An imaging spectrograph is based on prism-grating- prism principle. It has a slit 2.8 nm long and 30  $\mu\text{m}$  wide. Through the slit, light from the scanned line is dispersed by a prism-grating- prism device and projected onto the pixels of EMCCD detector. A two-dimensional image is generated with the spatial dimension along the horizontal axis and the spectral dimension along the vertical axis of the EMCCD. When the citrus sample is moved perpendicularly to the scanning direction by the motorized positioning table, one thousand seven hundred and forty line scans were performed for each leaf sample, and four hundred pixels covering the scene of the sample at each scan were saved. Therefore, a three-dimensional image cube (1740x400 for each band) was created.

The hyperspectral imaging software to transfer data and parameterization was developed using the Andor Software Development Kit (SDK, Luca, Andor Technology Inc., CT, USA) for the hyperspectral line scan imaging system. An Hg-Ne spectral calibration lamp (Oriel Instruments, Stratford, CT, USA) was used to investigate spectral

calibration of the system. Because of low light output in the visible region less than 450 nm, and low quantum efficiency of the EMCCD in the NIR region beyond 930 nm, the wavelength range between 451.67 nm and 927.71 nm was used (totaling 92 bands with a spectral resolution of 5.2 nm).

### **Hyperspectral Image Processing**

A detailed illustration of SAM and SID classification algorithms is given in Figure 8-2. After the relative hyperspectral reflectance images for all the wavelengths (483 to 959 nm) was acquired, the pixel value of the relative reflectance image was adjusted to the value in the range of 0 to 10,000 due to the range of original data from 0-16383 (the 14-bit EMCCD). The adjusted value could reduce rounding errors for further data analysis. Next, the mask template of the relative reflectance image was created by the threshold value. The threshold value was decided from the contrast image between the leaf surfaces with the background. Once the mask was covered with images, the image resolution was reduced to half of the image resolution. The reduction of the image resolution reduced the computational burden of the redundant features, and improved the performance of classification algorithms. The reduction brought comparable spatial resolutions for vertical and horizontal dimensions. Next, the spectral similarity for SAM and SID classification algorithms was determined between endmember and target spectrum by calculating the angle and divergence. The endmember spectra were extracted directly from sample images. In this study, the mean reflectance spectra from ROIs of 10 black spot samples were collected for describing endmember spectra. The mean and standard deviation from endmember spectra were illustrated in Figure 8-3. After applying SID and SAM mappings for the hyperspectral image of the test samples, rule images to separate black spot lesions from the fruit peel were obtained.

Representative rule images from SID and SAM mapping for the hyperspectral images are shown in Figure 8-4.

## **Results and Discussion**

### **Spectral Characteristics of Black Spot and Other Conditions**

The mean reflectance spectra and standard deviation of the samples with black spot, normal and different symptoms over the wavelength range between 483 and 959 nm are represented in Figures 8-5 and 8-6. This figure was developed using the mean of 10 spectra from the 10 different samples. Each plot in Figure 8-5 is generated by average region of interest (ROI) window covering 3 x 3 pixels extracted from each hyperspectral image. Figure 8-6 shows similar patterns which are regardless of the surface conditions.

All the spectra were presented as two local minima around 500 nm and 675 nm because of difference of light absorption of chlorophyll and carotenoid in the surface. As shown in Figure 8-6, the spectra from normal has the highest peak in wavelength from 500 nm to 600 nm and other diseases except black spot gradually are increased. Due to the low absorptions of chlorophyll and carotenoid, the spectra from black spot were not as outstanding as other peel conditions. Since any defect and contaminant was not appeared on the normal fruit surface, chlorophyll and carotenoid absorptions peak of the spectra region from 500 nm to 675 nm appear to be higher and more prominent. The spectra values of other disease generally are contained between the spectral of normal and black spot in the range of 12%-20% and 52%-75%. Reflectance spectra from hyperspectral images was extracted using ENVI 4.3 (ITT Visual Information Solutions, Boulder, CO, USA)

## **SID and SAM-based on Classifications**

Using the SID and SAM mapping, rule images of all the fruit samples were obtained. Since the SID and SAM mappings were based on the end spectra of black spots, the resulting rule images enhanced the black spot regions. Figure 8-7 shows a detailed comparison of the value distribution between black spot and non black spot regions. The SAM and SID pixel values of the black spot ranged from 0.01 to about 0.1. Based on these intensity values, a thresholding algorithm to separate the black spot was applied by setting the pixel as “1” for black spot region while “0” for the other classes. Figure 8-8 shows a sample binary image resulting from the thresholding operation where the black spots are represented by the white pixels.

### **The SID and SAM Threshold Values for Classification Accuracy**

The results shown in Table 8-2 is the classification summary using 6 threshold values for differentiating black spot from other conditions using SID mapping of hyperspectral images. The threshold value was increased from the default value of 0.01 to 0.06 with a step size of 0.01. As indicated in Figure 8-9(a), the value of 0.04 provided the best performance of 97.14%. When the threshold was changed from 0.01 to 0.04, the classification accuracy increased from 74.28% to 97.14%, respectively. However, the accuracy decreased when the threshold value was increased to 0.05. Based on the results shown in Table 8-4, the classification accuracy for 'Black Spot' class was 98%, and accuracy for 'No Black Spot' class was 96.92%. Only 3 black spot samples were misclassified. For the non black spot class, 12 samples were misclassified which included 8 greasy spot samples and 4 wind scar samples. The melanose samples and the market samples were correctly classified.

In the classification using SAM mapping, the threshold values were changed from 0.06 to 0.11 with a 0.01 increment. Figure 8-9(b) shows the effect of changing the threshold to classification accuracy, which increased from 81% (0.06) and peaked to 97.9% (0.09) and decreased to about 90%. It showed a similar characteristic as the SID mapping's threshold-classification accuracy, however SAM mapping had a higher accuracy in the threshold range that it was tested. As shown in Table 8-3 and Table 8-5 for the best classification result, the classification accuracies for 'Black Spot' class, and 'No Black Spot' class were 98% and 97.95%, respectively. There were three misclassified samples for 'Black spot' class, and eight misclassified samples for 'Non Black Spot' class which consisted of the greasy spot samples only. The other three non-black spot classes ('Normal', 'Wind Scar', and 'Melanose') have perfect classification results (100%). The plot of reflectance spectra in Figure 8-3 illustrated that greasy spot were very close to those of the black spot condition which contributed to the misclassification. Imaging parameters that will enhance the difference between the black spot samples and other confounding conditions such as greasy spot will be investigated in future studies. These will include looking into the effect of varying the illuminating source and changing the optical device (e.g. lens) to improve both reflectance and resolution of the images.

### **Summary**

Citrus black spot is the latest disease that has invaded the Florida citrus industry. Symptoms of black spot appear on the fruit surface as black lesions. The detection of fruits infected with black spot will greatly help in the controlling the spread of the disease. In this study, a hyperspectral imaging system was developed to recognize fruits exhibiting symptoms of citrus black spots along with fruits with other peel

conditions and market quality fruits. Five fruit classes were evaluated; 1) citrus black spot, 2) greasy spot, 3) melanose, 4) wind scar, and 5) market quality. The fruit samples were collected in the field and hyperspectral images of the leaf samples were collected in the laboratory with the spectral range of 400 nm to 900 nm. Reference spectrum of black spot was obtained from the ROIs that were selected from the reference hyperspectral images of the black spot samples.

In the image processing algorithm, two classification approaches were developed. First, Spectral Angle Mapper (SAM) was used to calculate the angle between the reflectance spectra of black spot and the reflectance spectra of an unknown fruit sample. The second method uses the Spectral Information Divergence (SID), which quantifies the similarities between the black spot's reference spectra and the spectra of the unknown fruit sample. The hyperspectral images of the tested fruit samples were mapped using both SAM and SID methods. A simple thresholding approach was applied to the rule images, which resulted from the mapping, to segment the black spot regions and eventually classify the fruits with black spots.

Based on the results, a black spot classification accuracy of 97.9% was obtained using the SAM approach with an optimal threshold value of 0.09. The SID mapping had a black spot classification accuracy of 97.14% with a 0.04 optimal threshold. All the melanose and market quality fruit samples were correctly classified using the two mapping approach, while the accuracy for greasy spot was about 91% and wind scar was over 96%. Overall, the performances of both classification approaches in detecting black spots along with other peel conditions were very good. However, it was found that SAM was superior to SID in threshold sensitivity; SID's performance deteriorated faster

as the threshold value moves away from the optimal threshold value while in SAM's classification performance was not greatly affected as the threshold was changed.

This research demonstrated that hyperspectral imaging combined with an appropriate image processing algorithm such as SAM and SID mapping could be used for detecting citrus black spots. Future studies will explore the identification of significant wavelengths from the reference spectrum to develop a multispectral imaging approach that could be applied in real time.

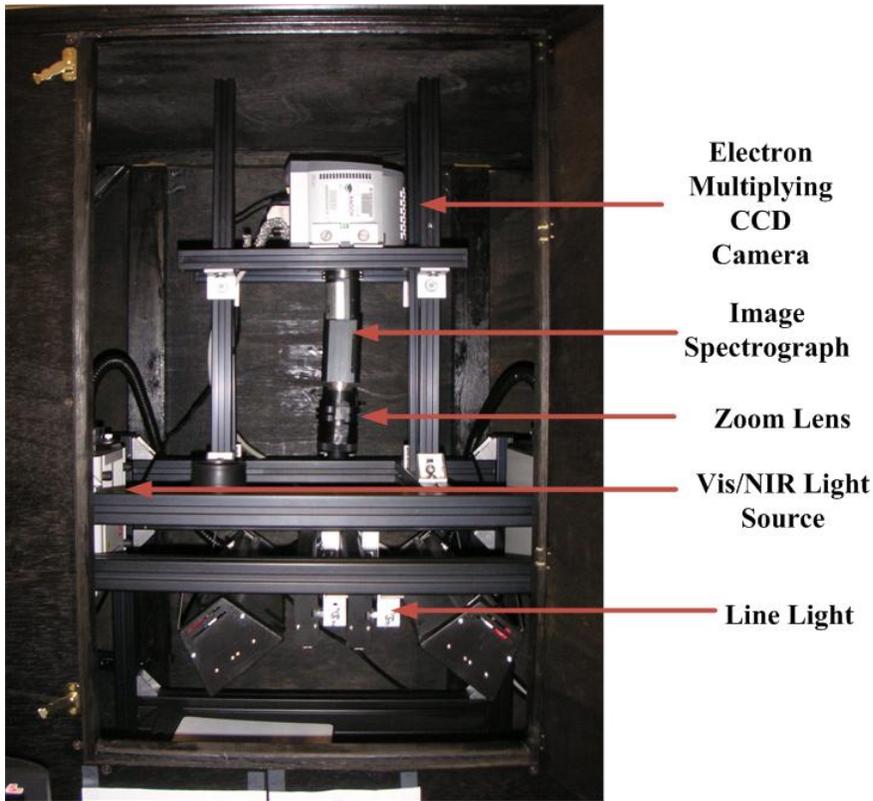


Figure 8-1. Hyperspectral line-scan imaging system

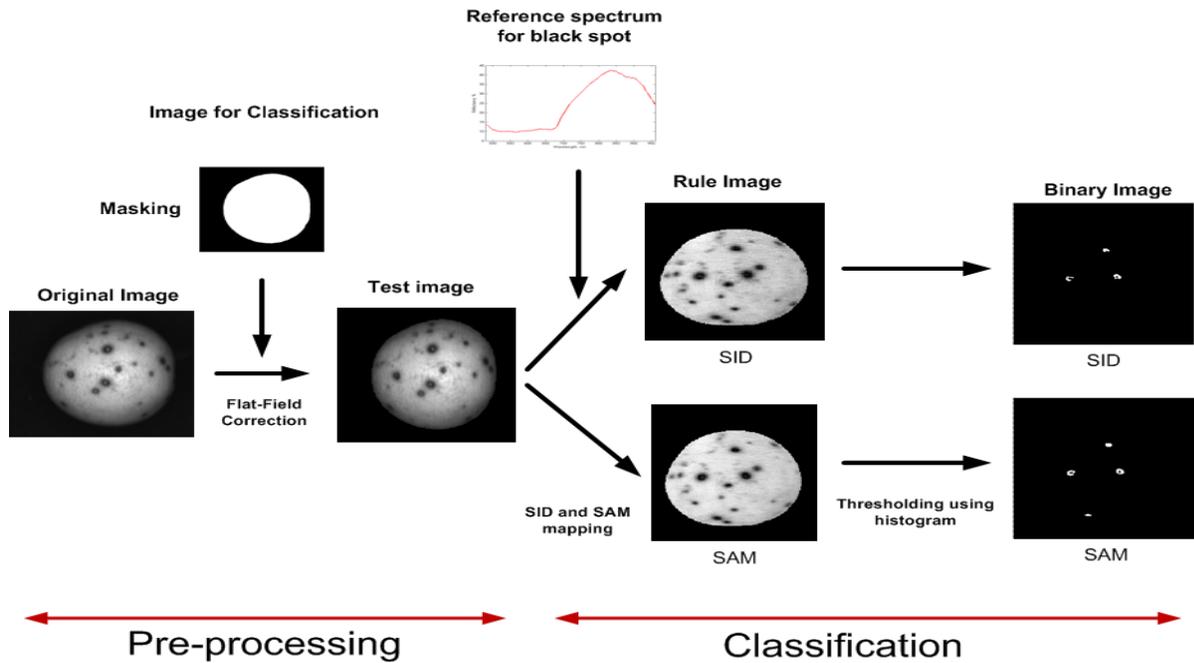


Figure 8-2. A detailed illustration of SAM and SID classification algorithms

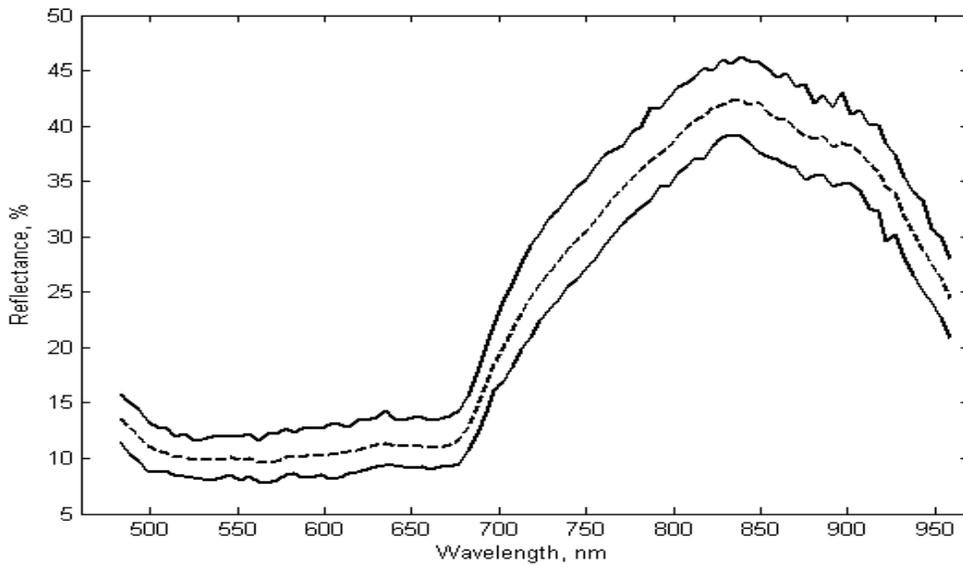


Figure 8-3. The mean and standard deviation from endmember spectra

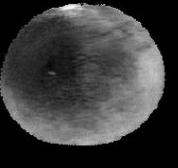
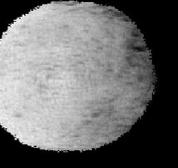
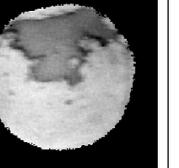
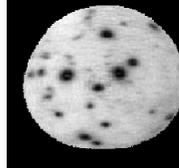
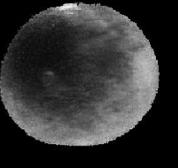
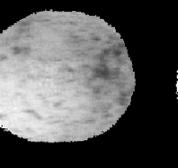
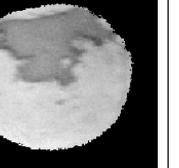
	<b>Black Spot</b>	<b>Greasy Spot</b>	<b>Melanose</b>	<b>Market</b>	<b>Wind Scar</b>
<b>S A M</b>					
	<b>Black Spot</b>	<b>Greasy Spot</b>	<b>Melanose</b>	<b>Market</b>	<b>Wind Scar</b>
<b>S I D</b>					

Figure 8-4. Representative rule images from SID and SAM mapping the hyperspectral images

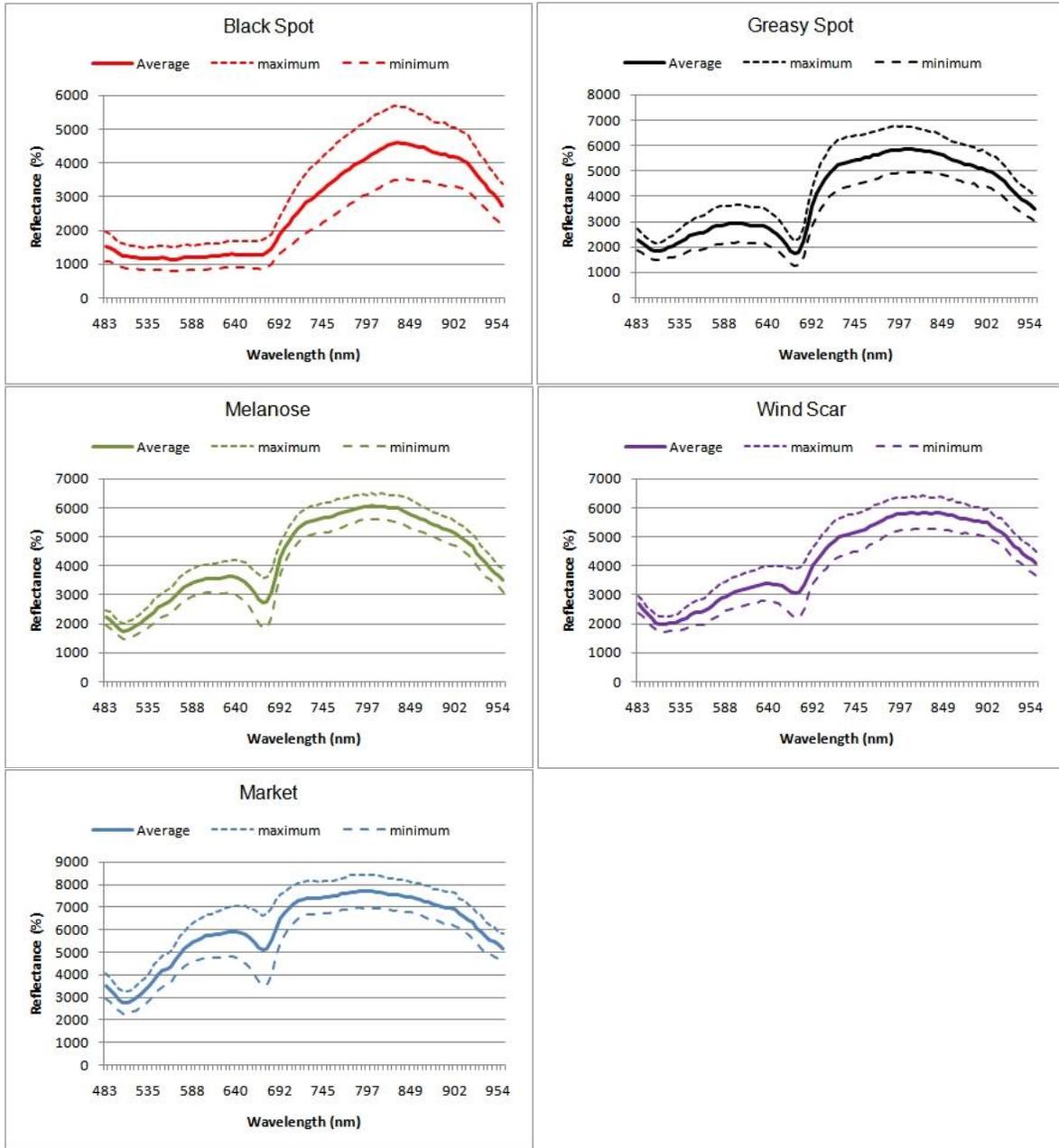


Figure 8-5. Reflectance mean spectra and corresponding standard deviation of each ROI from black spot, greasy spot, melanose, wind scar, and market

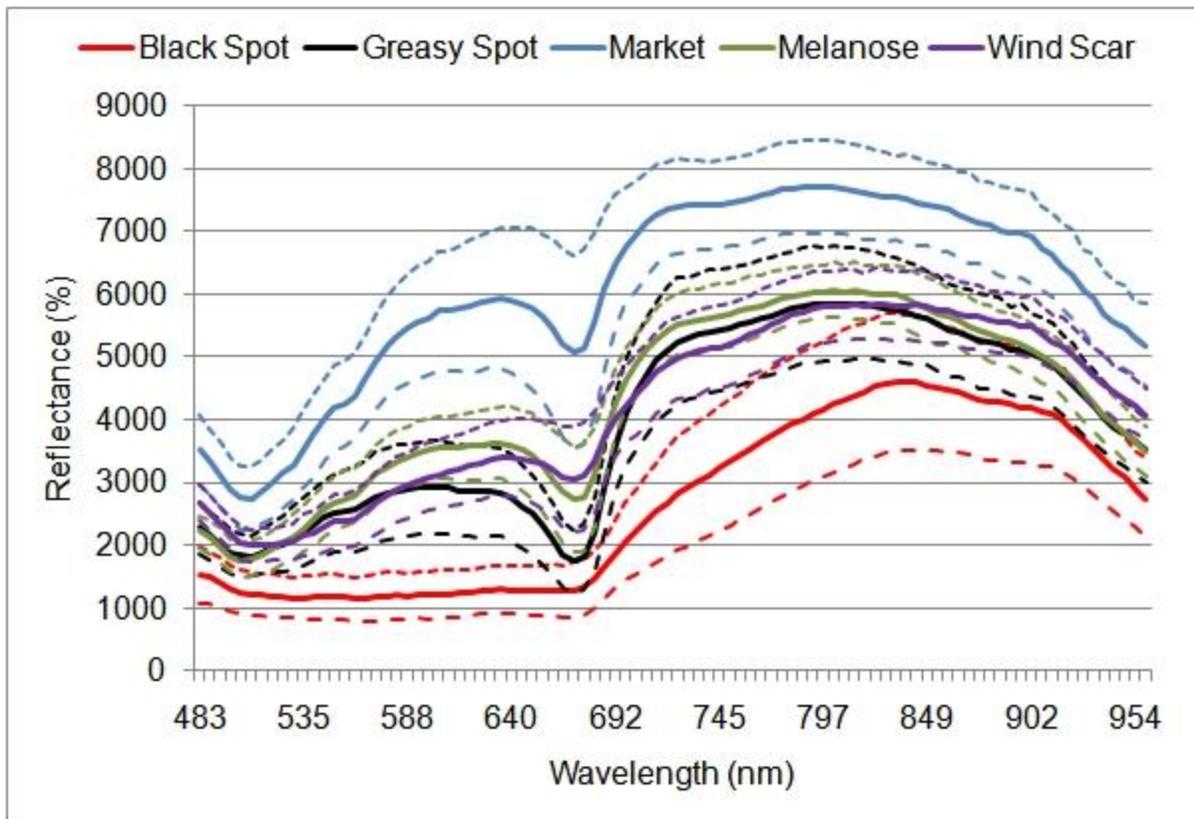


Figure 8-6. The reflectance spectra of the samples with black spot, normal and different symptoms over the wavelength range between 483 and 959 nm

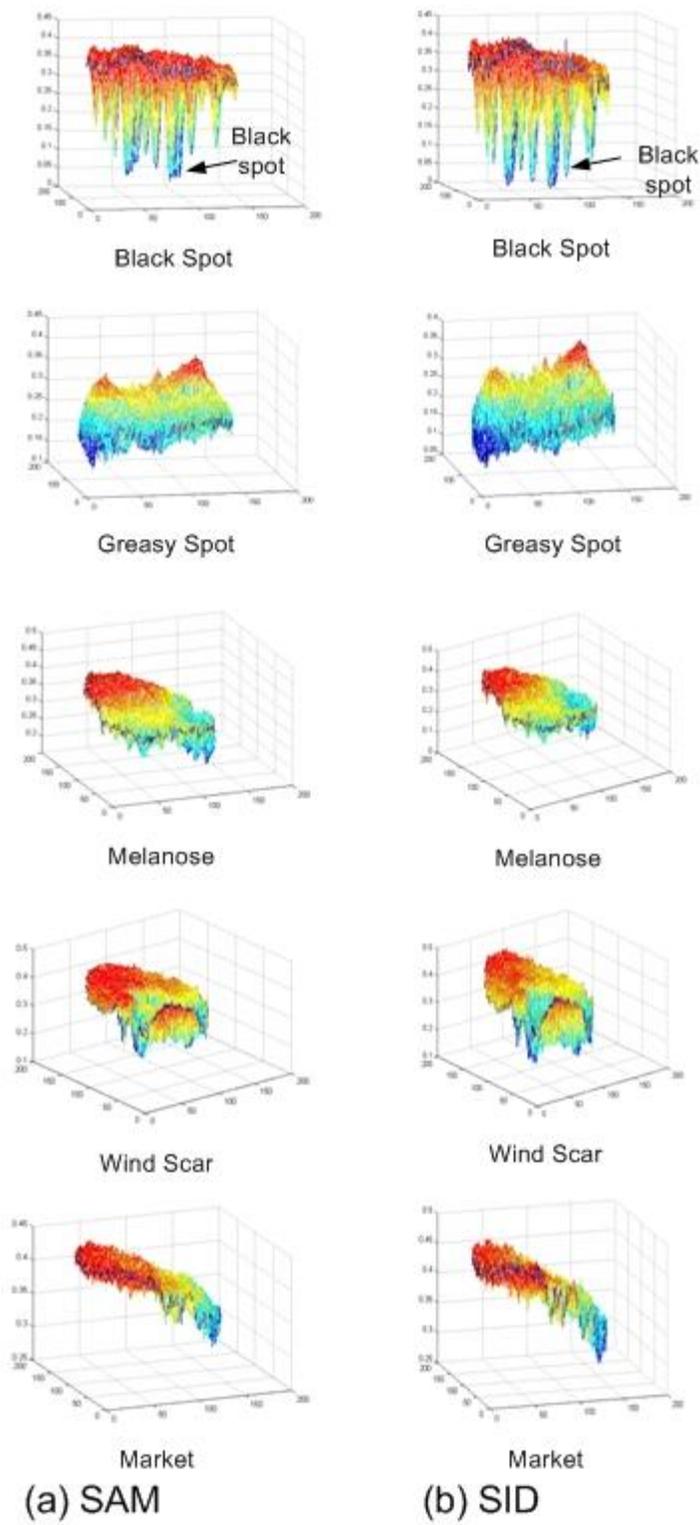


Figure 8-7. Detailed comparison of the value distribution between black spot and non black spot region

	Black Spot	Greasy Spot	Melanose	Market	Wind Scar
S A M					
S I D					
	Black Spot	Greasy Spot	Melanose	Market	Wind Scar
S I D					

Figure 8-8. A sample binary image resulting from the thresholding operation where the black spots are represented by the white pixels

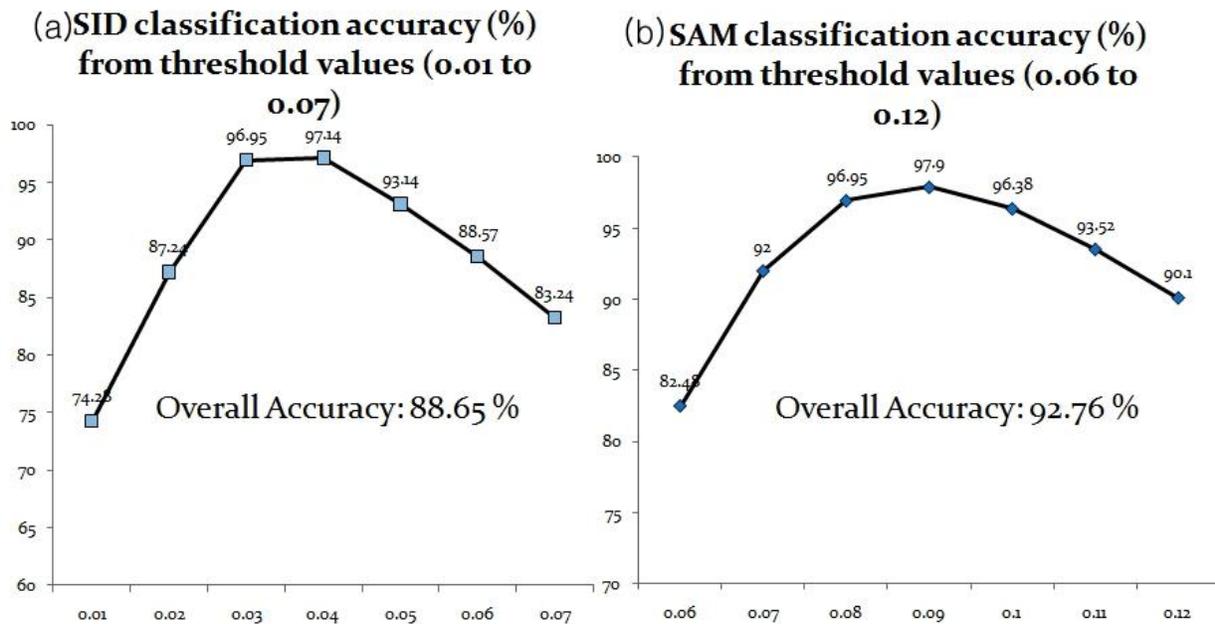


Figure 8-9. Comparison with SID and SAM classification accuracy

Table 8-1. The number of black spot and other conditions

<b>Class</b>	<b>Symptoms</b>	<b>Number</b>	<b>Total</b>
Greening Disease	Black spot	135	100
No Greening Disease	Greasy spot	90	390
	Market	90	
	Melanose	105	
	Wind scar	105	

Table 8-2. The classification summary using 6 threshold values for differentiating black spot from other conditions using SID mapping of hyperspectral images

<b>Spectral Information Divergence (SID)</b>							
<b>Thread Value</b>	<b>0.01</b>	<b>0.02</b>	<b>0.03</b>	<b>0.04</b>	<b>0.05</b>	<b>0.06</b>	<b>0.07</b>
<b>Black spot</b>	135	66	12	<b>3</b>	2	2	0
<b>Greasy spot</b>	0	1	3	<b>8</b>	22	38	53
<b>Market</b>	0	0	0	<b>0</b>	1	1	1
<b>Melanose</b>	0	0	0	<b>0</b>	0	0	1
<b>Wind scar</b>	0	0	1	<b>4</b>	11	19	33
<b>Accuracy (%)</b>	74.28	87.24	96.95	<b>97.14</b>	93.14	88.57	83.24
<b>Overall accuracy (%)</b>	<b>88.65</b>						

Table 8-3. The classification summary using 6 threshold values for differentiating black spot from other conditions using SAM mapping of hyperspectral images

<b>Spectral Angle Mapper (SAM)</b>							
<b>Thread Value</b>	<b>0.06</b>	<b>0.07</b>	<b>0.08</b>	<b>0.09</b>	<b>0.10</b>	<b>0.11</b>	<b>0.12</b>
<b>Black spot</b>	92	42	14	<b>3</b>	1	1	1
<b>Greasy spot</b>	0	0	2	<b>8</b>	16	29	44
<b>Market</b>	0	0	0	<b>0</b>	0	0	1
<b>Melanose</b>	0	0	0	<b>0</b>	0	0	0
<b>Wind scar</b>	0	0	0	<b>0</b>	2	4	6
<b>Accuracy (%)</b>	82.48	92.00	96.95	<b>97.90</b>	96.38	93.52	90.10
<b>Overall accuracy (%)</b>	<b>92.76</b>						

Table 8-4. The best overall classification accuracy using the SID threshold value of 0.04

<b>Class</b>	<b>Symptoms</b>	<b>Misclassified</b>	<b>Accuracy (%)</b>	
Black spot Disease	Black spot	3	98.00	98.00
No Black spot Disease	Greasy Spot	8	91.11	96.92
	Melanose	0	100.0	
	Wind scar	4	96.19	
	Market	0	100.0	
Total		15	97.14	

Table 8-5. The best overall classification accuracy using the SAM threshold value of 0.09

<b>Class</b>	<b>Symptoms</b>	<b>Misclassified</b>	<b>Accuracy (%)</b>	
Black spot Disease	Black spot	3	98.00	98.00
No Black spot Disease	Greasy Spot	8	91.11	97.95
	Melanose	0	100.0	
	Wind scar	0	100.0	
	Market	0	100.0	
Total		11	97.90	

## CHAPTER 9 HYPERSPSPECTRAL BAND SELECTION FOR DETECTION OF BLACK SPOT USING MULTISPECTRAL ALGORLTHMS

Citrus black spot is a fungal disease that threatens marketability of fresh market citrus production in Florida. Caused by *Guignardia citricarpa* (sexual stage) and *Phyllosticta citrucarpa* (asexual stage), this fungal disease is a significant threat to Florida citrus groves. The symptoms of black spots include hard spot lesions, virulent spot, cracked spot, and false melanose, and all appear on the fruit surface (The Institute of Food and Agricultural Sciences Extension at University of Florida, 2010). After citrus trees are infected with citrus black spot, the fruit yield, and quality are greatly reduced. Moreover, fruits infected with citrus black spots are not acceptable for fresh market export and domestic trade (Chung, 2009). Therefore, it is important to establish early control system to achieve profitable production in Florida. Detecting fruits infected with citrus black spot can help in controlling the spread of this disease specifically in areas that are black spot free. The design and implementation of computer vision technologies that efficiently detect citrus black spot would be an important aid in the control effort.

The identification of various crops and plant parts using computer vision and image processing techniques has been studied by several researchers. Jimenez et al. (2000) surveyed several computer vision approaches for locating fruit in trees. Regunathan and Lee (2005) indentified fruit count and size using machine vision and an ultrasonic sensor. Burks et al. (2000) developed a method for classification of weed species using color texture features and discriminate analysis. Tang et al. (1999) developed a texture-based weed classification method using Gabor wavelets and neural networks for real-time selective herbicide application. Pydipati et al. (2006) identified citrus disease using the co-occurrence matrix method, (CCM) texture feature method,

and a discriminant analysis. Du et al [9] described five different texture feature methods, including the common first-order gray-level statistics (FGLS), run length matrix (RLM), gray-level co-occurrence matrix (GLCM), fractal dimension (FD), and wavelet transform (WT) based method. After that, both simple correlation analysis and partial least squares regression (PLSR), to investigate tenderness of cooled pork ham.

Optical techniques have been used widely in the food processing and inspection application in recent years. In particular, hyperspectral imaging technologies have a growing interest in quality evolution and inspection of food and agricultural product (Sun, 2008). Previous research has demonstrated hyperspectral imaging technologies and applications for agricultural product. Jiang et al. (2007) used hyper-spectral fluorescence imaging to analyze the differences between walnut shells and meat. Hyper-spectral fluorescence imaging system scanned samples at 79 different wavelengths ranging from 425nm to 775nm with 4.5 nm increments, and data redundancy of data was reduced by principal component analysis (PCA). Zhang et al. (2005) suggested a creative classification approach for distinguishing healthy and fungal infected wheat kernels during storage. The research showed the potential use of NIR hyper-spectral imaging in grain quality assessment. The research used NIR hyper-spectral imaging and support vector machine (SVM) for identifying the fungi that caused the infection. Kim et al. (2002) researched a method for using hyper-spectral data to identify wavebands to be used in multispectral detection systems, and evaluated spatial and spectral responses of hyper-spectral reflectance images of fecal-contaminated apples. Lee et al (2005) used the hyper-spectral imaging technique to detect defects on apple peel after harvest. They developed a proper wavelength selection method for

detecting the defects. In hyperspectral image classification methods, a spectral angle mapper (SAM) and spectral information divergence (SID) classification methods were applied to identify defected agricultural productions. Park et al. (2007) used SAM algorithms to detect fecal and ingesta contaminants on the surface of poultry carcasses. Qin et al. (2009) introduced the detection of citrus canker using SID classification methods.

Hyperspectral imaging algorithms described above are very useful to inspect citrus diseases, but they would limit their applications such as rapid on-line disease detection because to large amount of 3-D hyperspectral imaging data. Therefore, algorithms for optimal band selection and multispectral image classification are needed to fulfill the goal of real-time black spot inspection. The overall objective of this study was to develop a multispectral algorithm based method to detect citrus black spot on citrus peel. Specific objectives implemented to accomplish the overall objective were to:

- Use a hyperspectral imaging system to collect 3-D hyperspectral images from fruits with black spot, greasy spot, melanose, wind scar, market;
- Determine optimal wavelength bands based on principal component analysis (PCA), singular vector decomposition (SVD), and correlation analysis (CA); and
- Develop algorithms for multispectral imaging analysis and classifying the citrus peel conditions based selected wavelengths.

## **Materials and Methods**

### **Hyperspectral Image Acquisition**

Figure 9-1 shows a hyperspectral line-scan imaging system for acquiring hyperspectral images of the fruit samples. The imaging system consisted of an electron-multiplying charge-coupled-device (EMCCD) camera (Luca, Andor Technology Inc., CT, USA) with imaging spectrograph (ImSpector V10E, Spectral Imaging Ltd., Oulu,

Finland) and a C-mount lens (Rainbow CCTV S6X11, International Space Optics, S.A., Irvine, CA, USA), a pair of halogen line lamp (21V, 150W) powered with a DC voltage regulated power supply (Dolan-Jenner Industries, Inc., Lawrence, MA, USA), and a programmable motorized positioning table (XN10-Xslide, Velmex Inc., Bloomfield, NY, USA). This equipment was placed inside in a dark box to eliminate stray external light.

The hyperspectral imaging software to transfer data and parameterization was developed using the Andor Software Development Kit (SDK, Luca, Andor Technology Inc., CT, USA) for the hyperspectral line scan imaging system. An Hg-Ne spectral calibration lamp (Oriel Instruments, Stratford, CT, USA) was used to investigate spectral calibration of the system. Because of low light output in the visible region less than 450 nm, and low quantum efficiency of the EMCCD in the NIR region beyond 930 nm, the wavelength range between 451.67 nm and 927.71 nm was used (totaling 92 bands with a spectral resolution of 5.2 nm).

### **Correlation Analysis Algorithms for Band Selection**

Correlation analysis algorithms using two wavelength ratios was used to find the best pair of wavelengths for differentiating between "black spot" and "no black spot" conditions. Referring to Figure 9-2, 35 and 80 representative spectra of 'Black Spot' classes and 'No Black Spot' classes were extracted from the hyperspectral images, respectively. All the spectral data were extracted from region of interest (ROI) hyperspectral image data, and at least one region of interest (ROI) was selected from the representative spot for each sample. Average spectrum calculated from the merged ROIs within each sample was used as individual reflectance spectrum for the wavelength selections.

For each of the "black spot" and "no black spot" spectra, two-band ratio values were calculated for all possible two-wavelength combinations. In the algorithms, the quality index for "black spot" spectral was labeled to be "1", and the quality index for "no black spot" was labeled to be "0". Correlation analyses were performed to determine correlation coefficient between two band ratios and fruit peel quality index values of the samples. The correlation analysis procedures described above were executed using program developed in MATLAB R2007B (MathWorks, Natick, MA, USA) and extraction of reflectance spectra from the hyperspectral images was finished using ENVI 4.3 (ITT Visual Information Solutions, Boulder, CO, USA).

### **Principal Component Analysis Algorithms for Band Selection**

Each score images from principal component analysis represent a linear summation of the original hyperspectral single-band images at different wavelengths weighted by the corresponding spectral loadings. The weighting coefficients from spectral loading values are useful for identifying important wavelengths that are responsible for the unique features appeared in the score images. In Figure 9-3, the PC-3 score images showed great potential for indentifying black spot from fruit peel conditions. Figure 9-4 shows the detailed procedures for band selection using PCA algorithms. First, with the random number generator, the random sample of 20 was selected, and average weight coefficients for the optimal principal component from 20 grapefruit samples with black spot lesions were selected. Next, optimal wavelengths were determined by large absolute weight coefficients. After important wavelengths were identified, image classification was performed.

## **Singular Vector Decomposition Algorithms for Band Selection**

A detailed illustration of singular vector decomposition classification algorithms is given in Figure 9-5. The hyperspectral reflectance images for singular vector decomposition (SVD) analysis were extracted from ROI hyperspectral image data using MATLAB R2007b (Mathworks, Natick, MA, USA). The ROI selection was started manually by determining a point on the hyperspectral reflectance image, and then was finished for extracting a square portion with the dimension of 10×10 pixels centered on the determined point. Representative ROI images for each black spot peel conditions used in this study are shown in Figure 9-6.

After the ROI hyperspectral reflectance images were extracted, the ROIs were analyzed with SVD. All analysis was done using MATLAB R2007b (The Mathworks, Inc., Natick, Mass.) and the Hyperspectral Image Analysis Toolbox (HIAT) (Research Thrust R3, Northeastern University, USA). HIAT is a toolkit for MATLAB R2007b. This package is released for the developed and evaluation of hyperspectral image analysis algorithms.

## **Result and Discussion**

### **Mean Relative Reflectance Spectra of Black Spot**

The reflectance mean spectra and corresponding standard deviation of the samples with 5 different conditions over the wavelength range between 451.67 nm and 927.71 nm are represented in Figure 9-7. The reflectance spectra were calculated using the mean of 10 spectra from 10 randomly selected samples. Each plot in Figure 9-7 is generated by the average ROI window covering 3 x 3 pixels extracted from each hyperspectral image.

In mean spectra obtained from the hyperspectral images of citrus greening with other conditions, the chlorophyll absorption peaks of wavelength from 400 nm (blue) to 600 nm (green) appear to be higher and more prominent. The relative reflectance increase steadily over the wavelengths 720 - 900 nm, but the reflectance rate is low (10-30%). Reflectance spectra from hyperspectral images was extracted using ENVI 4.3 (ITT Visual Information Solutions, Boulder, CO, USA)

### **Hyperspectral Band Selection using CA**

Figure 9-8 shows the contour plot the correlation coefficients ( $r$ ) between two band ratio and fruit peel conditions. The highest correlation coefficient ( $r$ ) of 0.92 occurred at the visible regions (724 nm and 698 nm, specifically). The band ratio with the second highest correlation value of 0.91 was in near-infrared (NIR) and visible regions (854 nm and 598 nm, specifically). Referring Figure 9-6, the reflectance values of black spot at 724 nm and 698 nm were similar to other disease conditions. However, the wavelength of 598 nm has the lowest reflectance values of back spot.

### **Hyperspectral Band Selection using PCA**

Figure 9-9 shows mean weighting coefficients for the third principal component from 20 black spot ROIs. The general pattern of the PC-3 loading revealed that the wavelengths in blue, green, red, and near-infrared regions were significant factors for generating the PC-3 score images. Local peaks appeared on the loading curve across the entire spectral region from 400 to 900 nm, and they had larger absolute values than those at other wavelengths.

As shown in Figure 9-9, four optimal wavelengths were selected, and they centered at 870 nm, 802 nm, 724 nm, and 671 nm. The wavelength of 671 nm likely represents the low chlorophyll content in the black spot lesions, and the wavelength in

the NIR region (870 nm, 802 nm) would provide useful information that is insensitive to colors. These four wavelengths can be adopted by a multispectral imaging solution for real-time black spot detection on commercial sorting line.

### **Hyperspectral Band Selection using SVD**

In singular vector decomposition algorithms, the number of features is not always easy to determine. The energy fraction could be used to argue for the usage of a given number of features. The number of features could also be determined from the characteristics of the singular values. When the singular values stabilize, the remaining features are usually contaminated with much noise and therefore not useful. Singular values of black spot are shown in Figure 9-10. From the number of feature are four and up, the singular values are almost constant.

Based on the Figure 9-10, four optimal wavelengths (e.g., 907 nm, 724 nm, 677 nm, and 572 nm) were determined by SVD. The optimal wavelengths selected by SVD showed that the wavelengths in red, green, and short-wavelength NIR regions could be important for black spot identification. Specifically, the reflectance values of black spot at wavelength of 677 nm have the lower chlorophyll content than other disease conditions. Their effectiveness for black spot detection also needs to be evaluated by image classification results.

### **The Hyperspectral Images Selected by Band Selection Methods**

Figure 9-11 (a) and (b) show the representative single band reflectance images of oranges with five different conditions at the wavelengths with highest correlation coefficients and second highest correlation coefficients. As observed in Figure 9-11 (a) and (b), the images at wavelengths of visible regions (e.g., 598 nm, 698 nm, and 724 nm) had more distinctive spots and higher contrast on the sample surface. Fruit

samples with normal conditions did not have the contrasting regions because of the absence of disease. Moreover, the other images at wavelengths of NIR regions (e.g., 854 nm) had low contrast between the diseased and the non-diseased surface.

Representative spectral reflectance images of diseased peel and normal conditions at four wavelength selected by PCA and SVD are shown in Figures 9-12 and 9-13. They also had more distinctive and higher contrast on the sample surface in the visible spectral region below 730 nm. Due to relatively high reflectance characteristics of diseased surface at long wavelengths, the contrast between diseased and non-disease spot surface was reduced in the NIR region.

### **Classification of Black Spot using Band Ratio Images**

The ratio images of black spot using wavelengths selected by CA, PCA and SVD algorithms were illustrated in Figures 9-14, 9-15, and 9-16. As shown in Figures 9-14 and 9-15, each PCA and SVD band selection has six bands ratio images calculated using all six possible combinations of four bands.

For the ratio images from CA band selections in Figure 9-16, the black spot revealed distinctive bright spots on the fruit surface due to their relatively high ratio values. The normal area of the fruit peel appeared darker than the disease owing to their small ratio values using selected wavelengths. However, two band ratio image from SVD and PCA band selections exhibited inconsistent patterns black spot identification. The ratio images (917 nm/729 nm, 917 nm/572 nm, and 729 nm/572 nm) from PCA and the (870 nm/802 nm, 870 nm/724 nm, and 802 nm/724 nm) ratio images from SVD showed similar features with those obtained using the bands from CA. In PCA, the stem-end and some edge area appeared bright in the images (802 nm/671 nm, 870 nm/671 nm, and 724 nm/671 nm) as well as the images in Figure 9-14. Moreover,

they showed inconsistent brightness and darkness patterns of the black spot lesions. Referring to Figure 9-15, inconsistent brightness patterns were also observed in ratio image images (907 nm/677 nm, and 724nm /677 nm) from SVD. The image (677 nm/572 nm) from SVD is not as important as other images because the bright area, other than black spot, could be the sources of false positive errors.

To separate the citrus black spots from the peel surfaces, a thresholding-based segmentation was applied to ratio images. In bi-level thresholding, the pixel scale less than threshold value are set as "1" (black spot condition) while others are set as "0"(non-black spot condition and back ground). To determine threshold values for automated classification of black spot, Figures 9-17, 9-18 and 9-19 illustrates histograms for ratio images of 802 nm/724 nm from PCA, 919 nm/572 nm from SVD, and 854 nm/598 nm from CA. The histograms shows that ratio values of black spot is larger than other peel diseases because black spot along with other four disease showed bright in the image and non-black spot lesions on fruit surfaces appeared darker than the black spot areas. Histograms for the other ratio images, as suggested in Figures 9-17, 9-18, and 9-19, failed to show high ratio value of black spot on fruit surface and are omitted for brevity. After image thresholding, some defects might be present in the images. As shown Figure 9-20, image-opening morphological filtering can remove noise a false black spot on the segmented images. This post-processing might reduce the misclassification of black spot diseases.

The classification summary for differentiating Black Spot from other diseases and normal conditions using PCA is shown in Table 9-1. As previously indicated, the test data consisted of 135 images from citrus greening and 390 images from others. The

best performance was 96.19% of the (802 nm/724 nm) ratio image (Table 9-1). The classification accuracy for the 'Black Spot' class was 94.81%, and for the 'No Black Spot' class was 96.67%, respectively. There were 7 misclassified samples for 'Black Spot' class, and 13 misclassified samples for 'Non Black Spot' class which consisted of greasy spot samples only. The other three non-black spot classes ('Normal', 'Wind Scar', and 'Melanose') have perfect classification results (100%). Based on Table 9-2, (724 nm/671 nm) ratio image had the lowest classification (58.1%). Figure 9-21 illustrates the plot of classification summary using 10 threshold values for differentiating black spot from other conditions using the (802 nm/724 nm) ratio image from PCA. The threshold value was increased from the value of 1.2 (Figure 9-21) to 1.6 with a step size of 0.05. As indicated in Figure 9-21, the value of 1.45 provided the best performance of 96.19%. When the threshold was changed from 1.2 to 1.45, the classification accuracy increased from 44.19% to 96.19%, respectively. However, the accuracy decreased when the threshold value was increased to 1.5.

In the classification using the SVD band selection as shown in Table 9-2, the ratio images were consist of the three pairs of two visible ratio bands, and three pairs of NIR and visible ratio bands. As shown in Table 9-2 for the best classification result, the classification accuracies for 'Black Spot' class, and 'No Black Spot' class were 97.04% and 93.59%, respectively. Four samples were misclassified in the 'Black Spot' classes, and there were 24 samples that were misclassified in the 'No Black Spot' class which included 23 'Greasy Spot' diseases, 2 'Wind Scar' diseases. This misclassification demonstrated the confusion in classifying black spot between greasy spot and wind scar. On the other hand, all 'Market' and 'Melanose' conditions were correctly classified.

As shown in Table 9-2, three pairs of NIR and visible ratio images generated slightly worst result than that of two visible ratio images. Two pair of NIR and visible (907 nm/724 nm, and 917 nm/572 nm) ratio images achieved the high classification accuracies (above 90%) being slightly higher than those of two visible ratio images. The above SVD results suggested that NIR and visible ratio bands necessarily improve the classification performance. Figure 9-22 shows the effect of changing the threshold to classification accuracy, which increased from 82.48% (2.70) and peaked to 94.48% (3.20) and decreased to about 91.43%.

The overall classification accuracy peaked at the threshold of 3.40, illustrated in Figure 9-23, by which the accuracy achieved 94.86%. Table 9-3 summarizes detailed classification results for identifying black spot from normal and other diseased peel conditions using all ratio images from CA wavelength selections. As shown in Table 9-3, 98.51% of the black spot were correctly classified as 'Black Spot' class, and accuracy for the 365 samples in 'No Black Spot' class was computed as 93.59%. A total of 27 samples were misclassified among all 525. 'Greasy Spot' samples were classified with the low accuracy of 72.22% due to their relatively high ratio values using the ratio bands selected by CA, and all other conditions were correctly classified.

### **Summary**

Our previous studies showed that disease detection based on hyperspectral reflectance imaging is useful for detection of black spot disease. Although good results for classification of citrus diseases were obtained with hyperspectral image techniques, disease detection using multispectral algorithms should be evaluated to further improve accuracies and reduce computational burden for real-time detection system. Therefore, band selection of hyperspectral imaging is important to fulfill the goal of rapid on-line

disease inspection using multispectral techniques. In this study, Principal component analysis, singular value decomposition, and correlation analysis algorithms were developed to find a few optimal wavelengths for multispectral detection of black spot.

Based on the results, a black spot classification accuracy of 96.19% was obtained using the PCA approach with the NIR and visible ratio image (802 nm/724 nm) and an optimal threshold value of 1.5. The SVD approach had a black spot classification accuracy of 94.48% with the NIR and visible ratio image (917 nm/572 nm) and an optimal threshold value (3.20). In CA algorithms, the (854 nm/598 nm) ratio images that had second highest correlation coefficient (9.1) provided the best classification accuracy (94.86%). As shown in all results, the classification accuracies for 'Greasy Spot' were relatively lower than other disease conditions because, reflectance spectra of greasy spot illustrated in Figure 9-7 were very close to those of the black spot condition which contributed to the misclassification. Thus, the hyperspectral imaging parameters that will enhance the difference between the black spot samples and other confounding conditions such as greasy spot will be considered in next studies.

Referring to Tables 9-1, 9-2, and 9-3, the short-wavelength NIR region and visible region (NIR/visible) ratio bands gave the good performances of three classification approaches among all combinations of the ratio images. Thus, this band pair (NIR/visible) has great potential to be adopted by a two band multispectral imaging system for real-time black spot detection. In the control of threshold values, the best overall classification accuracies were determined based on the assumption that the incorrect classification of a fruit without black spot and with black spot (i.e., false negative and false positive) for was equally weighted. In practice, the control of the two

types errors will be needed for adjusting the threshold values to meet the requirements of real-time system.

This research demonstrated that multispectral algorithms using hyperspectral imaging could be used for differentiating black spot from other disease conditions. Future studies will explore the utility of these algorithms in a real-time black spot detection system.

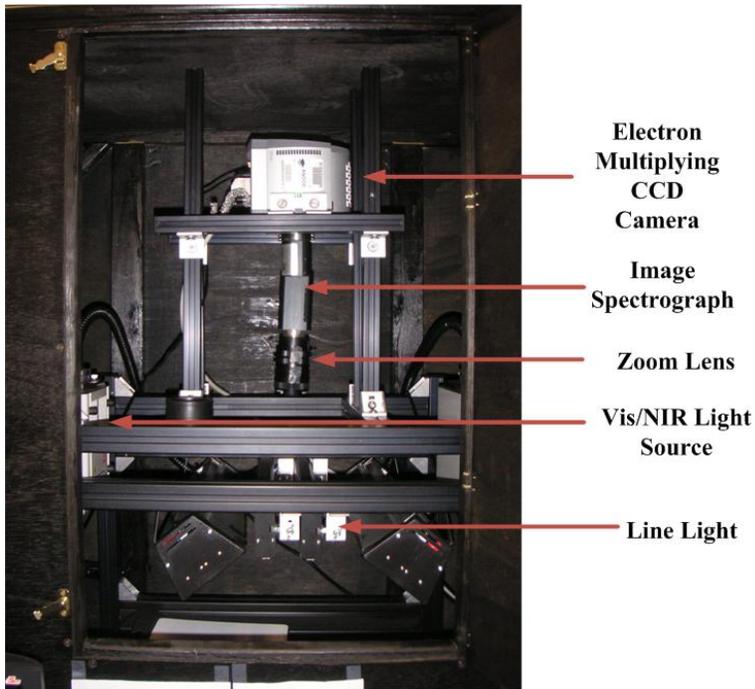


Figure 9-1. Hyperspectral line-scan imaging system

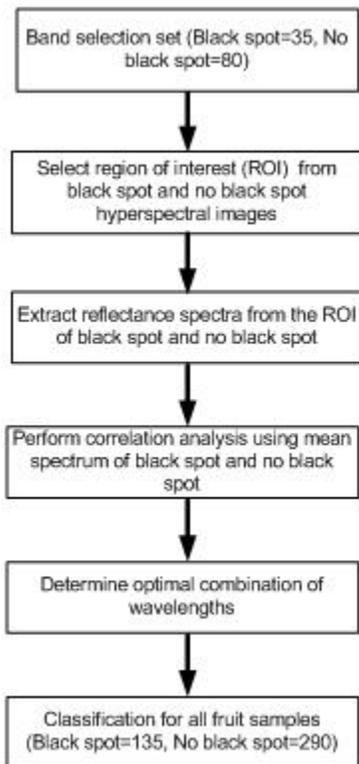


Figure 9-2. Data analysis and classification algorithms for CA

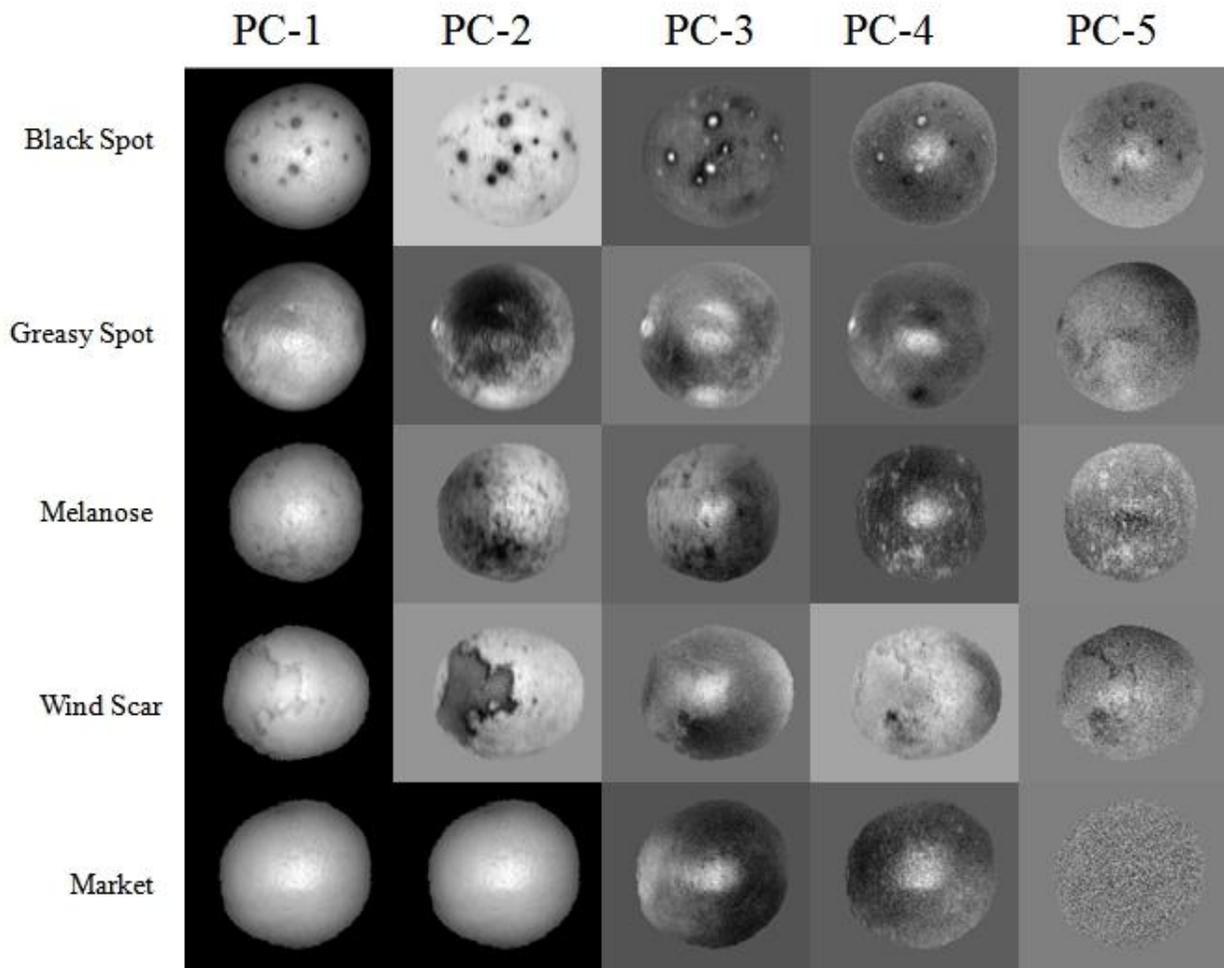


Figure 9-3. First five score images obtained from principal component analysis for hyperspectral images of grapefruit samples

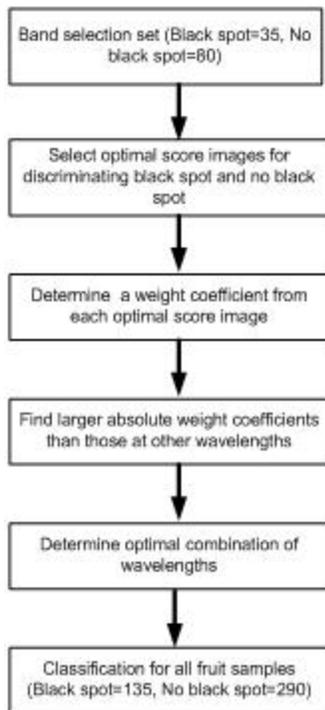


Figure 9-4. Data analysis and classification algorithms for PCA

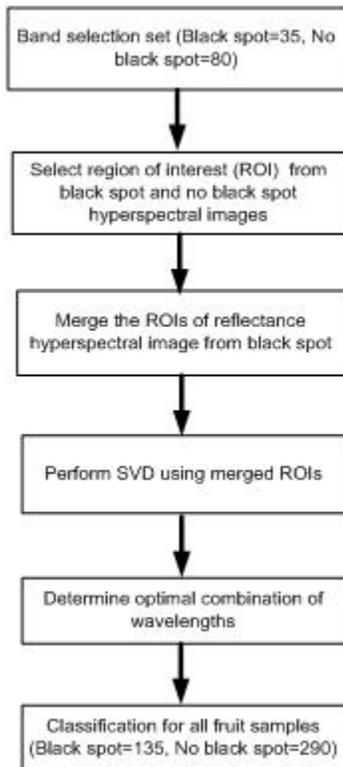


Figure 9-5. Data analysis and classification algorithms for SVD

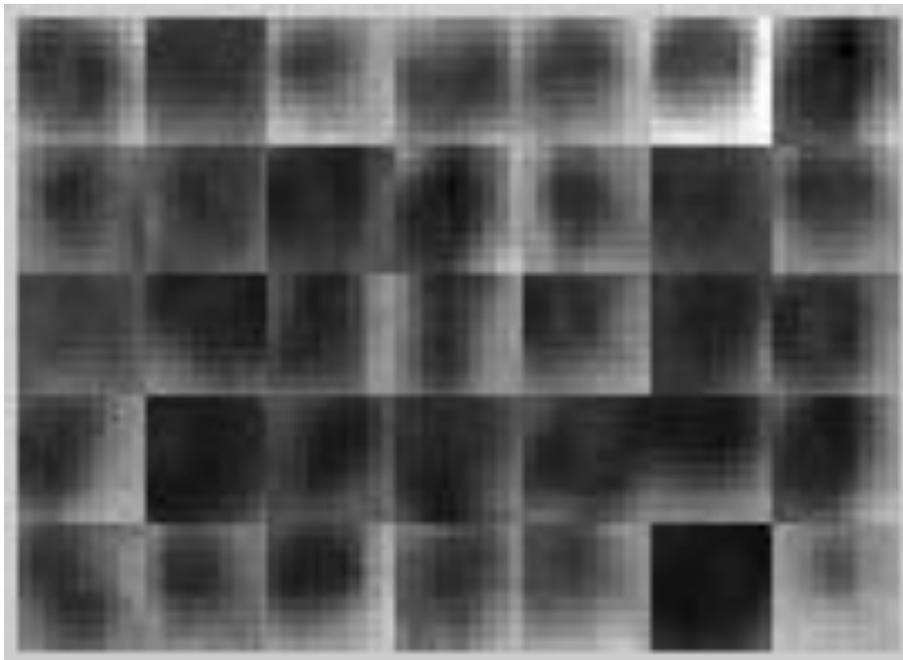


Figure 9-6. Representative ROI images for each black spot peel conditions

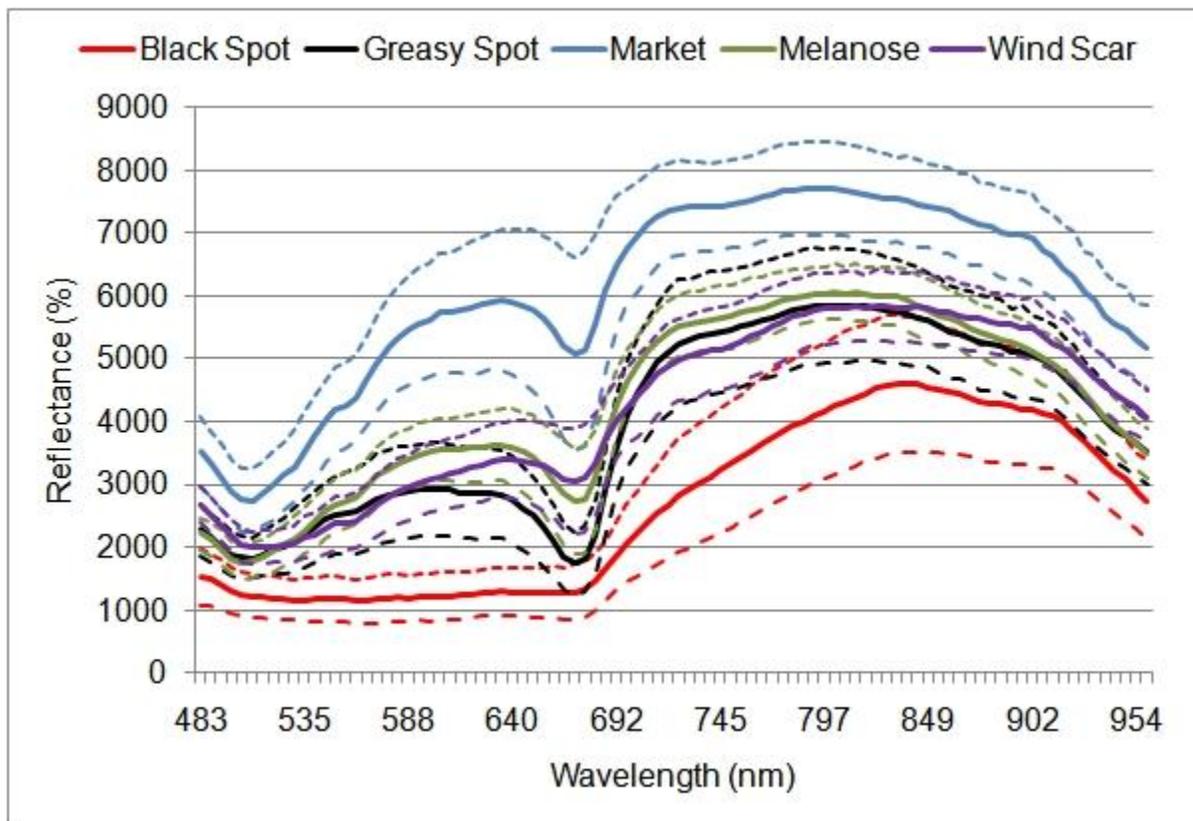


Figure 9-7. The reflectance spectra of the samples with black spot, normal and different symptoms over the wavelength range between 483 and 959 nm

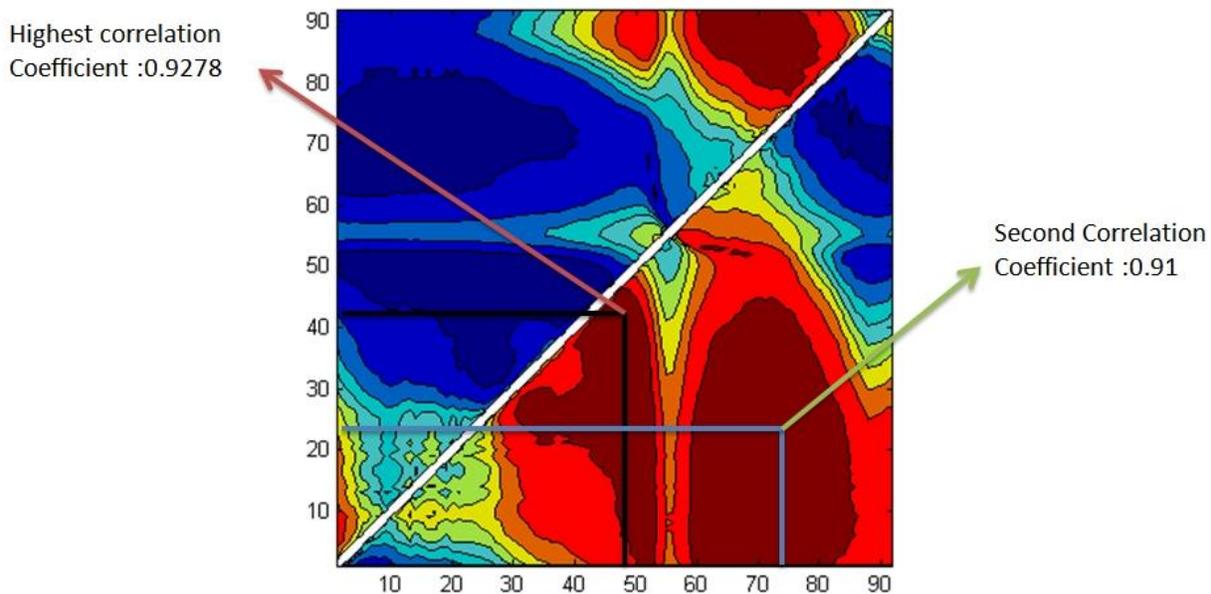


Figure 9-8. The contour plot the correlation coefficients ( $r$ ) between two band ratio and fruit peel conditions

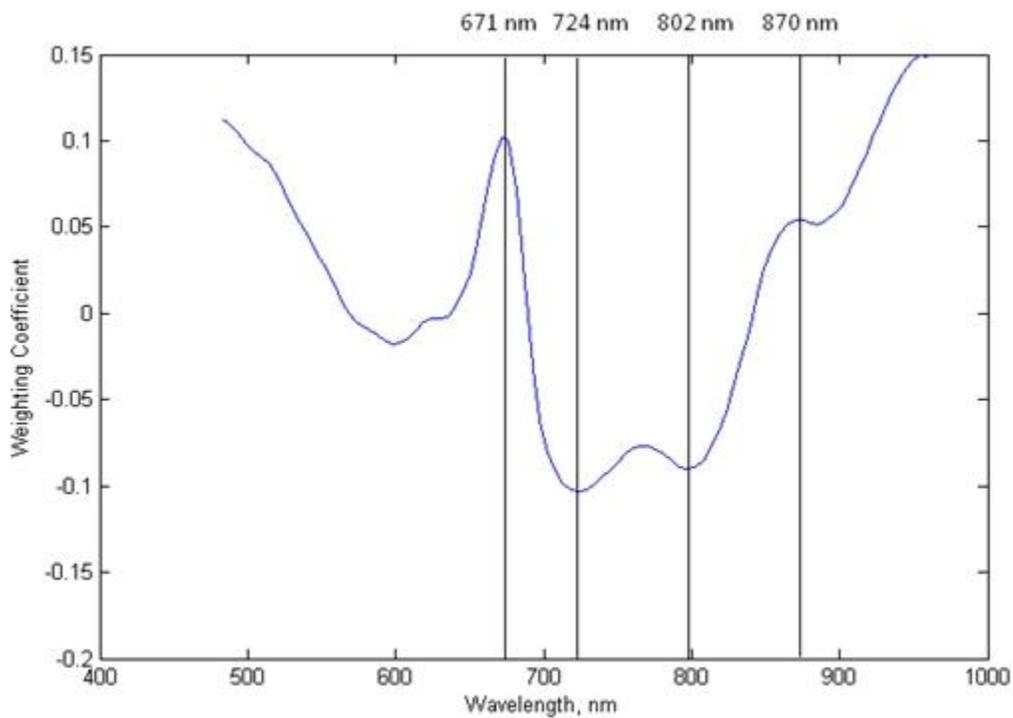


Figure 9-9. Mean weighting coefficients for the third principal component from 20 black spot ROIs

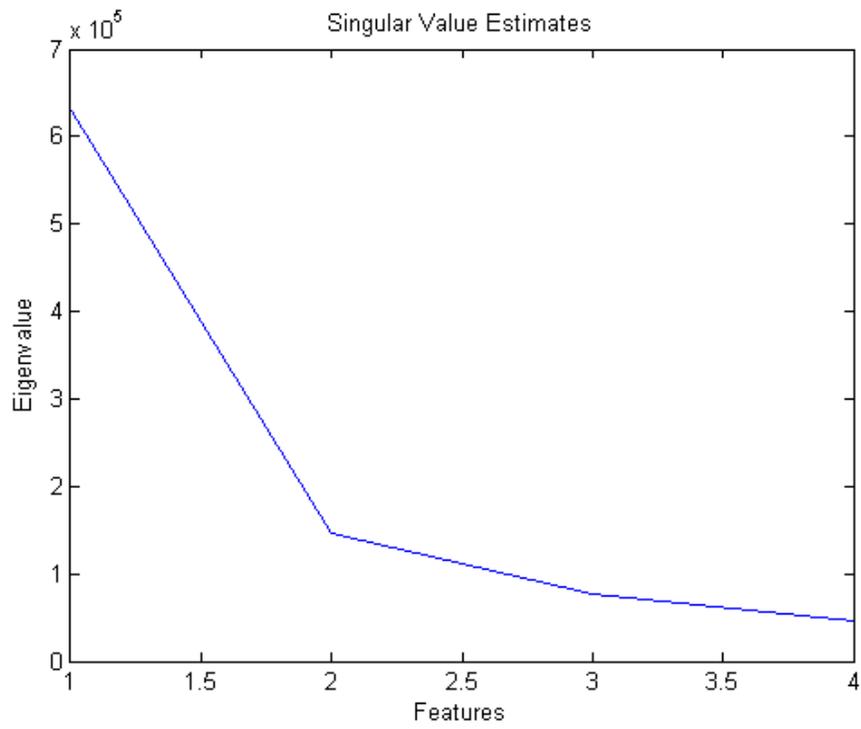


Figure 9-10. Singular value estimates of black spot conditions

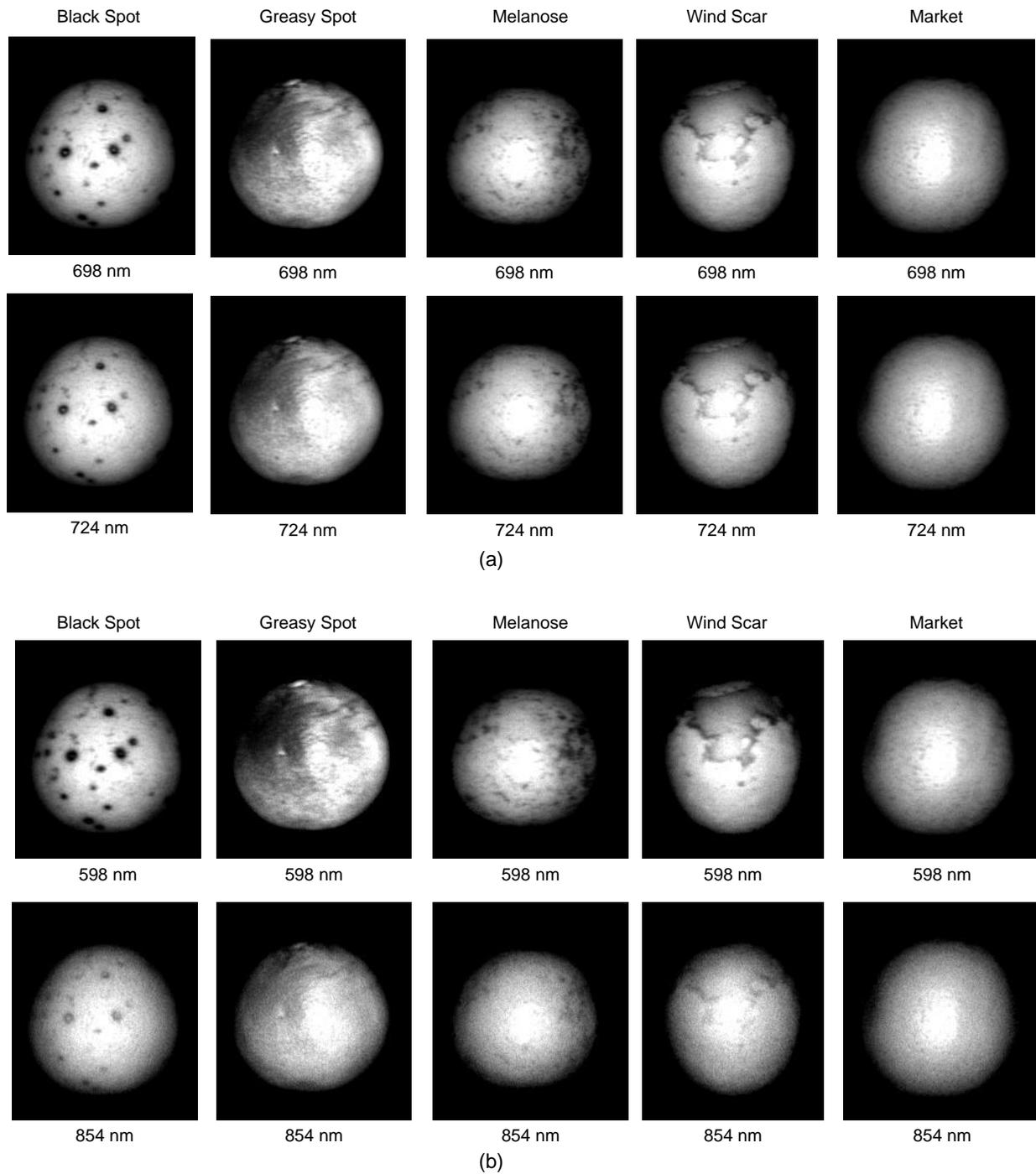


Figure 9-11. The representative single band reflectance images with five different conditions at the wavelengths with (a) highest correlation coefficients and (b) second highest correlation coefficients by CA

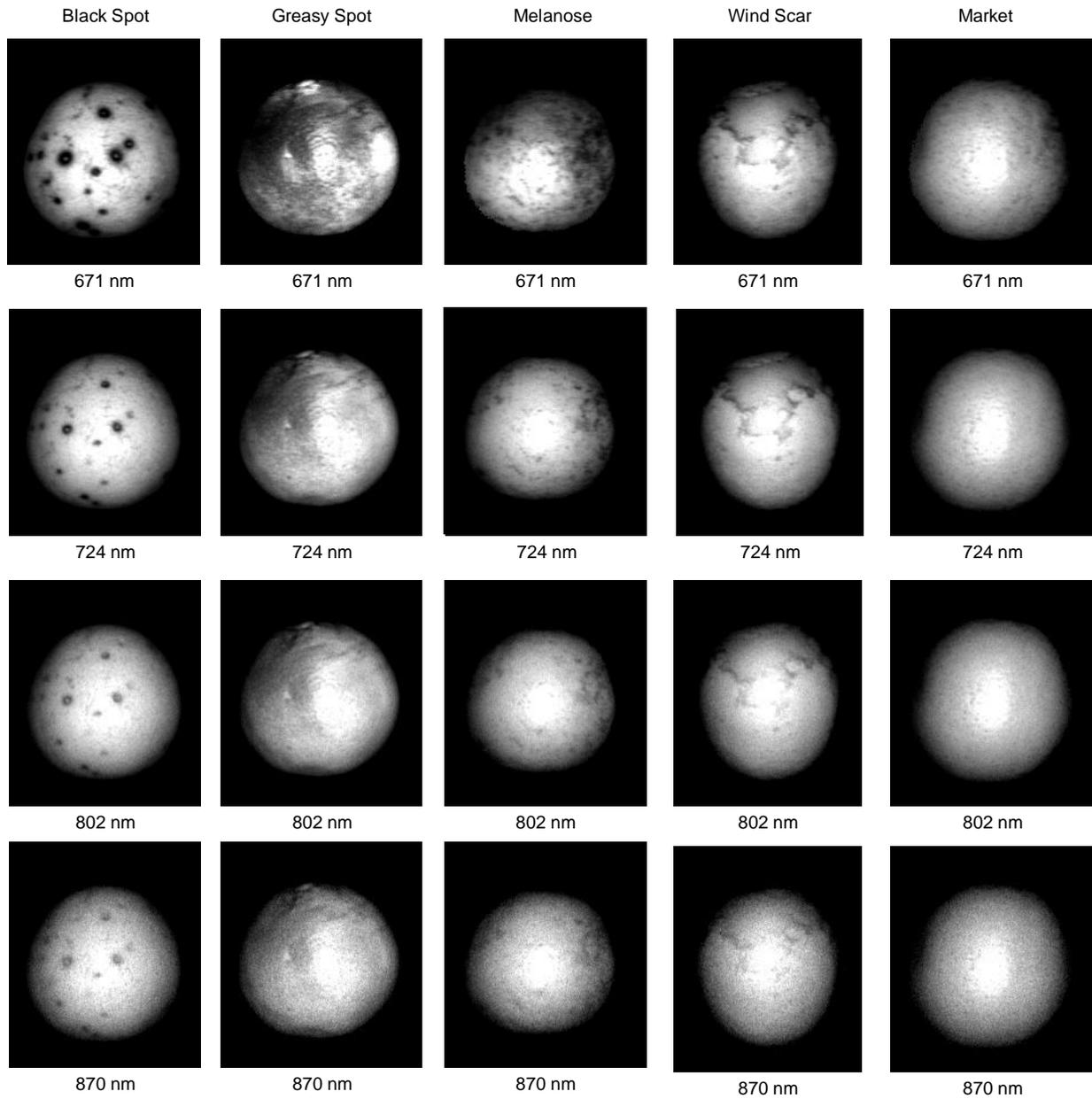


Figure 9-12. Representative spectral reflectance images of diseased peel and normal conditions at four wavelength selected by PCA

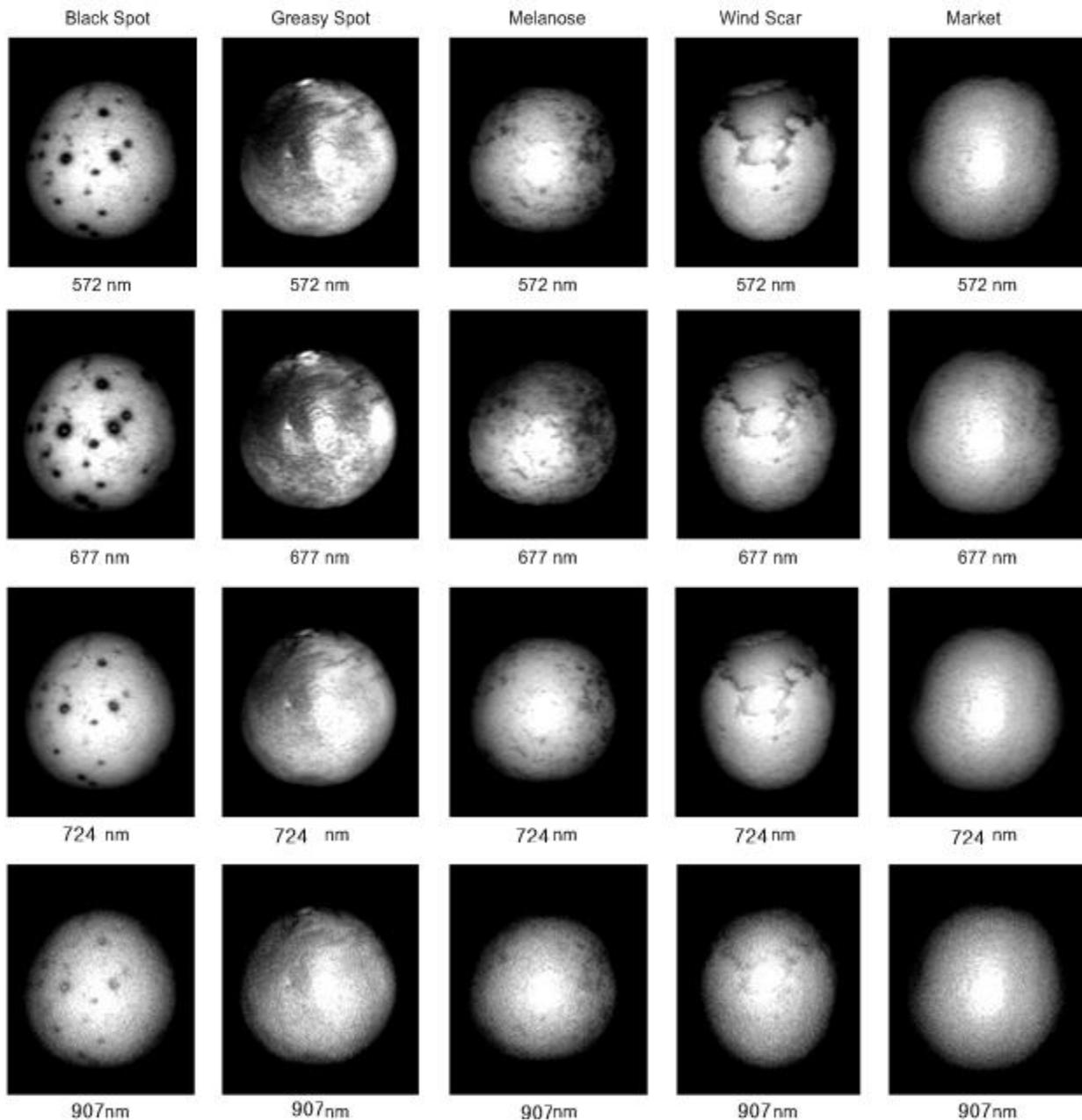


Figure 9-13. Representative spectral reflectance images of diseased peel and normal conditions at four wavelength selected by SVD

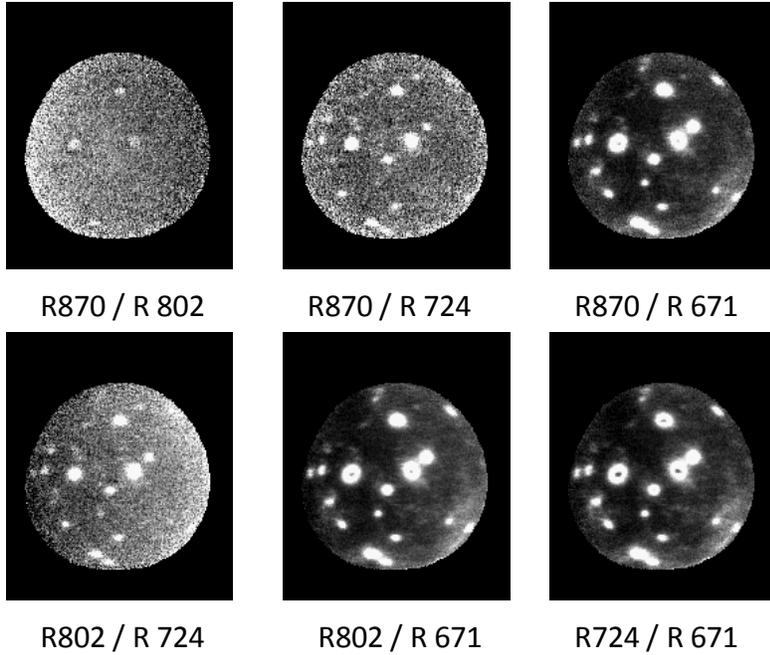


Figure 9-14. The ratio images of black spot using wavelengths selected by PCA

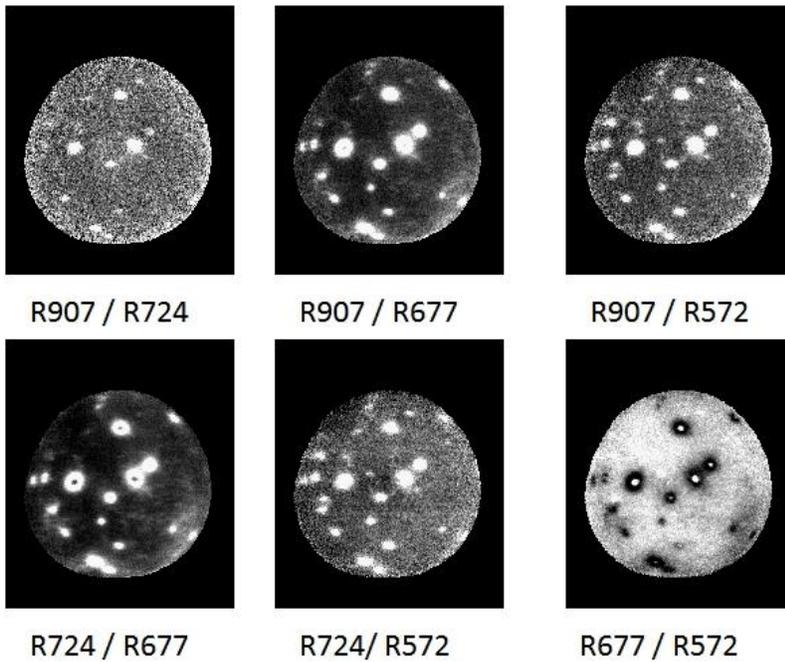
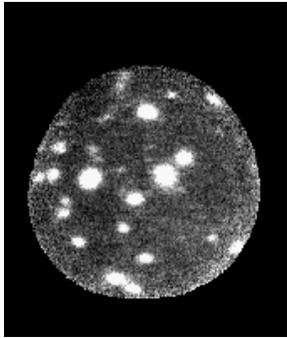
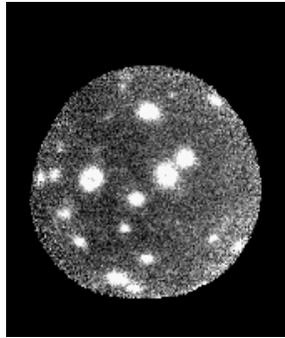


Figure 9-15. The ratio images of black spot using wavelengths selected by SVD



R854 / R 598



R724 / R 698

Figure 9-16. The ratio images of black spot using wavelengths selected by CA

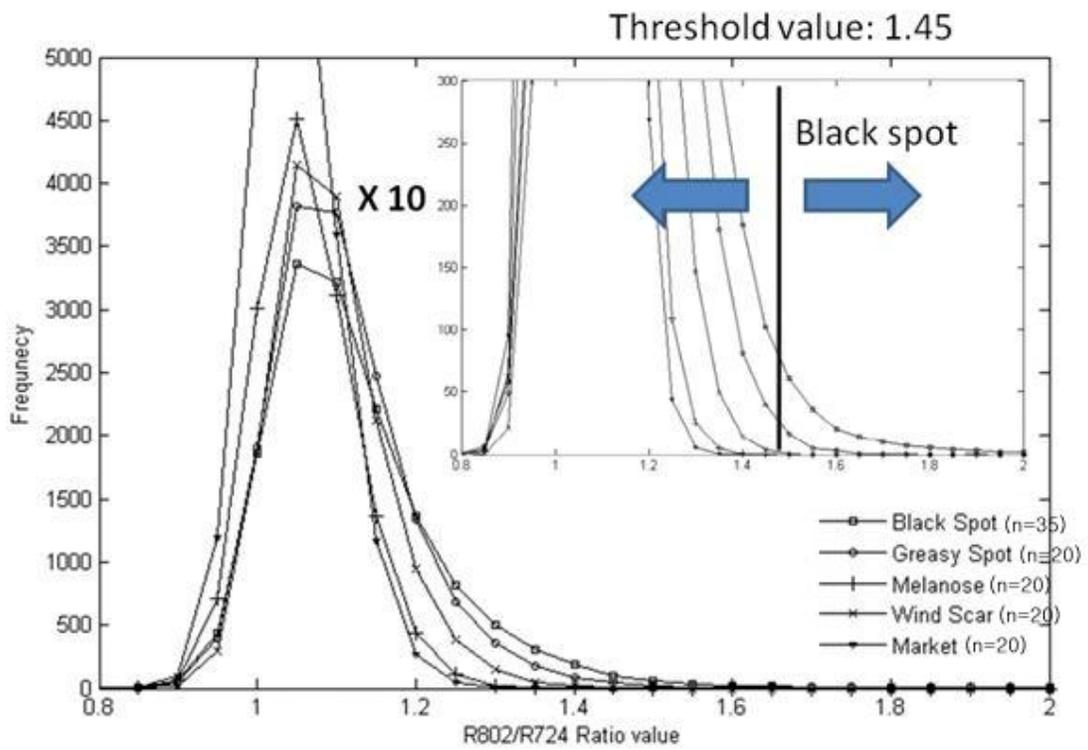


Figure 9-17. The histogram for ratio images of 802 nm/724 nm from PCA

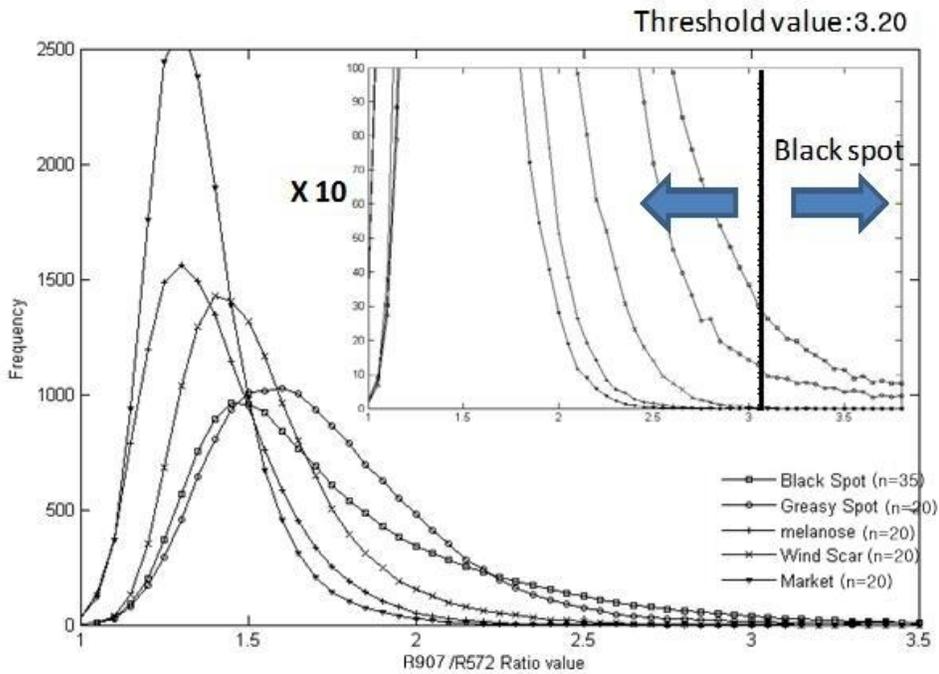


Figure 9-18. The histogram for ratio images of 919 nm/572 nm from SVD

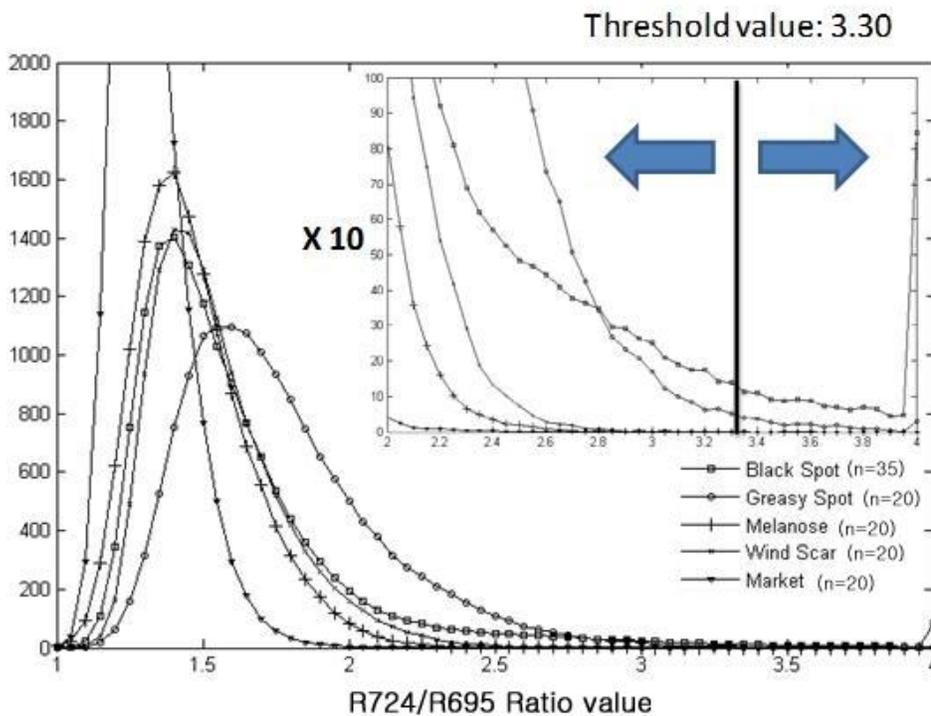


Figure 9-19. The histogram for ratio images of 854 nm/598 nm from CA

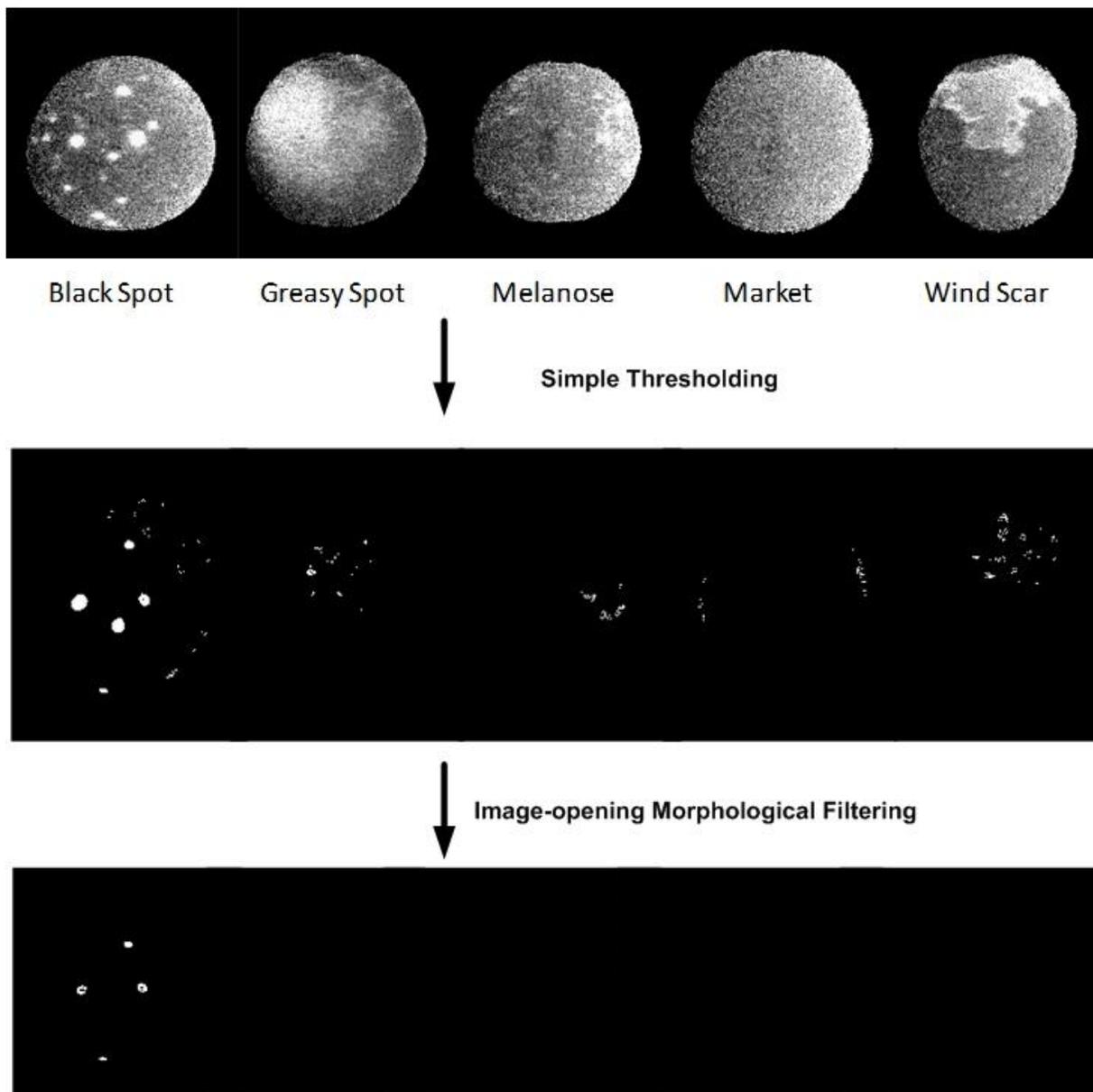


Figure 9-20. Identification of black spot lesions on the fruit peel based on two band ratio images using wavelengths selected by PCA (R802/R724)

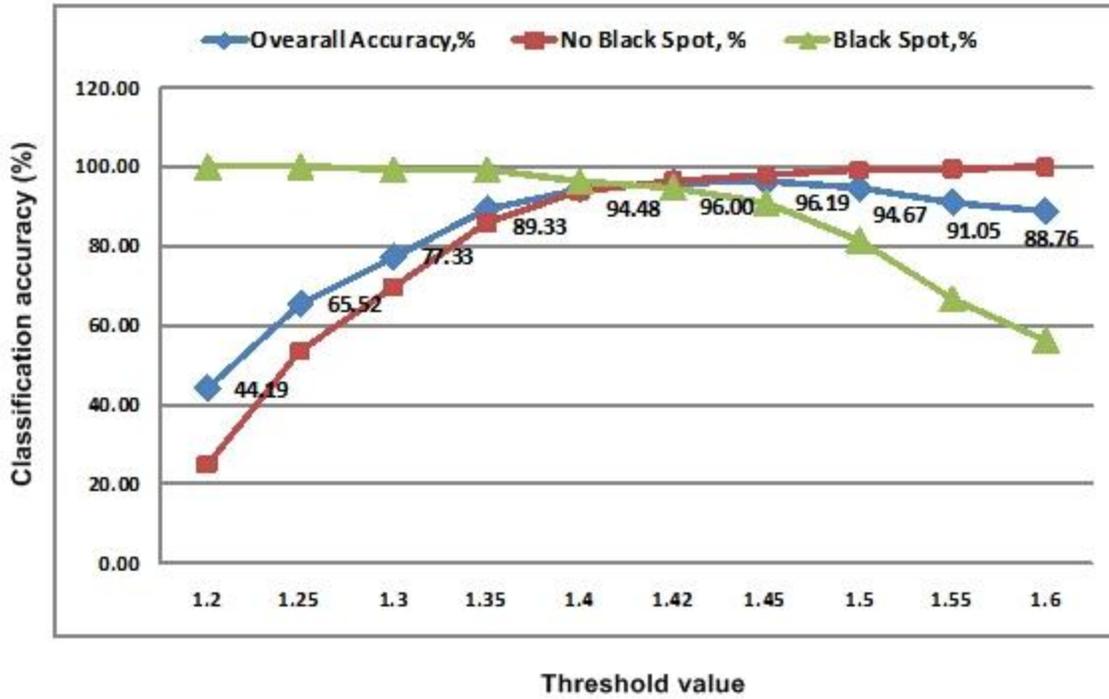


Figure 9-21. Classification accuracies obtained from the ratio image (802 nm/724 nm) selected by PCA

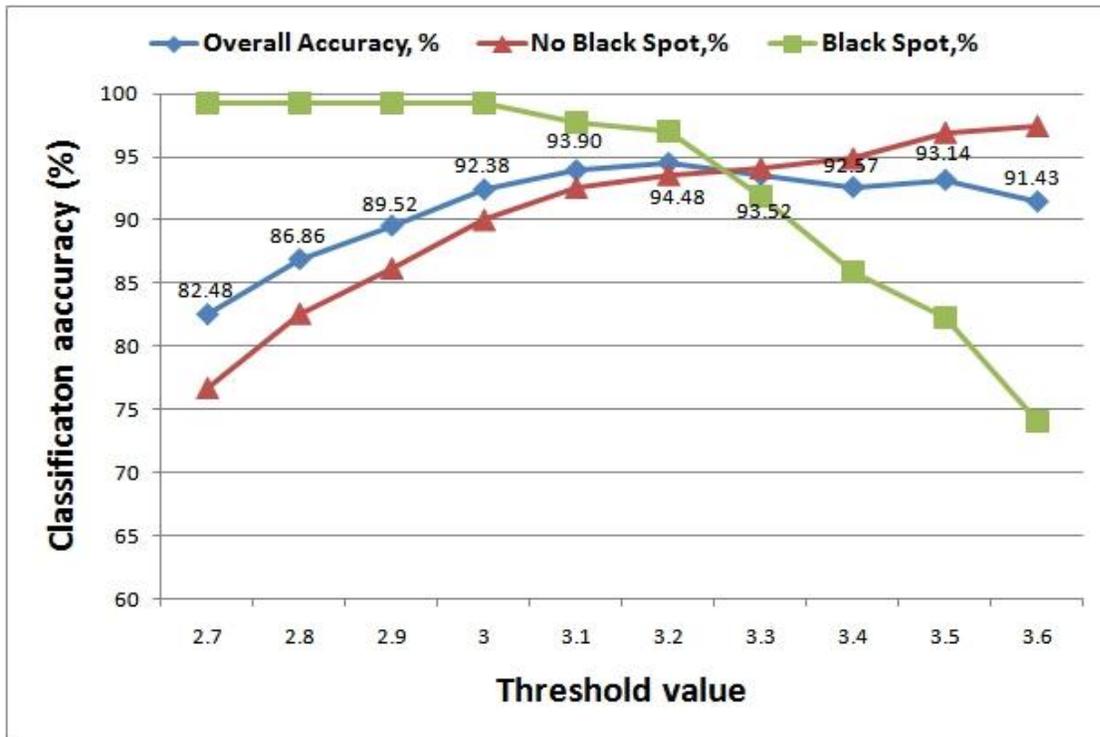


Figure 9-22. Classification accuracies obtained from the ratio image (919 nm/572 nm) selected by SVD

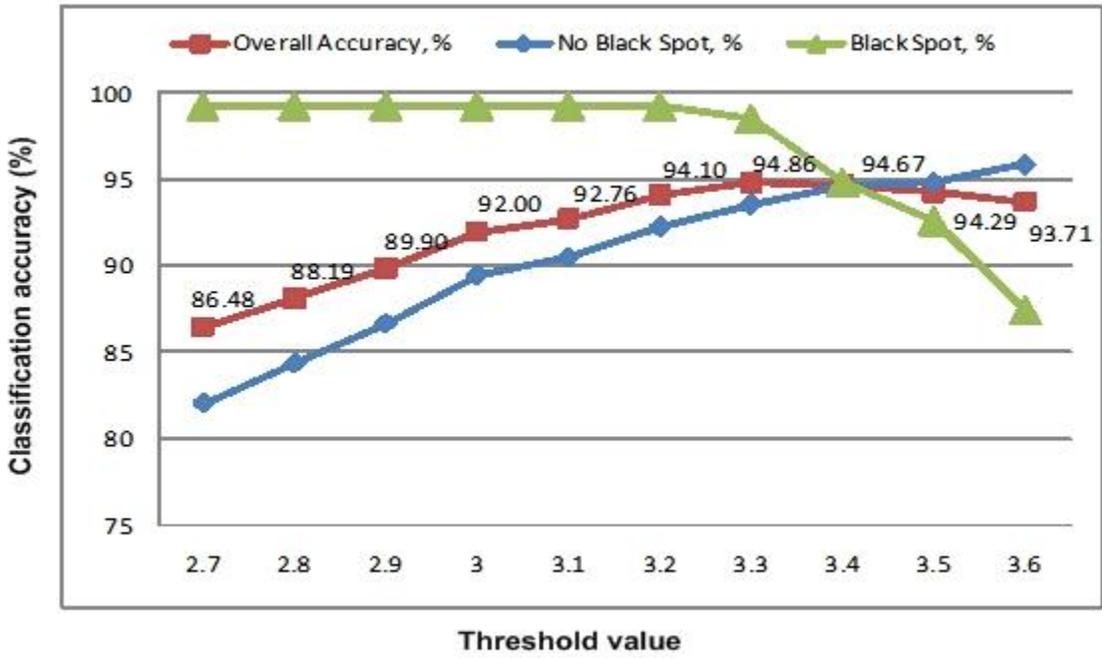


Figure 9-23. Classification accuracies obtained from the ratio image (854 nm/598 nm) selected by CA

Table 9-1. The classification summary for differentiating black spot from other diseases and normal conditions using PCA

Range	NIR/NIR	NIR/Visible				Vis./Vis.
Ratio	R870/ R802	R870/ R724	R820/ R671	R802/ R724	R802/ R671	R724/ R671
Misclassified						
Black Spot (n=135)	102	30	45	7	50	92
Greasy Spot (n=90)	71	3	68	13	75	68
Melanose (n=105)	0	0	16	0	29	29
Wind Scar (n=105)	52	0	22	0	20	19
Market (n=90)	5	0	7	0	14	12
Classification Accuracy, %						
'Black Spot' Class (n=135)	8.88	77.78	66.67	94.81	62.96	31.85
'No Black Spot' Class (n=390)	67.18	99.23	71.03	96.67	64.62	67.18
Overall Accuracy	77.14	93.52	69.90	96.19	64.19	58.1

Table 9-2. The classification summary for differentiating black spot from other diseases and normal conditions using SVD

Range	NIR/Visible			Visible/Visible		
Ratio	R907/ R724	R907/ R677	R917/ R572	R724/ R677	R724/ R572	R677/ R572
Misclassified						
Black Spot (n=135)	44	74	4	35	20	50
Greasy Spot (n=90)	7	52	23	87	31	4
Melanose (n=105)	0	3	0	70	0	9
Wind Scar (n=105)	1	18	2	45	1	24
Market (n=90)	0	3	0	40	1	0
Classification Accuracy, %						
'Black Spot' Class (n=135)	67.40	45.18	97.04	74.07	85.19	62.96
'No Black Spot' Class (n=390)	97.95	80.51	93.59	37.95	91.54	90.51
Overall Accuracy	90.10	74.43	94.48	47.23	89.90	83.43

Table 9-3. The classification summary for differentiating black spot from other diseases and normal conditions using CA

Range	NIR/Visible	Visible/Visible
Ratio	R854/R598	R724/R698
<b>Misclassified</b>		
Black Spot (n=135)	2	59
Greasy Spot (n=90)	25	20
Melanose (n=105)	0	0
Wind Scar (n=105)	0	6
Market (n=90)	0	2
<b>Classification Accuracy, %</b>		
'Black Spot' Class (n=135)	98.51	22.96
'No Black Spot' Class (n=390)	93.59	92.82
Overall Accuracy	94.86	73.14

## CHAPTER 10 CONCLUSION AND FUTURE WORK

### **Conclusions**

Florida has two important citrus markets: the fresh citrus fruits market and the processed citrus products market. However, citrus market in Florida is currently threatened by three major citrus diseases; citrus greening, canker, and black spot. For these reasons, a lot of effort has been put into automating the detection of citrus diseases using machine vision inspection with great success.

This research started with a color imaging system. The color imaging system provided an efficient laboratory-based detection method for citrus canker and greening diseases, and showed such techniques image processing and pattern recognition have the potential for automated disease detection applications. Moreover, this research demonstrated that color imaging and texture feature analysis could be used for differentiating citrus diseases under the controlled laboratory lighting conditions. Model HSI at citrus canker and greening diseases conditions emerged as the best data model for classification. Model HSI provided good results averaging above 95% overall classification for both canker disease and citrus greening disease. Similarly, the HSI models gave best classification accuracy (above 90%) for the pattern recognition method comparison. Through a series of random tests, the stability of classification accuracy was demonstrated.

This research also presents new applications of hyperspectral imaging for measuring the optical properties of fruits and disease blemishes. In this study, a hyperspectral imaging system was developed to recognize fruits exhibiting symptoms of citrus diseases along with fruits with other peel conditions and market quality fruits. For

real time citrus disease detection using multispectral algorithms, band selection algorithms for the hyperspectral image data were developed. Selected bands have great potential to be adopted by a multispectral imaging system for real-time citrus disease detection. As shown in Chapter 9, a black spot classification accuracy of 96.19% was obtained using the PCA approach with the NIR and visible ratio image (802 nm/724 nm) and an optimal threshold value of 1.5. The SVD approach had a black spot classification accuracy of 94.29% with the NIR and visible ratio image (907 nm/572 nm) and an optimal threshold value (3.20). In CA algorithms, the (854 nm/598 nm) ratio images that had second highest correlation coefficient (9.1) provided the best classification accuracy (94.86%).

### **Future Work**

This dissertation provided the development of the efficient laboratory-based detection method for citrus diseases. Although the overall performance of this research was good, significant future work will develop new outdoor-based algorithms that could be applied to autonomous scouting. The research on computer vision algorithms for citrus diseases detection encompassed various disease samples and includes numerous computer vision and image processing techniques for the overall detection system.

The selection of disease sample was guided by season and disease symptoms. The initial leaf samples for citrus greening were collected during the fall season, while later samples were completed in the spring season. The samples in spring season were collected in March, 2008 but the samples were not suitable for research, because the number of samples was small and some samples were dead.

For further work, an extensive research of detection for citrus canker and greening is planned with new spring samples. This work will build on the research data presented in spring and fall seasons. It is anticipated that the focus of this study will establish an improved image processing approach and classification method.

The effect of the size of the disease lesions was not studied in this research. Morphological filtering in the image processing step has the risk of missing the fruit with small size disease lesions. Therefore, future work will be performed to find the detection limit for the size of the canker lesions as well as the effects of fruit harvest time on citrus disease identification. The development of a real-time citrus disease detection system is the one of important further works. This system will be implemented on a small-scale fruit transport machine. The optical inspection module will be designed based on a two-band spectral imaging system.

APPENDIX A  
MATLAB CODE FILES FOR EDGE DETECTION

```
%% Matlab code for Edge Detection
%% Program developed by Dae Gwan Kim, Research Assistant

clear;
for i=1:62% The number of image samples
    cd('C:\Documents and
Settings\student\Desktop\Research\greening\greening_0903\mn');
    % read the image name (load the images numbers)
    % ex) image_1.bmp, image_2.bmp, image_3.bmp, image_4.bmp, .....
    original_name='mn_st_f_';
    % [X,map] = imread([original_name '.jpg']);

    % Preprocessing images for edge detection
    X = double(imread([original_name int2str(i) '.jpg']))/255;
    X=imresize(X,0.5);

    % converting RGB images to Binary images
    hsvIm = rgb2hsv(X);
    im =hsvIm(:,:,2);
    oim=imadjust(im);
    oim = oim/max(max(oim));
    mim2 = immultiply(im, oim);
    mim3=imadjust(mim2);
    oim3 = mim3/max(max(mim3));
    mim4 = immultiply(mim3, oim3);
    mim5=imadjust(mim4);
    BW = double(im2bw(im,0.5));

    % Create morphological structuring element
    se1 = strel('diamond', 2);
    se2 = strel('diamond',2);

    % control image edge using 'imdilate' and 'imerode' commands
    BW = imdilate(BW,se1);
    BW = imerode(BW,se2);
    BW = imdilate(BW,se1);
    BW = imdilate(BW,se1);
```

```

BW = imdilate(BW,se1);
BW = imerode(BW,se2);
% Fill image regions and holes
BW=imfill(BW,'holes');
fim(:,:,1)=immultiply(X(:,:,1),BW);
fim(:,:,2)=immultiply(X(:,:,2),BW);
fim(:,:,3)=immultiply(X(:,:,3),BW);
figure(3)
imshow(fim)
cd('C:\Documents and
Settings\student\Desktop\Research\greening\greening_0903\mn\edge');
% write resized image
% ex) 64x64_image_1.png, 64x64_image_2.png, ...
imwrite(fim, ['edge_' original_name int2str(i) '.jpg'], 'jpg');
end

```

APPENDIX B  
MATLAB CODE FILES FOR PATTERN RECOGNITION METHODS

```
%% Code Files for Pattern Recognition Methods
%% Citrus Greening Identification Project for Models
%% Program developed by Dae Gwan Kim, Research Assistant

% '1' represents Blotch Mottle
% '2' represents Iron Deficiency
% '3' represents Islands
% '4' represents MN Deficiency
% '5' represents Normal leaf
% '6' represents Zinc Deficiency

%% Blotchy mottle

clear;
% load each data sets
data=load('blotch_0907.mat');
h=data.blotch_0907_h;
s=data.blotch_0907_s;
i=data.blotch_0907_i;
% load reduced texture features
% HSI_15
tr_bm_hsi=[s(5,1:30);i(2,1:30);h(7,1:30);h(2,1:30);s(6,1:30);s(4,1:30);h(9,1:30);s(8,1:30);
i(6,1:30);s(13,1:30);h(4,1:30);i(4,1:30);i(13,1:30);s(7,1:30);i(7,1:30)];
ts_bm_hsi=[s(5,31:60);i(2,31:60);h(7,31:60);h(2,31:60);s(6,31:60);s(4,31:60);h(9,31:60);
s(8,31:60);i(6,31:60);s(13,31:60);h(4,31:60);i(4,31:60);i(13,31:60);s(7,31:60);i(7,31:60)];
% HS_9
tr_bm_hs=[s(5,1:30);h(7,1:30);h(5,1:30);h(12,1:30);s(4,1:30);s(7,1:30);h(8,1:30);s(8,1:30);
h(3,1:30);s(11,1:30)];
ts_bm_hs=[s(5,31:60);h(7,31:60);h(5,31:60);h(12,31:60);s(4,31:60);s(7,31:60);h(8,31:60);
s(8,31:60);h(3,31:60);s(11,31:60)];
% I_11
tr_bm_i=[i(2,1:30);i(8,1:30);i(9,1:30);i(6,1:30);i(5,1:30);i(7,1:30);i(10,1:30);i(1,1:30)];
ts_bm_i=[i(2,31:60);i(8,31:60);i(9,31:60);i(6,31:60);i(5,31:60);i(7,31:60);i(10,31:60);i(1,31:60)];
% HSI_39
tr_bm_hsi39=[h(:,1:30);s(:,1:30);i(:,1:30)];
ts_bm_hsi39=[h(:,31:60);s(:,31:60);i(:,31:60)];

%% Iron Deficiency

% load each data sets
data=load('iron_0907.mat');
h=data.iron_0907_h;
```

```

s=data.iron_0907_s;
i=data.iron_0907_i;
% load reduced texture features
% HSI_15
tr_iron_hsi=[s(5,1:30);i(2,1:30);h(7,1:30);h(2,1:30);s(6,1:30);s(4,1:30);h(9,1:30);s(8,1:30)
;i(6,1:30);s(13,1:30);h(4,1:30);i(4,1:30);i(13,1:30);s(7,1:30);i(7,1:30)];
ts_iron_hsi=[s(5,31:60);i(2,31:60);h(7,31:60);h(2,31:60);s(6,31:60);s(4,31:60);h(9,31:60)
;s(8,31:60);i(6,31:60);s(13,31:60);h(4,31:60);i(4,31:60);i(13,31:60);s(7,31:60);i(7,31:60)];
% HS_9
tr_iron_hs=[s(5,1:30);h(7,1:30);h(5,1:30);h(12,1:30);s(4,1:30);s(7,1:30);h(8,1:30);s(8,1:30);
h(3,1:30);s(11,1:30)];
ts_iron_hs=[s(5,31:60);h(7,31:60);h(5,31:60);h(12,31:60);s(4,31:60);s(7,31:60);h(8,31:60);
s(8,31:60);h(3,31:60);s(11,31:60)];
% I_11
tr_iron_i=[i(2,1:30);i(8,1:30);i(9,1:30);i(6,1:30);i(5,1:30);i(7,1:30);i(10,1:30);i(1,1:30)];
ts_iron_i=[i(2,31:60);i(8,31:60);i(9,31:60);i(6,31:60);i(5,31:60);i(7,31:60);i(10,31:60);i(1,31:60)];
% HSI_39
tr_iron_hsi39=[h(:,1:30);s(:,1:30);i(:,1:30)];
ts_iron_hsi39=[h(:,31:60);s(:,31:60);i(:,31:60)];

```

%% Islands

```

% load each data sets
data=load('island_0907.mat');
h=data.island_0907_h;
s=data.island_0907_s;
i=data.island_0907_i;
% load reduced texture features
% HSI_15
tr_is_hsi=[s(5,1:30);i(2,1:30);h(7,1:30);h(2,1:30);s(6,1:30);s(4,1:30);h(9,1:30);s(8,1:30);i(
6,1:30);s(13,1:30);h(4,1:30);i(4,1:30);i(13,1:30);s(7,1:30);i(7,1:30)];
ts_is_hsi=[s(5,31:60);i(2,31:60);h(7,31:60);h(2,31:60);s(6,31:60);s(4,31:60);h(9,31:60);s(
8,31:60);i(6,31:60);s(13,31:60);h(4,31:60);i(4,31:60);i(13,31:60);s(7,31:60);i(7,31:60)];
% HS_9
tr_is_hs=[s(5,1:30);h(7,1:30);h(5,1:30);h(12,1:30);s(4,1:30);s(7,1:30);h(8,1:30);s(8,1:30);
h(3,1:30);s(11,1:30)];
ts_is_hs=[s(5,31:60);h(7,31:60);h(5,31:60);h(12,31:60);s(4,31:60);s(7,31:60);h(8,31:60);
s(8,31:60);h(3,31:60);s(11,31:60)];
% I_11
tr_is_i=[i(2,1:30);i(8,1:30);i(9,1:30);i(6,1:30);i(5,1:30);i(7,1:30);i(10,1:30);i(1,1:30)];
ts_is_i=[i(2,31:60);i(8,31:60);i(9,31:60);i(6,31:60);i(5,31:60);i(7,31:60);i(10,31:60);i(1,31:
60)];
% HSI_39
tr_is_hsi39=[h(:,1:30);s(:,1:30);i(:,1:30)];
ts_is_hsi39=[h(:,31:60);s(:,31:60);i(:,31:60)];

```

```
%% MN Deficiency
```

```
% load each data sets
```

```
data=load('mn_0907.mat');
```

```
h=data.mn_0907_h;
```

```
s=data.mn_0907_s;
```

```
i=data.mn_0907_i;
```

```
% load reduced texture features
```

```
% HSI_15
```

```
tr_mn_hsi=[s(5,1:30);i(2,1:30);h(7,1:30);h(2,1:30);s(6,1:30);s(4,1:30);h(9,1:30);s(8,1:30);  
i(6,1:30);s(13,1:30);h(4,1:30);i(4,1:30);i(13,1:30);s(7,1:30);i(7,1:30)];
```

```
ts_mn_hsi=[s(5,31:60);i(2,31:60);h(7,31:60);h(2,31:60);s(6,31:60);s(4,31:60);h(9,31:60);  
s(8,31:60);i(6,31:60);s(13,31:60);h(4,31:60);i(4,31:60);i(13,31:60);s(7,31:60);i(7,31:60)];
```

```
% HS_9
```

```
tr_mn_hs=[s(5,1:30);h(7,1:30);h(5,1:30);h(12,1:30);s(4,1:30);s(7,1:30);h(8,1:30);s(8,1:30)  
);h(3,1:30);s(11,1:30)];
```

```
ts_mn_hs=[s(5,31:60);h(7,31:60);h(5,31:60);h(12,31:60);s(4,31:60);s(7,31:60);h(8,31:60)  
);s(8,31:60);h(3,31:60);s(11,31:60)];
```

```
% I_11
```

```
tr_mn_i=[i(2,1:30);i(8,1:30);i(9,1:30);i(6,1:30);i(5,1:30);i(7,1:30);i(10,1:30);i(1,1:30)];
```

```
ts_mn_i=[i(2,31:60);i(8,31:60);i(9,31:60);i(6,31:60);i(5,31:60);i(7,31:60);i(10,31:60);i(1,31  
:60)];
```

```
% HSI_39
```

```
tr_mn_hsi39=[h(:,1:30);s(:,1:30);i(:,1:30)];
```

```
ts_mn_hsi39=[h(:,31:60);s(:,31:60);i(:,31:60)];
```

```
%% Normal Leaves
```

```
% load each data sets
```

```
data=load('no_0907.mat');
```

```
h=data.no_0907_h;
```

```
s=data.no_0907_s;
```

```
i=data.no_0907_i;
```

```
% load reduced texture features
```

```
% HSI_15
```

```
tr_nl_hsi=[s(5,1:30);i(2,1:30);h(7,1:30);h(2,1:30);s(6,1:30);s(4,1:30);h(9,1:30);s(8,1:30);i(  
6,1:30);s(13,1:30);h(4,1:30);i(4,1:30);i(13,1:30);s(7,1:30);i(7,1:30)];
```

```
ts_nl_hsi=[s(5,31:60);i(2,31:60);h(7,31:60);h(2,31:60);s(6,31:60);s(4,31:60);h(9,31:60);s(  
8,31:60);i(6,31:60);s(13,31:60);h(4,31:60);i(4,31:60);i(13,31:60);s(7,31:60);i(7,31:60)];
```

```
% HS_9
```

```
tr_nl_hs=[s(5,1:30);h(7,1:30);h(5,1:30);h(12,1:30);s(4,1:30);s(7,1:30);h(8,1:30);s(8,1:30);  
h(3,1:30);s(11,1:30)];
```

```
ts_nl_hs=[s(5,31:60);h(7,31:60);h(5,31:60);h(12,31:60);s(4,31:60);s(7,31:60);h(8,31:60);  
s(8,31:60);h(3,31:60);s(11,31:60)];
```

```

% I_11
tr_nl_i=[i(2,1:30);i(8,1:30);i(9,1:30);i(6,1:30);i(5,1:30);i(7,1:30);i(10,1:30);i(1,1:30)];
ts_nl_i=[i(2,31:60);i(8,31:60);i(9,31:60);i(6,31:60);i(5,31:60);i(7,31:60);i(10,31:60);i(1,31:
60)];
% HSI_39
tr_nl_hsi39=[h(:,1:30);s(:,1:30);i(:,1:30)];
ts_nl_hsi39=[h(:,31:60);s(:,31:60);i(:,31:60)];

%% Zinc Deficiency

% load each data sets
h=data.zinc_0907_h;
s=data.zinc_0907_s;
i=data.zinc_0907_i;
% load reduced texture features
% HSI_15
tr_zinc_hsi=[s(5,1:30);i(2,1:30);h(7,1:30);h(2,1:30);s(6,1:30);s(4,1:30);h(9,1:30);s(8,1:30
);i(6,1:30);s(13,1:30);h(4,1:30);i(4,1:30);i(13,1:30);s(7,1:30);i(7,1:30)];
ts_zinc_hsi=[s(5,31:60);i(2,31:60);h(7,31:60);h(2,31:60);s(6,31:60);s(4,31:60);h(9,31:60)
;s(8,31:60);i(6,31:60);s(13,31:60);h(4,31:60);i(4,31:60);i(13,31:60);s(7,31:60);i(7,31:60)];
% HS_9
tr_zinc_hs=[s(5,1:30);h(7,1:30);h(5,1:30);h(12,1:30);s(4,1:30);s(7,1:30);h(8,1:30);s(8,1:3
0);h(3,1:30);s(11,1:30)];
ts_zinc_hs=[s(5,31:60);h(7,31:60);h(5,31:60);h(12,31:60);s(4,31:60);s(7,31:60);h(8,31:6
0);s(8,31:60);h(3,31:60);s(11,31:60)];
% I_11
tr_zinc_i=[i(2,1:30);i(8,1:30);i(9,1:30);i(6,1:30);i(5,1:30);i(7,1:30);i(10,1:30);i(1,1:30)];
ts_zinc_i=[i(2,31:60);i(8,31:60);i(9,31:60);i(6,31:60);i(5,31:60);i(7,31:60);i(10,31:60);i(1,3
1:60)];
% HSI_39
tr_zinc_hsi39=[h(:,1:30);s(:,1:30);i(:,1:30)];
ts_zinc_hsi39=[h(:,31:60);s(:,31:60);i(:,31:60)];

%% PRTools for Pattern Recognition

% make each training set
hsi_training=dataset([tr_bm_hsi;tr_iron_hsi;tr_is_hsi;tr_mn_hsi;tr_nl_hsi;tr_zinc_hsi],genlab
([30 30 30 30 30 30],[1 2 3 4 5 6]));
hs_training=dataset([tr_bm_hs;tr_iron_hs;tr_is_hs;tr_mn_hs;tr_nl_hs;tr_zinc_hs],genlab
([30 30 30 30 30 30],[1 2 3 4 5 6]));
i_training=dataset([tr_bm_i;tr_iron_i;tr_is_i;tr_mn_i;tr_nl_i;tr_zinc_i],genlab([30 30 30 30
30 30],[1 2 3 4 5 6]));
hsi39_training=dataset([tr_bm_hsi39;tr_iron_hsi39;tr_is_hsi39;tr_mn_hsi39;tr_nl_hsi39;t
r_zinc_hsi39],genlab([30 30 30 30 30 30],[1 2 3 4 5 6]));

% make each test set for HSI_15 model

```

```

hsi_test_bm=dataset([ts_bm_hsi],genlab([30],[1]));
hsi_test_iron=dataset([ts_iron_hsi],genlab([30],[2]));
hsi_test_is=dataset([ts_is_hsi],genlab([30],[3]));
hsi_test_mn=dataset([ts_mn_hsi],genlab([30],[4]));
hsi_test_nl=dataset([ts_nl_hsi],genlab([30],[5]));
hsi_test_zinc=dataset([ts_zinc_hsi],genlab([30],[6]));

% make each test set for HS_9 model
hs_test_bm=dataset([ts_bm_hsi],genlab([30],[1]));
hs_test_iron=dataset([ts_iron_hs],genlab([30],[2]));
hs_test_is=dataset([ts_is_hs],genlab([30],[3]));
hs_test_mn=dataset([ts_mn_hs],genlab([30],[4]));
hs_test_nl=dataset([ts_nl_hs],genlab([30],[5]));
hs_test_zinc=dataset([ts_zinc_hs],genlab([30],[6]));

% make each test set for I_11 model
i_test_bm=dataset([ts_bm_i],genlab([30],[1]));
i_test_iron=dataset([ts_iron_i],genlab([30],[2]));
i_test_is=dataset([ts_is_i],genlab([30],[3]));
i_test_mn=dataset([ts_mn_i],genlab([30],[4]));
i_test_nl=dataset([ts_nl_i],genlab([30],[5]));
i_test_zinc=dataset([ts_zinc_i],genlab([30],[6]));

% make each test set for HSI_39 model
hsi39_test_bm=dataset([ts_bm_hsi39],genlab([30],[1]));
hsi39_test_iron=dataset([ts_iron_hsi39],genlab([30],[2]));
hsi39_test_is=dataset([ts_is_hsi39],genlab([30],[3]));
hsi39_test_mn=dataset([ts_mn_hsi39],genlab([30],[4]));
hsi39_test_nl=dataset([ts_nl_hsi39],genlab([30],[5]));
hsi39_test_zinc=dataset([ts_zinc_hsi39],genlab([30],[6]));

% Three Pattern Recognition methods for HSI_15 model
hsi_fisher=fisherc(hsi_training);
hsi_bp=bpxnc(hsi_training);
hsi_svc=svc(hsi_training,'p',2);

% Three Pattern Recognition methods for HS_9 model
hs_fisher=fisherc(hs_training);
hs_bp=bpxnc(hs_training);
hs_svc=svc(hs_training,'p',2);

% Three Pattern Recognition methods for I_11 model
i_fisher=fisherc(i_training);
i_bp=bpxnc(i_training);
i_svc=svc(i_training,'p',2);

```

```
% Three Pattern Recognition methods for HSI_15 model
hsi39_fisher=fisherc(hsi39_training);
hsi39_bp=bpxnc(hsi39_training);
hsi39_svc=svc(hsi39_training,'p',2);
```

```
%% classification results using FDA
```

```
% Classification results for HSI_15 Model
blotchy_mottle_hsi=testc(hsi_test_bm*hsi_fisher)
iron_hsi=testc(hsi_test_iron*hsi_fisher)
island_hsi=testc(hsi_test_is*hsi_fisher)
mn_hsi=testc(hsi_test_mn*hsi_fisher)
normal_leave_hsi=testc(hsi_test_nl*hsi_fisher)
zinc_hsi=testc(hsi_test_zinc*hsi_fisher)
```

```
% Classification results for HS_9 Model
blotchy_mottle_hs=testc(hs_test_bm*hs_fisher)
iron_hs=testc(hs_test_iron*hs_fisher)
island_hs=testc(hs_test_is*hs_fisher)
mn_hs=testc(hs_test_mn*hs_fisher)
normal_leave_hs=testc(hs_test_nl*hs_fisher)
zinc_hs=testc(hs_test_zinc*hs_fisher)
```

```
% Classification results for I_11 Model
blotchy_mottle_i=testc(i_test_bm*i_fisher)
iron_i=testc(i_test_iron*i_fisher)
island_i=testc(i_test_is*i_fisher)
mn_i=testc(i_test_mn*i_fisher)
normal_leave_i=testc(i_test_nl*i_fisher)
zinc_i=testc(i_test_zinc*i_fisher)
```

```
% Classification results for HSI_39 Model
blotchy_mottle_hsi39=testc(hsi39_test_bm*i_fisher)
iron_hsi39=testc(hsi39_test_iron*hsi39_fisher)
island_hsi39=testc(hsi39_test_is*hsi39_fisher)
mn_hsi39=testc(hsi39_test_mn*hsi39_fisher)
normal_leave_hsi39=testc(hsi39_test_nl*hsi39_fisher)
zinc_hsi39=testc(hsi39_test_zinc*hsi39_fisher)
```

```
%% classification results using BP Neural Network
```

```
% Classification results for HSI_15 Model
blotchy_mottle_hsi=testc(hsi_test_bm*hsi_bp)
iron_hsi=testc(hsi_test_iron*hsi_bp)
island_hsi=testc(hsi_test_is*hsi_bp)
mn_hsi=testc(hsi_test_mn*hsi_bp)
```

```
normal_leave_hsi=testc(hsi_test_ni*hsi_bp)
zinc_hsi=testc(hsi_test_zinc*hsi_bp)
```

```
% Classification results for HS_9 Model
blotchy_mottle_hs=testc(hs_test_bm*hs_bp)
iron_hs=testc(hs_test_iron*hs_bp)
island_hs=testc(hs_test_is*hs_bp)
mn_hs=testc(hs_test_mn*hs_bp)
normal_leave_hs=testc(hs_test_ni*hs_bp)
zinc_hs=testc(hs_test_zinc*hs_bp)
```

```
% Classification results for I_11 Model
blotchy_mottle_i=testc(i_test_bm*i_bp)
iron_i=testc(i_test_iron*i_bp)
island_i=testc(i_test_is*i_bp)
mn_i=testc(i_test_mn*i_bp)
normal_leave_i=testc(i_test_ni*i_bp)
zinc_i=testc(i_test_zinc*i_bp)
```

```
% Classification results for HSI_39 Model
blotchy_mottle_hsi39=testc(hsi39_test_bm*hsi_bp)
iron_hsi39=testc(hsi39_test_iron*hsi39_bp)
island_hsi39=testc(hsi39_test_is*hsi39_bp)
mn_hsi39=testc(hsi39_test_mn*hsi39_bp)
normal_leave_hsi39=testc(hsi39_test_ni*hsi39_bp)
zinc_hsi39=testc(hsi39_test_zinc*hsi39_bp)
```

```
%% classification results using SVC
```

```
% Classification results for HSI_15 Model
blotchy_mottle_hsi=testc(hsi_test_bm*hsi_svc)
iron_hsi=testc(hsi_test_iron*hsi_svc)
island_hsi=testc(hsi_test_is*hsi_svc)
mn_hsi=testc(hsi_test_mn*hsi_svc)
normal_leave_hsi=testc(hsi_test_ni*hsi_svc)
zinc_hsi=testc(hsi_test_zinc*hsi_svc)
```

```
% Classification results for HS_9 Model
blotchy_mottle_hs=testc(hs_test_bm*hs_svc)
iron_hs=testc(hs_test_iron*hs_svc)
island_hs=testc(hs_test_is*hs_svc)
mn_hs=testc(hs_test_mn*hs_svc)
normal_leave_hs=testc(hs_test_ni*hs_svc)
zinc_hs=testc(hs_test_zinc*hs_svc)
```

```
% Classification results for I_11 Model
blotchy_mottle_i=testc(i_test_bm*i_svc)
iron_i=testc(i_test_iron*i_svc)
island_i=testc(i_test_is*i_svc)
mn_i=testc(i_test_mn*i_svc)
normal_leave_i=testc(i_test_nl*i_svc)
zinc_i=testc(i_test_zinc*i_svc)
```

```
% Classification results for HSI_39 Model
blotchy_mottle_hsi39=testc(hsi39_test_bm*hsi39_svc)
iron_hsi39=testc(hsi39_test_iron*hsi39_svc)
island_hsi39=testc(hsi39_test_is*hsi39_svc)
mn_hsi39=testc(hsi39_test_mn*hsi39_svc)
normal_leave_hsi39=testc(hsi39_test_nl*hsi39_svc)
zinc_hsi39=testc(hsi39_test_zinc*hsi39_svc)
```

APPENDIX C  
MATLAB CODE FILES FOR PATTERN HYPERSPECTRAL IMAGING

```
%Demonstrate procedures for hyperspectral imaging processing
% Read original hyperspectral images of dark current and white panel
R_white = hsi_read ('white_panel_051710');
R_dark = hsi_read ('dark_current_051710');
save bs_wd;
clear;
%% Read original hyperspectral images of fruit samples
for y=1:9
    R_sample = hsi_read (['bs_' int2str(y)]);
    save (['bs_sample_' int2str(y)]);
end
clear;
Generate mask for fruit
for z=1:9;
    bs_hsi_mask (['bs_sample_' int2str(z)], 'bs_wd');
end
%% Normalize hyperspectral images of fruit samples
for z=1:9
    hsi_norm (['bs_sample_' int2str(z)], 'bs_wd');
end
for z=1:80
    [R_score, R_loading] = hsi_pca (['Norm_gi_sample_' int2str(z)]);
    gi_R_score{z,1}=R_score;
    gi_R_loading{z,1}=R_loading;
end
```

APPENDIX D  
MATLAB CODE FILES FOR PRINCIPAL COMPONENT ANALYSIS

```
function [R_score, R_loading] = hsi_pca (filename_sample)
% Perform PCA for hyperspectral images
% Example: [R_score, R_loading] = hsi_pca ('Norm_R_RG_CK_W_05');

% Read image data

load (filename_sample);
index_band = [1:1:92];
sample_m = size(R_sample, 1);
sample_n = size(R_sample, 2);
R_sample = double(R_sample(1:1:sample_m, 1:1:sample_n, index_band)); % Kim

% Reshape image data
image_height = size(R_sample, 1);
image_width = size(R_sample, 2);
num_band = size(R_sample, 3);
R_sample = reshape (R_sample, image_height*image_width, num_band);

% Perform PCA
num_pcs = 6;
options_pca = pca('options');
options_pca.algorithm = 'svd';
options_pca.display = 'off';
options_pca.plots = 'none';
model_pca = pca (R_sample, num_pcs, options_pca);
clear R_sample;
R_loading = model_pca.loads{2};
R_score_vector = model_pca.loads {1};
R_score= reshape (R_score_vector, image_height, image_width, num_pcs);
```

APPENDIX E  
MATLAB CODE FILES FOR WAVELELT TRANSFROM

```
%% Weight coefficients for wavelet transform

for i=1:50
cd('C:\Documents and Settings\student\Desktop\Research\wavelet
analysis\bm_image');
band_number=84;
level=4;
n=256;
load(['Norm_bm_sample_' int2str(i)]);
img=R_sample(:, :, band_number);
mn = min(min(img));
    mx = max(max(img));
    range = mx-mn;
    % "g22" is Normalized "Principle Component 2"
    norm_wc_img = (img-mn)/range;
cd('C:\Documents and Settings\student\Desktop\Research\wavelet
analysis\wc_image');
imwrite(norm_wc_img, ['wc_img_' int2str(i) '.png'], 'png');

[LL,HL,LH,HH]=wavelet(['wc_img_' int2str(i)],level, n);
ELLP_bm(:,i)=LL;
EHLP_bm(:,i)=HL;
ELHP_bm(:,i)=LH;
EHHP_bm(:,i)=HH;

end
for i=1:50
cd('C:\Documents and Settings\student\Desktop\Research\wavelet analysis\gi_image');
band_number=84;
level=4;
n=256;
load(['Norm_gi_sample_' int2str(i)]);
img=R_sample(:, :, band_number);
mn = min(min(img));
    mx = max(max(img));
    range = mx-mn;
    % "g22" is Normalized "Principle Component 2"
    norm_wc_img = (img-mn)/range;
cd('C:\Documents and Settings\student\Desktop\Research\wavelet
analysis\wc_image');
imwrite(norm_wc_img, ['wc_img_' int2str(i) '.png'], 'png');

[LL,HL,LH,HH]=wavelet(['wc_img_' int2str(i)],level, n);
```

```

ELLP_gi(:,i)=LL;
EHLP_gi(:,i)=HL;
ELHP_gi(:,i)=LH;
EHHP_gi(:,i)=HH;

end
for i=1:40
cd('C:\Documents and Settings\student\Desktop\Research\wavelet analysis\ir_image');
band_number=84;
level=4;
n=256;
load(['Norm_ir_sample_' int2str(i)]);
img=R_sample(:, :, band_number);
mn = min(min(img));
    mx = max(max(img));
    range = mx-mn;
    % "g22" is Normalized "Principle Component 2"
    norm_wc_img = (img-mn)/range;
    cd('C:\Documents and Settings\student\Desktop\Research\wavelet
analysis\wc_image');
imwrite(norm_wc_img, ['wc_img_' int2str(i) '.png'], 'png');

[LL,HL,LH,HH]=wavelet(['wc_img_' int2str(i)],level, n);
ELLP_ir(:,i)=LL;
EHLP_ir(:,i)=HL;
ELHP_ir(:,i)=LH;
EHHP_ir(:,i)=HH;

end
for i=1:40
cd('C:\Documents and Settings\student\Desktop\Research\wavelet
analysis\mn_image');
band_number=84;
level=4;
n=256;
load(['Norm_mn_sample_' int2str(i)]);
img=R_sample(:, :, band_number);
mn = min(min(img));
    mx = max(max(img));
    range = mx-mn;
    % "g22" is Normalized "Principle Component 2"
    norm_wc_img = (img-mn)/range;
    cd('C:\Documents and Settings\student\Desktop\Research\wavelet
analysis\wc_image');
imwrite(norm_wc_img, ['wc_img_' int2str(i) '.png'], 'png');

```

```

[LL,HL,LH,HH]=wavelet(['wc_img_' int2str(i)],level, n);
ELLP_mn(:,i)=LL;
EHLP_mn(:,i)=HL;
ELHP_mn(:,i)=LH;
EHHP_mn(:,i)=HH;

end

for i=1:40
cd('C:\Documents and Settings\student\Desktop\Research\wavelet
analysis\nm_image');
band_number=84;
level=4;
n=256;
load(['Norm_nm_sample_' int2str(i)]);
img=R_sample(:,band_number);
mn = min(min(img));
mx = max(max(img));
range = mx-mn;
% "g22" is Normalized "Principle Component 2"
norm_wc_img = (img-mn)/range;
cd('C:\Documents and Settings\student\Desktop\Research\wavelet
analysis\wc_image');
imwrite(norm_wc_img, ['wc_img_' int2str(i) '.png'],'png');

[LL,HL,LH,HH]=wavelet(['wc_img_' int2str(i)],level, n);
ELLP_nm(:,i)=LL;
EHLP_nm(:,i)=HL;
ELHP_nm(:,i)=LH;
EHHP_nm(:,i)=HH;

end

for i=1:40
cd('C:\Documents and Settings\student\Desktop\Research\wavelet analysis\ny_image');
band_number=84;
level=4;
n=256;
load(['Norm_ny_sample_' int2str(i)]);
img=R_sample(:,band_number);
mn = min(min(img));
mx = max(max(img));
range = mx-mn;
% "g22" is Normalized "Principle Component 2"
norm_wc_img = (img-mn)/range;
cd('C:\Documents and Settings\student\Desktop\Research\wavelet
analysis\wc_image');

```

```

imwrite(norm_wc_img, ['wc_img_' int2str(i) '.png'],'png');

[LL,HL,LH,HH]=wavelet(['wc_img_' int2str(i)],level, n);
ELLP_ny(:,i)=LL;
EHLP_ny(:,i)=HL;
ELHP_ny(:,i)=LH;
EHHP_ny(:,i)=HH;

end

for i=1:40
cd('C:\Documents and Settings\student\Desktop\Research\wavelet analysis\zc_image');
band_number=84;
level=4;
n=256;
load(['Norm_zc_sample_' int2str(i)]);
img=R_sample(:,band_number);
mn = min(min(img));
mx = max(max(img));
range = mx-mn;
% "g22" is Normalized "Principle Component 2"
norm_wc_img = (img-mn)/range;
cd('C:\Documents and Settings\student\Desktop\Research\wavelet
analysis\wc_image');
imwrite(norm_wc_img, ['wc_img_' int2str(i) '.png'],'png');

[LL,HL,LH,HH]=wavelet(['wc_img_' int2str(i)],level, n);
ELLP_zc(:,i)=LL;
EHLP_zc(:,i)=HL;
ELHP_zc(:,i)=LH;
EHHP_zc(:,i)=HH;

end
for i=0:3
wt_features_bm(1+(3*i),:)=EHLP_bm((i+1),:);
wt_features_bm(2+(3*i),:)=ELHP_bm((i+1),:);
wt_features_bm(3+(3*i),:)=EHHP_bm((i+1),:);
end
wt_features_bm(13,:)=ELLP_bm(1,:);
for i=0:3
wt_features_gi(1+(3*i),:)=EHLP_gi((i+1),:);
wt_features_gi(2+(3*i),:)=ELHP_gi((i+1),:);
wt_features_gi(3+(3*i),:)=EHHP_gi((i+1),:);
end
wt_features_gi(13,:)=ELLP_gi(1,:);
for i=0:3

```

```

wt_features_ir(1+(3*i),:)=EHLP_ir((i+1),:);
wt_features_ir(2+(3*i),:)=ELHP_ir((i+1),:);
wt_features_ir(3+(3*i),:)=EHHP_ir((i+1),:);
end
wt_features_ir(13,:)=ELLP_ir(1,:);
for i=0:3
wt_features_mn(1+(3*i),:)=EHLP_mn((i+1),:);
wt_features_mn(2+(3*i),:)=ELHP_mn((i+1),:);
wt_features_mn(3+(3*i),:)=EHHP_mn((i+1),:);
end
wt_features_mn(13,:)=ELLP_mn(1,:);
for i=0:3
wt_features_nm(1+(3*i),:)=EHLP_nm((i+1),:);
wt_features_nm(2+(3*i),:)=ELHP_nm((i+1),:);
wt_features_nm(3+(3*i),:)=EHHP_nm((i+1),:);
end
wt_features_nm(13,:)=ELLP_nm(1,:);
for i=0:3
wt_features_ny(1+(3*i),:)=EHLP_ny((i+1),:);
wt_features_ny(2+(3*i),:)=ELHP_ny((i+1),:);
wt_features_ny(3+(3*i),:)=EHHP_ny((i+1),:);
end
wt_features_ny(13,:)=ELLP_ny(1,:);
for i=0:3
wt_features_zc(1+(3*i),:)=EHLP_zc((i+1),:);
wt_features_zc(2+(3*i),:)=ELHP_zc((i+1),:);
wt_features_zc(3+(3*i),:)=EHHP_zc((i+1),:);
end
wt_features_zc(13,:)=ELLP_zc(1,:);

```

## LIST OF REFERENCES

- Aleixos, N., J. Blasco, F. Navarron, and E. Molto. 2002. Multispectral inspection of citrus in real-time using machine vision and digital signal processors. *Comput. Electron. Agri.*, Volume 33(2), Page(s): 121-137.
- Arivazhagan, S., Ganesan, L., 2003. Texture segmentation using wavelet transform *Pattern Recognition Letters*, vol. 24, no. 16, pp. 3197-3203.
- Bishop, C.M., 2006. *Pattern Recognition and Machine Learning*. Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA. pp. 242-244.
- Burgers, C., 1998. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2, 121-167.
- Burks, T.F., Shearer, S.A., Payne, F.A., 2000. Classification of weed species using color texture features and discriminant analysis. *Trans. ASAE*, 43(2), 441-448.
- Burks, T.F., Shearer, S.A., Gates, R.S, Donohue, K.D., 2000. Back-propagation neural network design and evaluation for classifying weed species using color image texture. *Trans. ASAE*, 43(4), 1029-1037.
- Burks, T. F., 1997. *Color Image Texture Analysis and Neural Network Classification of Weed Species*. Ph.D. Thesis, University of Kentucky, Lexington, Kentucky.
- Chang, C.I., 1999 Information divergence for hyperspectral image analysis. *Geoscience and Remote Sensing Symposium*, vol. 1, pp. 509-511
- Cheng, X., Chen, Y.R., Tao, Y., Chan, D., Wang, C.Y., 2003. Hyper-spectral Imaging and feature extraction methods in fruit and vegetable defect inspection, 2003 ASAE Annual meeting. pp. 03119.
- Chung, K.R., Brlansky, R.H., 2006. Citrus Disease Exotic to Florida: Huanglongbing (Citrus Greening), Institute of Food and Agricultural Sciences, University of Florida, pp. 210.
- Chung, K.T., Peres, N.A., Timmer, L.W., 2009. Citrus disease exotic to Florida: Black spot, one of a series of the Plant Pathology Department, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida, pp-213.
- Citrus Greening Continues To Spread In Growing Areas [Internet], ScienceDaily LLC (US); [Updated: 2007 Jul 13; Cited: 2008 Oct 24].  
<http://www.sciencedaily.com/releases/2007/07/070711001507.htm>.

- Dave, S., Runtz, K., 1995. Image Processing Methods for Identifying Species of Plants, IEEE CAT. NO. 95CH3581-6/0-7803-2741-1.
- Dennison, P.E., Halligan, K.Q., Roberts, D.A., 2004. A comparison of error metrics and constraints for multiple endmember spectral mixture analysis and spectral mapper. Remote sensing of environment, 93, 359-367.
- Dewdney, M.M., 2010. Citrus Black Spot. Citrus Research and Education Center, University of Florida, Citrus Industry 91 (2), 19-20.
- Donohue, K.D., Huang, L., Burks, T., Forberg, F., Piccoli, C.W., 2001. Tissue classification with generalized spectrum parameters. Ultrasound Med. Biol. 27 (11), 1505-1514.
- Du, Y., Chang, C.I., Ren, H., Chang, C.-C., Jensen, J.O., 2004. New hyperspectral discrimination measure for spectral characterization. Optical engineering, 43 (8), 1777-1788.
- Du, C.-J., Sun, D.-W., 2006. Correlating Image Texture Features Extracted by Five Different Methods with the Tenderness of Cooked Pork Ham: A Feasibility Study. Trans. ASAE, 49(2), 441-448.
- Duda, R.O., Hart, P.E., Stork, D.G., 2001. Pattern Classification, 2nd Edition, A Wiley-Interscience Publication, New York.
- Edwards, J.G., Sweet, C.H., 1986. Citrus blight assessment using a microcomputer: quantifying damage using an apple computer to solve reflectance spectra of entire trees. Florida Scientist 49 (1), 48-53.
- Florida Statistics Site [Internet], USDA's National Agricultural Statistics Service (US); [Update: 2007 June; cited 2008 Nov].  
[http://www.nass.usda.gov/Statistics\\_by\\_State/Florida/Publications/Broc/index.asp](http://www.nass.usda.gov/Statistics_by_State/Florida/Publications/Broc/index.asp)
- Fukagawa, T., Ishii, K., Noguchi, N., Terao, H., 2003. Detecting crop growth by a multi-spectral image sensor. Presented at an ASAE Meeting presentation, pp. 033125.
- Gaffney, J.J., 1973. Reflectance properties of citrus fruits. Trans. ASAE 16(2), 310-314.
- Gonzalez, R.C., Woods, R.E., Digital Image Processing, 3rd Edition, 2008, Pearson Prentice Hall, New Jersey.
- Gottwald, T.R., Graham, J.H., Schibert, T.S., 2002. Citrus Canker: The Pathogen and Its Impact. Online. Plant Health Progress doi:10.1094/PHP-2002-0812-01-RV.

- Gottwald, T.R., Graça, J.V., Bassanezi, R.B., 2007. Citrus Huanglongbing: The Pathogen and Its Impact. Online. Plant Health Progress doi:10.1094/PHP-2007-0906-01-RV.
- Gowen, A.A., O'Donnell, C.P., Cullen, P.J., Downey, G., Frias, G.M., 2007. Hyperspectral imaging – an emerging process analytical tool for food quality and safety control. Trends in Food Science & Technology. 18 (2007) : 590-598.
- Gunn, S.R., [Internet]. Support Vector Machines for Classification and Regression. Technical Report. University of Southampton(US); [Updated 1998 May 10; cited 2008 Oct 12]. <http://users.ecs.soton.ac.uk/srg/publications/pdf/SVM.pdf>
- Haralick, R.M., 1979. Statistical and Structural Approaches to Texture. Processing of the IEEE, 67, 768-804.
- Haralick, R.M., Shanmugam, K., 1974. Combined spectral and spatial processing of ERTS imagery data. J. Remote Sens. Environ. 3, 3-13.
- Haralick, R.M., Shanmugam, K., Dinstein., 1973. Texture Features for Image classification, IEEE Trans. on systems, Man, and Cybernetics. 3, 610:621
- Hodge, A., Philippakos, E., Mulkey, D., Spreen, T., Muraro, R., 2001. Economic impact of Florida's citrus industry, Extension Digital Information Source (EDIS) FE307. Gainesville, Fla.: University of Florida, Department of Food and Resource Economics, 1999-2000.
- Ionescu, R., Liobet, E., 2002. Wavelet transform-based fast feature extraction from temperature modulated semiconductor gas sensors. Sens. Actuators B 81 (2002), pp. 289–295.
- Jacobs, J.A., 1994. Cooperative in the U.S. Citrus Industry, Agricultural Economist, U.S. Department of Agriculture, Rural Business and Cooperative Development Service, RBCDS Research Report 137, December, 1994.
- Jiang, L., Zhu, B, Rao, X., Berney, G., Tao, Y., 2007. Black walnut shell and Meat Discrimination using Hyper-spectral Fluorescence Imaging, 073089.
- Jimenez, A.R., R Ceres, Pons, J.L., 2000. A survey of computer vision methods for locating fruit on trees. Trans. ASAE 43(6), 1911-1920.
- Kataoka, T., Kaneko, T., Okamoto, H., Hata, S., 2003. Crop growth estimation system using machine vision. Presented at the 2003 IEEE/ASME International conference on advanced Intelligent mechatronics.AIM 2003. 2, b1079- b1083.

- Kawamura, S., Tsukahara, M., Natsuga, M., Itoh, K., 2003. On-line near infrared spectroscopic sensing technique for assessing milk quality during milking. 2003 ASAE Annual international meeting. 033026.
- Kim, M. S., Chen, Y. R., Mehl, P. M., 2001. Hyperspectral reflectance and fluorescence imaging system for food quality and safety. *Trans. ASAE*, 44(3), 721-729.
- Kim, M.S., Lefcourt, A.M., Chao, K., Chen, Y.R., Kim, I., Chan, D.E., 2002. Multispectral detection of fecal contamination on apple based on hyper-spectral imagery. *Trans. ASAE*, 45(6), 2027-2037.
- Lee, K. J., Kang, S., Kim, M.S., Noh, S. H., 2005. Hyper-spectral imaging for detection defect on apple. 2005 ASAE Annual international meeting, 053075.
- Lee, W.S., Slaughter, D., 1998. Plant recognition using hardware-based neural network. *Trans. ASAE*. pp. 98030.
- Mehl, P.M., Chao, K., Kim, M., Chen, Y.R., 2002. Detection of defects on selected apple cultivars using hyper and multi image analysis. *Applied Engineering in Agriculture*. 18(2), 219-226.
- Miller, W.M. Drouillard, G.P., 2001. Multiple feature analysis for machine vision grading of Florida citrus. *Applied Eng. Agri*. 17(5), 627-633.
- Mishra, A., Ehsani, F., Albrigo, G., Lee, W.S., 2007. Spectral characteristics of citrus greening (Huanglongbing). ASABE Annual international meeting, 073056.
- Moshou, D., Virindts, E., Ketelaere, B.D., Baerdemaeker, J.D., Ramon, H., 2001. A neural network based plant classifier. *Computers and electronics in agriculture*. 31(1), 5-16.
- Muhammed, H., 2005. Hyperspectral Crop Reflectance Data for characterizing and estimating Fungal Disease Severity in Wheat. *Biosystems Engineering*. (2005) 91(1) : 9-20.
- Ng, H.F., Wilcke, W.F., Morey, R.V., Lang, J.P., 1998. Machine evaluation of corn kernel mechanical and mold damage. *Trans. ASAE*, 41(2), 415-420.
- Nilsson, H.-E., 1995. Remote sensing and image analysis in plant pathology. *Canadian journal of plant pathology*. 17, 154 -166
- Ohta, Y. 1985. Knowledge-Based Interpretation of Outdoor Natural Color Scenes. Pitman Publishing Inc., Marshfield, M.A.

- Park, B., Yoon, S.C., Lawrence, K.C., Windham, W.R., 2007. Fisher Linear Discriminant Analysis for Improving Fecal Detection Accuracy with Hyperspectral Images. *Trans. ASAE* 50(6), 2257-2283.
- Park, B., Windham, W.R., Lawrence, K.C., Smith, D.P., 2007. Contaminant classification of poultry hyperspectral imagery using a spectral angle mapper algorithm. *Biosystems engineering*, 96(3), 323-333.
- The Plant Pathology / Plant Disease Online [Internet], The American Phytopathological Society (US), [Update: 2008; Cited: 2008 Oct]. <http://www.apsnet.org/>
- Polek, M., Vidalakis, G., Godfrey, K., 2007. Citrus Bacterial Canker Disease and Huanglongbing(Citrus Greening), University of California, Agricultural and Natural Resources, Publication 8218, ISBN-13: 978-1-60107-439-3, ISBN-10: 1-60107-439-5.
- Powell, N.B., Spencer, S.R., Boyette, M.D., 2005. Machine Vision for Autonomous Machine Guidance, 2005 ASAE Annual international meeting 053091.
- Pydipati, R., 2004. Automatic Disease Detection in Citrus Trees. MS Thesis, University of Florida, Gainesville, FL.
- Pydipati, R., Burks, T.F., Lee, W.S., 2005. Statistical and Neural Network Classifiers for Citrus Disease Detection Using Machine Vision. *Trans. ASAE*. 48(5), 2007-2014.
- Pydipati, R., Burks, T.F., Lee, W.S., 2006. Identification of citrus disease using color texture features and discriminant analysis. *Computers and electronics in agriculture*. 52, 49-59.
- Qin, J., Burks, T.F., Kim, M.S., Chao, K., Ritenour, M.A., 2008. Citrus canker detection using hyperspectral reflectance imaging and PCA-based image classification method. *Sens. Instrum. Food Qual. Saf.*, doi: 10.1007/s11694-008-9043-3.
- Qin, J., Burks, T.F., Ritenour, M.A., Bonn, W.G., 2009, Detection of citrus canker using hyperspectral reflectance imaging with spectral information divergence, *Journal of Food Engineering*, 93:183-191.
- Rao, C. R., 1973. Linear statistical inference and its application. John Wiley and Sons, New York, NY, USA.
- Requathan, M., Lee, W.S., 2005. Citrus Fruit Identification and Size Determination Using Machine Vision and Ultrasonic Sensors. 2005 ASAE Annual international meeting 053017.
- Ruangwong, O., Akrapisan, A., 2006. Detection of Candidatus Liberibacter asiaticus causing Citrus Huanglongbing disease. *Journal of Agricultural Technology* 2(1):111-120.

- Ruiz, L.A., Fdez-Sarria, A., Recio, J.A., 2002. Texture feature extraction for classification of remote sensing data using wavelet decomposition: A comparative study. IAPRS, vol. 35.
- SAS Institute. Inc., 1985. SAS Introductory Guild (3rd ed.), Cary, North Carolina. USA
- SAS Institute. Inc., 2004. SAS/STAT 9.1 User's Guide. SAS Institute Inc., Cary, North Carolina, USA.
- Scherz, C.E., Brown, G.K., 1968. Basic considerations in mechanizing citrus harvest. Trans. ASABE. 343-346.
- Sepulcre-Canto, G., Zarco-Tejada, P.J., Jimenez-Munoz, J.C., Sobrino, J.A., Soriano, M.A., Fereres, E., Vega, V., Pastor, M., 2007. Monitoring yield and fruit quality parameters in open-canopy tree crops under water stress. Implications for ASTER. Remote Sensing of Environment. 107 , 455 -470.
- Sevier, B.J. and Lee, W.S., 2003. Adoption trends and attitudes towards precision agriculture in florida citrus: preliminary results from a citrus producer survey. Trans. ASAE. pp. 031100.
- Shearer, S.A., Holmes, R.G., 1990. Plant identification using co-occurrence matrices. Trans. ASAE. 33, 2037-2044.
- Shearer, S.A., 1986, Plant identification using color co-occurrence matrices derived from digitized images. PhD Thesis, Ohio State University, Columbus, OH.
- Shearer, S.A., Holmes, R.G., 1990. Plant identification using co-occurrence matrices. Trans. ASAE. 33, 2037-2044.
- Shimada, H., Shima, E., Tanaka, K., Nagayoshi, T., 2008. Use of personal remote sensing system in grazing land. 2008 ASABE Annual international meeting, 084473.
- Shippert, P., 2003. Introduction to hyperspectral image analysis. Online Journal of space communication, Issue No.3.
- Sun, D.W., 2008. Computer Vision Technology for food quality evaluation. Elsevier Inc. Burlington, MA 01803, USA.
- Tang, L., Tian, L.F., Stward, B.L., Reid, J.F., 1999. Texure based weed classification using gabor wavelets and neural networks for real time selective herbicide application. ASAE/CSAE-SCGR Annual international metting, Toronto, Canada. pp. 993036.
- Tuceryan, M., Jain, A.K., 1998. Texture Analysis. The handbook of Pattern Recognition and Computer Vision (2nd Edition), by C. H. Chen, L. F. Pau, P. S. P. Wang (eds.), 207-248.

- Welling, M., [Internet]. Fisher Linear Discriminant Analysis. Department of Computer Science, University of Toronto, 10 King's College Road, Toronto, M5S 3G5 Canada; [cited 2008 Oct 12].  
[http://www.ics.uci.edu/~welling/classnotes/papers\\_class/Fisher-LDA.pdf](http://www.ics.uci.edu/~welling/classnotes/papers_class/Fisher-LDA.pdf)
- Wyszecki, G., Stiles, W.S., 1992. Color science: concepts and methods. In: Quantitative Data and Formulae, second ed. John Wiley & Sons, New York, pp. 117-137.
- Yang, C., Everitt, J.H., Bradford, J.M., 2006. Use of spectral angle mapper (SAM) and hyperspectral imagery for yield estimation. 2006 ASAE Annual meeting presentation:061168.
- Yang, T., Chen, Y., Cheng, X., 2001. Infrared imaging and wavelet-based segmentation method for apple defect inspection. 2001 ASAE Annual international meeting, pp. 01-3109.
- Ye, X., Skai, K., Okamoto, H., Garciano, L.O., 2008. Aground-based hyperspectral imaging system for characterizing vegetation spectral features. Computer and electronics in agriculture, 63, 13-21.
- Zarco-Tejada, P.J., Berjon, A., Lopez-Lozano, R., Miler, J.R., Martin, P., Cachorro, V., Gonzalez, M.R., de Frutos, A., 2005. Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy. Remote sensing of environment. 99, 271-287.
- Zhang, H., Paliwal, J., Jayas, D.S., White, N.D.G., 2007. Classification of fungal Infected Wheat Kernels Using Near-Infrared Reflectance Hyperspectral Imaging and Support Vector Machine, Trans. ASABE, 50(5), 1779-1785.
- Zhang, Y., He, X.Y., Han, J.H., 2005. Texture feature-based image classification using wavelet package transform. ICIC 2005 part I, LNCS 3644, 165-173.

## BIOGRAPHICAL SKETCH

Dae gwan Kim was born in 1979 in Tong Young, South Korea. He graduated with a mechanical engineering degree in March 2006 from the Dong-A University, Busan, South Korea. He started his graduate studies in the Department of Mechanical and Aerospace Engineering at the University of Florida. He transferred to the Department of Agricultural and Biological Engineering in August 2007. He was a member of the Agricultural Robotics and Mechatronics Group (ARMg) in the Department of Agricultural and Biological Engineering, where he worked as a research assistant under advisor Dr. Thomas F. Burks.