Detection of Citrus Greening Using Microscopic Imaging

Dae G. Kim[*], Thomas F. Burks[a], Arnold W. Schumann[b], Mongi Zekri[c], Xuhui Zhao[a], Jianwei Qin[a]

[a] Department of Agricultural and Biological Engineering, 233 Frazier Rogers Hall, P.O. Box 110570, University of Florida, Gainesville, FL 32611-0570, USA.
[b] Department of Soil & Water Science, Citrus Research & Education Center, 700 Experiment Station Road, Lake Alfred, FL, 33850,USA
[c] Hendry County Extension Office, PO Box 68, LaBelle, FL 33975-0068, USA
Corresponding author: dgkim030@ufl.edu

ABSTRACT

Citrus greening reduces fruit production and quality and will likely result in rapid tree decline and death. Because citrus greening symptoms are usually observed on the leaf surface, detection of citrus greening leaf symptoms can significantly aid in scouting for infected trees and managing the disease, thus reducing its spread and minimizing losses for citrus growers. This article presents the microscopic image analysis using color co-occurrence method to differentiate citrus leaves with eight conditions: greening blotchy mottle, green islands, iron deficiency, manganese deficiency, zinc deficiency, young flush leaves and normal mature leaves. Thirty-nine statistical features were extracted from transformed hue (H), saturation (S), and intensity (I) images using the color co-occurrence method for each leaf sample. The number of extracted texture features was reduced by a stepwise discriminant analysis. A discriminant function based on a measure of the generalized squared distance was used for classification. Three classification models were performed using (1) all leaf conditions, (2) all conditions except young flush leaves and (3) all conditions except young flush leaves and blotchy mottle. The three classification models obtained accuracies of 86.67 %, 95.60 % and 97.33 %, respectively. The overall performance was demonstrated in a confusion matrix. The model HSI_14, which used all conditions except young flush and blotchy mottle, resulted in the best accuracy for positive (96.67 %) and negative (97.5 %) symptoms.

Keywords: Citrus leaf, disease detection, machine vision, color co-occurrence method, texture features, discriminant analysis

1. INTRODUCTION

Huanglongbing (HLB), commonly known as citrus greening, is one of the most dangerous diseases that affect citrus production, and 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. 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 one pixel offset with four directions (0°, 45°, 90° and 135°). 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 matrix (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:

1) Transformations of an RGB color representation of an image to an equivalent HSI color representation.
2) Generation of color co-occurrence matrices from the HSI pixels
3) 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 (Burks et al., 2000). 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 was then 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:

1) 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).
2) Determine image texture features based on the color co-occurrence method (CCM).
3) Create a set of reduced feature data models through a stepwise elimination process and classify different citrus leaf conditions.
4) Compare the classification accuracies.

2. MATERIALS AND METHODS

2.1 Citrus Leaf Samples

The leaf samples used in this research were collected from two orange groves near Immokolee in southwest Florida, during summer and fall of 2008. Eight different classes of citrus leaves were selected for this study, and were graded manually into classes by an expert extension agent. The leaf sample conditions were blotchy mottle, green islands, iron deficiency, manganese deficiency, zinc deficiency, young flush and normal mature. Images of leaf samples are shown in figure 1.

| Iron Deficiency | Zinc Deficiency | Manganese Deficiency | Greening Blotch |

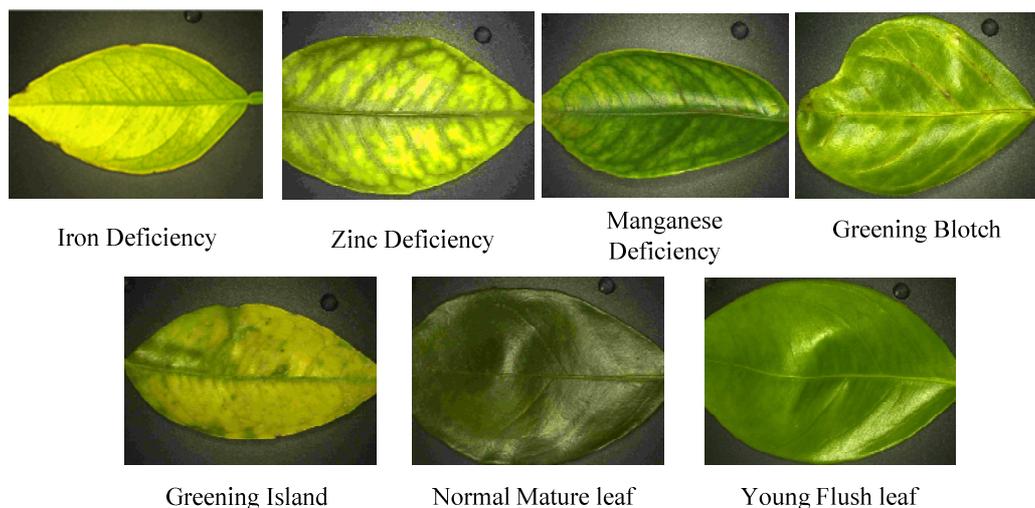| Greening Island | Normal Mature leaf | Young Flush leaf |

Figure 1. Citrus Leaf Conditions

Each nutritional deficiency of citrus has distinctive feature on the leaf surface. The leaf with iron deficiency has the dark green network of veins within the yellow leaf blade. The leaf with manganese deficiency has the symmetry of the yellowing across the mid-vein and the dark triangle at the leaf base. The leaf with zinc deficiency has the fairy symmetrical yellowing across the mid-vein (Polek *et al.*, 2005).

The citrus greening leaf symptoms are similar to other cultural conditions and diseases, but it has difference with others. A blotchy mottle pattern was most typical with light greening and dark green patches, and no symmetry. Green islands has non-symmetrical pattern on opposite sides of mid-vein. Figure 2 shows various canker condition images used in this study.

Figure 2. 15 images of blotch module conditions

Kim D.G., Burks T.F., Schumann A.W., Zekri M., Zhao X., Qin J. "Detection of Citrus Greening Using Microscopic Imaging" Agricultural Engineering International: the CIGR Ejournal. Manuscript 1194. Vol. XI. June, 2009.

The visual symptom observed varied between leaf samples. Leaf samples were from trees with petioles intact and then sealed in Ziploc® bags to maintain the moisture level of the leaves. Sixty samples of each of the eight classes of leaves were collected. The samples were brought to a laboratory. The leaf samples were then sealed in new bags with appropriate labels and put in environmental control chambers maintained at 4 ºC. They were removed from cold storage about 2 hours before imaging to allow them to reach room temperature. The leaf samples were then taken to an imaging station where images of the upper side of the leaf were acquired.

2.2 Color Image Acquisition

A digital microscope system (VHX-600K, Keyence, Japan) was used for acquiring color (RGB) images from citrus leaf samples, as shown in figure 3. The imaging system consisted of a halogen lamp (12V, 100W), a zoom lens (OP-51479, Keyence, Japan), a 2.11-million-pixel CCD image sensor (1/1.8-inch), a 15-inch Color LCD monitor (TFT, UXGA), and a computer installed with an image capture function and a hard disk drive unit (image format: JPEG and TIFF, Storage capacity: 700 MB). 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 capture 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 (1200×1600, 8 bit/channel).

Figure 3. Digital microscope system for acquiring color images from citrus leaf samples

2.3 Texture Analysis

2.3.1 Color Co-occurrence Matrix Methodology

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 consisted of three primary mathematical processes. First, the RGB images of leaves were converted to a hue, saturation and intensity (HSI) color space representation. Intensity was calculated using the mean value of the three RGB values. The hue and saturation values were determined using a geometrical transformation of the International Commission on Illumination's (CIE) chromaticity diagram (Ohta, 1985). In this process, the CIE chromaticity diagram represented 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 applied. The gray level co-occurrence methodology is a statistical way 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, a matrix $P(i,j)$ counts the number of times a pixel with grey-level i occurs at position p from a pixel with grey-level j. Normalization of the matrix P by the total number of pixels calculates values between 0 and 1, resulting in a gray-level co-occurrence matrix.
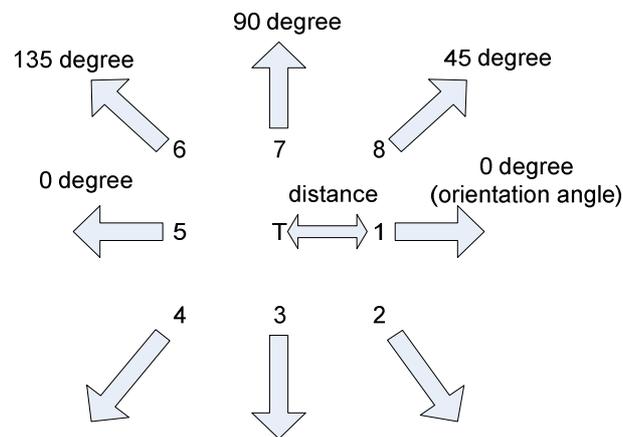

Figure 4. Nearest neighbor mask for calculating
spatial gray-level dependence matrices (SGDM's)

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 θ from location *(x,y)*. The nearest neighbor mask is shown in figure 4, where the reference pixel *(x,y)* is shown with 'T'. All eight neighbors shown are one pixel

distance from the reference pixel 'T' and are numbered in a clockwise direction from one to eight. The neighbors at positions one and five are both considered to be at an orientation angle equal to zero degree, while positions eight and four are considered to be at an angle of 45 degrees. The equation for normalizing the co-occurrence matrix is given in Eq.1, where $P(i,j,1,0)$ is the intensity co-occurrence matrix.

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

(1)

An example image matrix $I(x,y)$, with a gray scale ranges from zero to three is shown in Eq.2.

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

(2)

The hue, saturation and intensity CCM matrices were 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, and 11 texture feature equations. Donohue *et al.* (1985) added image contrast and modus texture features to those used by Ohta (1985), for a total of thirteen features when classifying cancer tissue. The same equations were 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 were identified by a coded variable name where the first letter represented whether it was a hue (H), saturation (S) or intensity (I) feature and the number following represented one of the thirteen texture features described in Shearer (1990). Intensity texture feature equations are presented in table 1.

Table 1. Intensity texture features

Feature	Description	Equation
I1	Uniformity (2nd Moment)	$I_1 = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}[p(i,j)]^2$
I2	Mean Intensity	$I_2 = \sum_{i=0}^{N_g-1}ip_x(i)$
I3	Variance	$I_3 = \sum_{i=0}^{N_g-1}(i-I_2)p_x(i)$
I4	Correlation	$I_4 = \dfrac{\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}ijp(i,j)-I_2^2}{I_3}$
I5	Product Moment	$I_5 = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}(i-I_2)(j-I_2)p(i,j)$

I6	Inverse Difference	$I_6 = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}\dfrac{p(i,j)}{1+(i-j)^2}$
I7	Entropy	$I_7 = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}p(i,j)\ln p(i,j)$
I8	Sum Entropy	$I_8 = \sum_{k=0}^{2(N_g-1)}p_{x+y}(k)\ln p_{x+y}(k)$
I9	Difference Entropy	$I_9 = \sum_{k=0}^{N_g-1}p_{x-y}(k)\ln p_{x-y}(k)$
I10	Information Correlation 1	$I_{10} = \dfrac{I_7 - HXY1}{HX}$
I11	Information Correlation 2	$I_{11} = \left[1 - e^{-2(HXY2-I_7)}\right]^{1/2}$
	HX	$HX = -\sum_{i=0}^{N_g-1}p_x(i)\ln p_x(i)$
	HXY1	$HXY1 = -\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}p(i,j)\ln\left[p_x(i)p_x(j)\right]$
	HXY2	$HXY2 = -\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}p_x(i)p_x(j)\ln\left[p_x(i)p_x(j)\right]$
I12	Contrast	$I_{12} = \sum_{|i-j|=0}^{N_g-1}(i-j)^2\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}p(i,j)$
I13	Modus	$I_{13} = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}\max\left[p(i,j)\right]$

As an example, the feature *(I7)* is a measure of the entropy in the intensity CCM matrix, which represents the amount of order in an image. 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 in the image would therefore have more uncertainty and thus a higher entropy measure.

2.3.2 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 minimized negative time dependant variability, and reduced potential for data selection bias between the training and test datasets.

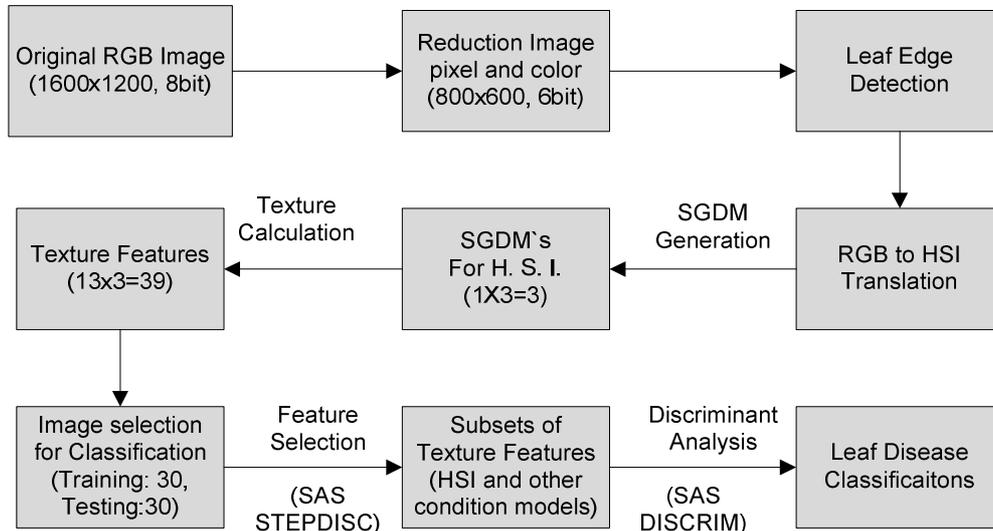|		
Original RGB Image (1600x1200, 8bit)	→ Reduction Image pixel and color (800x600, 6bit)	→ Leaf Edge Detection
Texture Features (13x3=39)	← Texture Calculation ← SGDM's For H. S. I. (1X3=3)	← SGDM Generation ← RGB to HSI Translation
Image selection for Classification (Training: 30, Testing:30)	→ Feature Selection (SAS STEPDISC) → Subsets of Texture Features (HSI and other condition models)	→ Discriminant Analysis (SAS DISCRIM) → Leaf Disease Classificaitons

Figure 5. Procedures for color image analysis

A detailed illustration of the image acquisition and classification process is given in Figure 5. Algorithms for image segmentation and texture feature generation were developed in MATLAB R2007b (MathWorks. Inc. USA). 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 size was reduced from 1600×1200 pixels to 800×600 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 6 exhibits a detailed edge detection process.

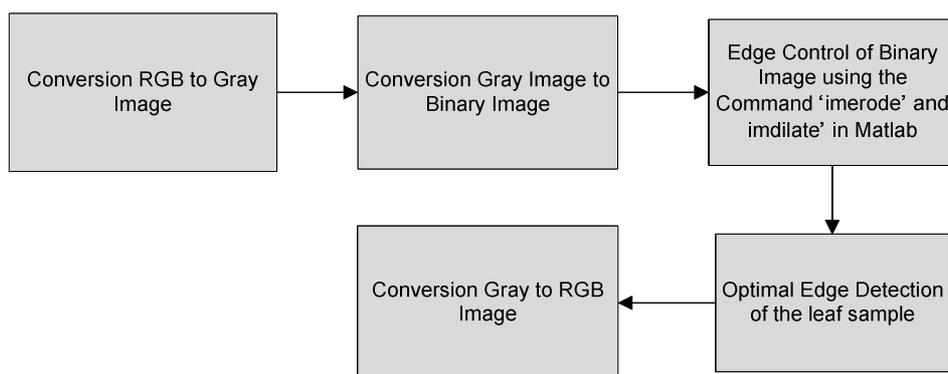|		
Conversion RGB to Gray Image	→ Conversion Gray Image to Binary Image	→ Edge Control of Binary Image using the Command 'imerode' and imdilate' in Matlab
	Conversion Gray to RGB Image	← Optimal Edge Detection of the leaf sample

Figure 6. Procedures for leaf edge detection

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 pixel offset.

2.3.3 Statistical Analysis

Once the texture statistics were generated for each image, statistical analyses were conducted using procedure STEPDISC (SAS, 1985) 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. PROC STEPDISC was used to reduce the number of texture features by a stepwise selection process. 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 Wilk's 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 2. Model HSI_18 consisted of all conditions, HSI_15 model consisted of all condition except normal young and the HSI_14 model consisted of all conditions except blotchy mottle and normal young leaves.

Table 2. Texture feature models selected by stepwise discriminant analysis for fall season

Condition	Model[1]	Color feature[2]	Texture feature set[3]
All disease condition	HSI_18	H, S, I	$S_4, I_2, H_7, S_{13}, H_2, H_9, S_5, I_7, S_7, I_9, S_8, I_1, I_{10}, H_4, I_6, S_6, H_8, I_{13}$
All conditions except young flush	HSI_15	H, S, I	$S_5, I_2, H_7, H_2, S_6, S_4, H_9, S_8, I_6, S_{13}, H_4, I_4, I_{13}, S_7, I_7$
All conditions except blotch mottle and young flush	HSI_14	H, S, I	$S_5, I_2, H_7, H_2, S_4, H_9, S_{13}, S_7, I_7, I_1, I_9, S_8, I_{10}, I_6$

[1] Classification model designation based on color features in the 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] Selected texture features, given in order of discriminant power.

3. RESULTS AND DISCUSSION

3.1 Classification of Citrus Disease Conditions

The texture feature dataset was generated by containing 39 texture features for each image. The dataset consisted of 420 elements, representing 60 samples from each of the seven classes of leaves. To compare classification accuracies under various disease conditions, three models were created which are shown in table 2. These models represent the compilation of three different leaf conditions sets, which isolate leaf conditions that are difficult to discriminate. Table 3 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 2 and 3 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.

Table 3. Texture feature models to all conditions except young flush for fall season

Classification model[1]	Color feature[2]	Texture feature set[3]
HSI_15	H, S, I	$S_5, I_2, H_7, H_2, S_6, S_4, H_9, S_8, I_6, S_{13}, H_4, I_4, I_{13}, S_7, I_7$
HS_10	H, S	$S_5, H_7, H_5, H_{12}, S_4, S_7, H_8, S_8, H_3, S_{11}$
I_8	I	$I_2, I_8, I_9, I_6, I_5, I_7, I_{10}, I_1$
HSI_39	H, S, I	All 39 texture features ($H_1 - H_{13}, S_1 - S_{13}, I_1 - I_{13}$)

[1] Classification model designation based on 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] Selected texture features, given in order of discriminant power.

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 3, 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. The results shown in table 4 are the classification summaries for models HSI_18, HSI_15 and HSI_14 given in table 2. 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 2. HSI_15 and HSI_14 models had high classification accuracies (95.60 % and 97.33 %).

Table 4. Classification summary in percent correct for selected models

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	95.60	97.33

Kim D.G., Burks T.F., Schumann A.W., Zekri M., Zhao X., Qin J. "Detection of Citrus Greening Using Microscopic Imaging" Agricultural Engineering International: the CIGR Ejournal. Manuscript 1194. Vol. XI. June, 2009.

Based on the results shown in table 5, the classification model using only intensity texture features presented the worst performance at 81.11 % for the I_8 model. When compared with other models, HS_10 model had 87.78 %, HSI_15 model had 95.60 % and HSI_39 model had 95.60 %. Therefore, other models provided better performance than the model that used only intensity texture features.

Table 5. Classification summary in percent correct for selected models

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 (%)	95.60	87.78	81.11	95.60

 In table 5, the highest overall performance was 95.60 % for HSI_15 and 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 6, most images were correctly classified into the appropriate category; however, young flush leaves had a very low classification at 23 %.

Table 6. Classification result for HSI_18 model

Actual leaf condition	Classified leaf condition							
	Blotchy mottle	Islands	Zinc deficiency	Iron deficiency	MN deficiency	Normal	Young flush	Accuracy (%)
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

The negative influence of young flush leaves was further demonstrated in the results from table 4 where the classification accuracy was 86.67 % while other leaf condition models that exclude

Kim D.G., Burks T.F., Schumann A.W., Zekri M., Zhao X., Qin J. "Detection of Citrus Greening Using Microscopic Imaging" Agricultural Engineering International: the CIGR Ejournal. Manuscript 1194. Vol. XI. June, 2009.

young flush leaves had accuracy above 95 %. Table 7 shows improved classification accuracy of 95.56 % and thus proved that young flush leaves affected overall performance result.

Table 7. Classification result for HSI_15 model

Actual leaf condition	Classified leaf condition						
	Blotchy mottle	Islands	Zinc deficiency	Iron deficiency	MN deficiency	Normal	Accuracy (%)
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 8 demonstrates that HSI_14 model obtained the best accuracy (97.33 %) in three leaf condition models (table 2). 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.

Table 8. Classification result for HSI_14 model

Actual leaf condition	Classified leaf condition					
	Islands	Zinc deficiency	Iron deficiency	MN deficiency	Normal	Accuracy (%)
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	0	0	0	0	30	100.00
Total	29	34	27	30	30	97.33

3.2 The Confusion Matrix for Greening Positive vs. Negative

Kim D.G., Burks T.F., Schumann A.W., Zekri M., Zhao X., Qin J. "Detection of Citrus Greening Using Microscopic Imaging" Agricultural Engineering International: the CIGR Ejournal. Manuscript 1194. Vol. XI. June, 2009.

The classification results for the confusion matrix obtained under the positive vs. negative greening model are shown in Tables 9, 10 and 11. In general, the HSI_18 model had the lowest classification accuracy among all symptom models. However, in the confusion matrix shown in Table 9, the discrimination of citrus greening symptoms had high success rate (96.7 %).

Table 9. Confusion matrix in percent correct for HSI_18 model

Actual value		Positive for greening (Blotch mottle, Islands)	Negative for greening (Young flush, Normal, MN, IR, ZN)	Total
	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 10. Confusion matrix in percent correct for HSI_15 model

Actual value		Positive for greening (Blotch mottle, Islands)	Negative for greening (Normal, MN, IR, ZN)	Total
	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 11. Confusion matrix in percent correct for HSI_14 model

Actual value		Positive for greening (Islands)	Negative for greening (Normal, MN, IR, ZN)	Total
	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 %)

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 had good classification accuracies between positive (91.67 %) and negative (97.50 %) as shown in table 10. 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 11, 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 than/with/etc. 97.3 %. When evaluating each model, it is likely that the similarity between young flush leaves and other conditions affected the detection accuracy of citrus greening disease.

3.2 Stability Test of the Greening Classification Model

Kim D.G., Burks T.F., Schumann A.W., Zekri M., Zhao X., Qin J. "Detection of Citrus Greening Using Microscopic Imaging" Agricultural Engineering International: the CIGR Ejournal. Manuscript 1194. Vol. XI. June, 2009.

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.

Table 12. 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

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 12 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.

4. CONCLUSIONS

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 investigated. 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 9 and 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. The 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 may explore the utility of these algorithms in outdoor conditions.

5. ACKNOWLDEGEMENTS

6. REFERENCES

Burks, T.F., S.A. Shearer and F.A. Payne. 2000. Classification of weed species using color texture features and discriminant analysis. *Transactions of the ASAE*, 43(2), 441-448

Donohue, K.D., L. Huang, T. Burks, F. Forberg and C.W. Piccoli. 2001.Tissue classification with generalized spectrum parameters. *Ultrasound in Medicine & Biology.*, 27(11), 1505-1514.

Haralick, R.M., K. Shanmugam and Its'Hak. Dinstein. 1973. Texture features for image classification. *Systems, Man, and Cybernetics. IEEE Transactions on*, 3, 610-621.

Haralick, R.M. and K. Shanmugam. 1974. Combined spectral and spatial processing of ERTS imagery data. *Journal of Remote Sensing if Environment*, 3, 3-13.

Polek, M., G. Vidalakis and K. Godfrey. 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.

Ohta, Y. 1985. *Knowledge-based interpretation of outdoor natural color scenes*. Pitman Publishing Inc., Marshfield, M.A.

Ruangwong, O. A. Akrapisan. 2006. Detection of *Candidatus liberibacter asiaticus* causing citrus Huanglongbing disease. *Journal of Agricultural Technology*, 2(1), 111-120.

Pydipati, R., T.F. Burks and W.S. Lee. 2006. Identification of citrus disease using color texture features and discriminant analysis. *Computers and Electronics in Agriculture*, 52, 49-59.

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. and R.G. Holmes. 1990. Plant identification using co-occurrence matrices. *American Society of Agricultural Engineers Transactions*, 33, 2037-2044.

SAS Institute. Inc. 1985. SAS Introductory Guild (3rd ed.), Cary, North Carolina. USA.

Tang, L., L.F. Tian, B.L. Stward and J.F. Reid. 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.

The plant Pathology / Plant Disease Online [Internet], The American Phytopathological Society (US), [Update: 2008; Cited: 2008 Oct]. Available from: http://www.apsnet.org/media/.

Gottwald, T.R., J.V. da Graça and R.B. Bassanezi**.** 2007. Citrus Huanglongbing: The pathogen and its impact. *Plant Health Progress*, doi:10.1094/PHP-2007-0906-01-RV.

Wyszecki, G. and W.S. Stiles. 1992. Color science: concepts and methods. In: *Quantitative Data and Formulae*, 2nd ed. John Wiley & Sons, New York, pp. 117-137.