2950 Niles Road, St. Joseph, MI 49085-9659, USA
269.429.0300 fax 269.429.3852 hq@asabe.org www.asabe.org

An ASABE Meeting Presentation

Paper Number: 1592802

Airborne hyperspectral imaging based citrus greening disease detection using different dimension reduction methods

Han Li, Ph.D. Candidate

China Agricultural University, Qinghua East Rd, 17#, Haidian, Beijing, P.R. China, 100083,

Agricultural and Biological Engineering, University of Florida, Gainesville, FL 32611

email: emmahhanhan@gmail.com

Written for presentation at the

2013 ASABE Annual International Meeting

Sponsored by ASABE

Kansas City, Missouri

July 21 – 24, 2013

Abstract. *Hyperspectral (HS) imaging has become an efficient tool for agriculture applications. While HS image can provide large amounts of information, it has high redundancy because the bands of HS image are highly correlated. To solve citrus greening disease (Huanglongbing, or HLB) detection problem using airborne HS image, an efficient dimension reduction method needs to be applied, to improve the accuracy of HLB detection and simplify the image acquisition step. There are many dimension reduction methods have been proposed. The objective of this study was to compare the classification performance when using different bands selected using different dimension reduction methods, utilizing the HS airborne images of citrus groves in Florida in 2011. In this study four dimension reduction methods including principal component analysis (PCA), maximum noise fraction (MNF) transformation, forward feature selection algorithm (FFSA) and Kullback-Leibler divergence (KLD) based method, were applied on the obtained HS image. While PCA and MNF are feature extraction methods, which means a new and reduced dataset representing the transformed initial information will be obtained by using them, FFSA and KLD based methods are feature selection, or band selection methods, which means a subset of bands from the original information will be selected. To analyze the band reduction results, both pixel based classification and tree based classification were applied on the 2011 HS data, and the results obtained from these four methods were compared. After applying K-nearest neighbor (KNN) classification on the pixel bands chosen by these four methods, they all showed a 100% accuracy in a calibration dataset for both HLB detection and healthy sample detection. KLD showed the highest HLB detection accuracy of 63.3%, while PCA and KLD based method showed the highest healthy detection accuracy of 74% using a validation dataset. After applying Mahalanobis distance (MahaDist) classification on the transformed image data, MNF and KLD gave the same tree based HLB detection accuracy, which was 93.3%.*

Keywords. *Dimension reduction, FFSA, KLD, KNN, MNF, PCA.*

Introduction

Citrus utilized production for the 2011-2012 season totaled 11.7 million tons throughout the U.S., while Florida accounted for 65 percent of total United States citrus production, according to the citrus statistics by National Agricultural Statistics Services (USDA, 2012). The whole citrus industry has about $9 billion economic impact in Florida where nearly 569,000 acres of citrus groves exist. Citrus greening, also known as Huanglongbing (HLB), caused by Asian citrus psyllids, is a catastrophic disease which has no cure reported yet. It is a disease which has no cure reported yet. Once the tree is infected, it tends to die within 3-5 years. The infection can cause substantial economic losses to the citrus industry by shortening the life span of infected trees and threaten the sustainability of citrus planting in Florida.

Thirty seven counties with over 4012 sections were infected as of August, 2011, (FDACS, DPI, 2011). This deadly disease was recently found in Texas, USA in January 2012 (Texas Department of Agriculture and the USDA, 2012) and also in California in March, 2012 (California Department of Food and Agriculture and the USDA, 2012). To efficiently manage this disease and reduce risks of the disease being spread, timely and location-specific detection and monitoring of the infected citrus trees are required.

Many studies on hyperspectral (HS) image processing have been conducted in recent years, since satellite and aircraft remote images are becoming easier to acquire. Many researchers have noticed the usefulness of HS image data for disease detection of various crops or citrus fruit, since it provides an easy and convenient method for capturing the spectral characteristics over a large area and can be used to detect abnormalities. Previous attempts at HS image classification have been successful in detecting citrus canker, bacterial leaf blight, and diseases in lettuce plants, etc. In this study, airborne HS image were utilized to implement the classification of HLB infected and healthy citrus trees.

The mixture tuned match filtering (MTMF), the spectral angle mapper (SAM) and linear spectral unmixing (LSU) methods were used for HLB detection by Kumar et al. (2010) and Kumar et al. (2012). A fairly high detection accuracy of 80% was achieved using MTMF on hyperspectral image of the SG E1- east site. SAM with multispectral images also gave a very high accuracy rate (87%). Li et al. (2010) and Li et al. (2012) analyzed and discussed spectral features derived from original reflectance, first derivative (1D), red edge position between different classes from both ground measurement and airborne images acquired in 2007 and 2010. They used several classification and spectral mapping methods to implement HLB detection in airborne multispectral (MS) and HS images and estimated their performances and adaptability to detect HLB infected canopy in citrus groves. Although a lot of research has been done to detect HLB pixels using HS images (Lee et al., 2008; Lee et al., 2009; Kumar et al., 2010; Li et al., 2011; Li et al., 2012; Kumar et al., 2012;), the accuracy of the results is either unstable or too low using all the bands or the bands chosen by using the common used method.

While HS image can provide large amounts of information, it has high redundancy because the bands of HS image are highly correlated. To solve this problem using airborne HS image, an efficient dimension reduction method need to be applied, to improve the accuracy of HLB detection and simplify the image acquisition step. However, a lot of dimension reduction methods have been proposed. Approaches such as principle components analysis (PCA) (Singh and Harrison, et al. 1985), maximum noise fraction (MNF) (Green et al., 1988) transformation, forward feature selection algorithm (FFSA) (Whitney, 1971; Kumar et al., 2001) and Kullback-Leibler divergence (KLD) based method (Martinez-Uso et al., 2007), etc. , could be used for noise reduction or bands selection. All these four methods have been used successfully in different studies. For example, PCA was successfully used to help classify blueberry fruits and leaves in the multinomial logistic regression (MNR) model by Yang et al. (2012), which had higher accuracy than that of the classification tree model. The accuracy of MNR model was 100% for the leaf and mature fruit class, and the lowest accuracy occurred in the detection of near-mature fruit, however, it was still higher than 94%. To classify HLB infected citrus trees from healthy citrus trees, the first 20 MNF components were chosen to carry out MTMF mapping method by Li X., et al. (2012), since they had more significant eigenvalues, which gave a 61.9 % accuracy using 2010 SG HS image in validation set. Similarly, FFSA showed significant improvements in classification accuracies while using a smaller number of features of AVIRIS dataset for a 12-class problem by Kumar et al. (2001), and KLD based method was proved to have a stable behavior for different image datasets and a noticeable accuracy, mainly when selecting small sets of bands by Marrinez-Uso et al. (2007). It is very important to find one specific dimension reduction method for our specific task because different methods may give different results for one problem. In this study, these four methods were applied on the HS image for dimension reduction. While PCA and MNF are feature extraction method, which means a new and reduced dataset representing the transformed initial information will be obtained by using them, FFSA and KLD based method are feature selection, or band selection methods, which means a subset of bands from the original information will be selected.

The overall objective of this research was to compare the performance of different dimension reduction methods for HLB detection using HS airborne image, which were achieved in the following three specific objectives:

1. Analyze spectral features of HLB infected and healthy canopies from ground truthing, HS images, to find the differences between them,

2. Conduct dimension reduction using PCA, MNF, FFSA and KLD method, to improve the computation complexity of the classification of HLB infected and healthy pixels, and determine optimal wavelengths which are most responsive to HLB affected citrus pixels in HS image,

3. Compare the pixel based classification results and tree based classification results obtained by these four dimension reduction methods.

Materials and Methods

Airborne image acquisition in 2011

A set of airborne HS image was acquired for Citrus Research and Education Center (CREC) grove, located in Lake Alfred, Central Florida, USA, on December, 14, 2011.

A 3.6 m by 3.6 m reference tarp was used for calibration of the reflectance value of HS data. An airborne hyperspectral camera unit – the AISA EAGLE VNIR Hyperspectral Imaging Sensor, ranging from 400 – 1,000 nm was used for imaging. This sensor has 128 spectral bands, and a spectral resolution of around 5 nm. The ground sampling distance (GSD) of the image was 0.5 m. The HS image was georeferenced to the UTM coordinate system in zone 17 N with the datum of WGS-84.

Ground Truth Measurement

In this experiment, the trees in block 8a in CREC were scouted, to identify if the tree is HLB infected or healthy on December, 14, 2011. Locations of all the measured trees were recorded with an RTK GPS receiver (HiPer XT, Topcon, Livermore, CA, USA). In total, the positions of 119 trees were collected for block 8a, in the CREC grove. The 119 pixels corresponding to the measured trees were extracted from the HS image to do the further pixel based analysis. The measured pixels were classified into two classes, which were hlb infected and healthy pixels, as shown in Table 1. The corresponding HS RGB of this block was shown in Figure1. The tree infection status was determined by experienced ground inspection crews at the CREC grove in July, 2011, which was five months earlier than the ground truth. Their inspection results were comparable to a PCR laboratory test, approximately 95% of the time.

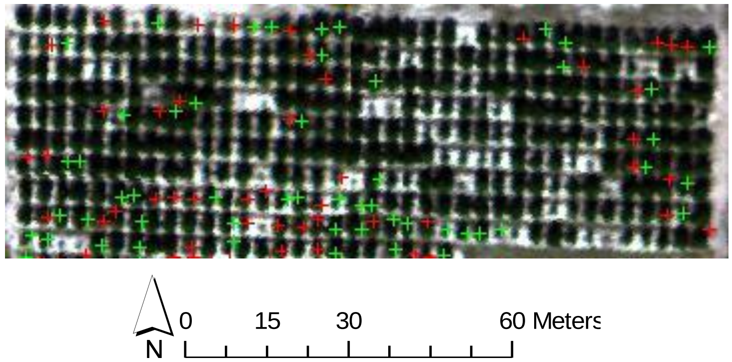


Figure 1. The airborne HS RGB image of the block in CREC. The green crosshairs and the red crosshairs indicated the locations of the healthy trees and HLB infected trees, respectively, determined by experienced ground inspection crews.

Table 1. Brief description of different classes of citrus canopies used in this study.

Class	Description	Number of pixels
8a-hlb	HLB infected canopy in the block	60
8a-healthy	Healthy canopy in the block	61

Methods

K-nearest neighbor classification (KNN)

Because one of the objectives of this paper was to compare the performance of different dimension reduction methods, only one classifier was needed to do the related classification accuracy analysis for the sake of uniformity. KNN classifier was chosen in this study since it was among the simplest of all machine learning algorithms. As the preliminary experimental results indicates that KNN performed the best compared with other supervised methods, such as Naïve Bayesian classifier (NBC), it was chosen for the further accuracy analysis. Given an unknown feature vector *x* and a distance measure (Euclidean distance in this paper), the KNN rule classifies *x* by assigning it the label most frequently represented among the *k* nearest neighbors and taking a vote. *K* was assigned 1 in this paper. When *k*=1, the feature vector is assigned to the class of its nearest neighbor (Duda et al., 2001).

Principal component analysis (PCA)

PCA (Singh and Harrison, et al. 1985) is mathematically defined (Jolliffe, 2002) as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. It can help find alternative uncorrelated variables, and each component as a vector is orthogonal to the others. Usually, the first few principal components contribute the greatest part of the information of the data.

Maximum Noise Fraction (MNF) transformation

The MNF (Green et al., 1988) transform as modified from Green et al. (1988) is a linear transformation that consists of the following two separate principal components analysis rotations. The first rotation uses the principal components of the noise covariance matrix to decorrelate and rescale the noise in the data, which is a process known as noise whitening, resulting in transformed data in which the noise has unit variance and no band-to-band correlations. The second rotation uses the principal components derived from the original image data after they have been noisewhitened by the first rotation and rescaled by the noise standard deviation. Since further spectral processing will occur, the inherent dimensionality of the data is determined by examining the final eigenvalues transformation coefficients in principle components analysis that can be used to determine the percentage of total variance explained by each of the principle components and the associated images. The data space can be divided into two parts. One part associated with large eigenvalues and coherent eigen images, and a complementary part with near-unity eigenvalues and noise-dominated images. Using only the coherent portions to separate the noise from the data, the spectral processing results can be improved.

Forward feature selection algorithm (FFSA)

FFSA was originally proposed by Whitney (1971), and was named FFSA by Kumar et al. (2001). In this method, the band(s) are added from the available set until the increase in the classification result is not significant. The subset that provides the best classification result will be selected. FFSA is the direct but nonparametric evaluation of subsets, and should be used with a classifier. The procedures are as following. Firstly, one specific classifier is chosen to do the classification by using the chosen bands. KNN classifier was chosen in this study. Then, it considers each band individually at the start of the search. By considering which measurement is best when adjoined to the previously selected subset, the procedure is responsive to dependencies between measurements as a function of sample classification. The new band is selected so that the increase in classification accuracy with its addition is the maximum. This procedure is repeated iteratively until all bands have been used up or the classification accuracy does not improve significantly. Although this technique employs a suboptimum procedure, instead of using all the subsets of the total measurements, it has the following advantages. This method permits tradeoffs between system complexity and performance. Zero percent increase was considered the threshold in the study.

Kullback-Leibler divergence (KLD) based band selection method

This Kullback-Leibler divergence measure (Martinez-Uso et al., 2007) can be used as a criterion to know how far two distributions are, and it can be interpreted as the cost of using one of the distributions instead of the other one. In the hyperspectral band selection framework, it can be used as a measure of dissimilarity between two image bands, which are represented by their corresponding probability distributions.

This divergence measure is one criterion that used as a distance for the clustering process, and it has been frequently used in order to compare different probability distributions, also in hyperspectral imaging to measure the overlapped information that is contained in a pair of image bands, as a band-decorrelation algorithm.

The symmetric Kullback-Leibler divergence can be expressed in the discrete domain as following,

$$D_{KL}(X_i, X_j) = \sum_{x\in\Omega} p_i(x)\log\frac{p_i(x)}{p(x)} + \sum_{x\in\Omega} p_j(x)\log\frac{p_j(x)}{p_i(x)} \tag{1}$$

The Kullback–Leibler divergence is always nonnegative, being zero when $p_i(x)$ and $p_j(x)$ are the same probability distribution. This divergence measure can be used as a criterion to know how far two distributions are, and it can be interpreted as the cost of using one of the distributions instead of the other one.

All these five methods mentioned above were realized in Matlab (Ver. R2010a, Mathworks, Natick, MA, USA). The dimension reduction results obtained from those four methods, including PCA, MNF, FFSA, and KLD, will be compared in three ways. Firstly, the bands chosen by them will be compared. Then, the pixel based classification results by using KNN classifier will be used. Thirdly, the tree based classification results will be compared.

Tree based classification

The reason why the tree based classification was needed here is that one HLB infected tree could have many pixels which were infected. The ones used in the former study is not necessarily the most infected one in the trees. That is why the detection accuracy of the pixel based analysis is not very high. For tree based classification, if any one of the pixels of the measured tree was detected as infected, the detection result of the tree would be infected. Support vector machine (SVM) was used to create one mask for the trees in the HS image, using the procedure used by Li et al (2011). Because Li et al. (2011) summarized that for tree based classification, Mahalanobis distance (MahaDist) had more balanced accuracy rates between the calibration and validation sets, it was chosen to test the tree based classification results for those four dimension reduction methods in this paper. The calibration set used by MahaDist was the same one used for pixel based classification. This procedure was done using ENVI (Exelis Visual Information Solutions, Inc., Boulder, Colorado, USA).

Results and discussion

Spectral feature analysis

The whole dataset, including 60 HLB infected pixels, and 61 healthy pixels, which were used in this paper was divided into two parts, calibration and validation datasets. 31 healthy pixels and 30 HLB pixels were used as a calibration dataset, while the rest of them were used as a validation dataset, respectively. The spectral analysis for the two calibration datasets was shown in Figure 2.

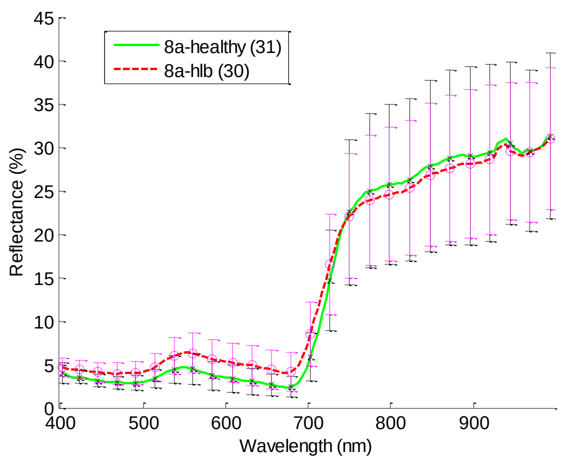
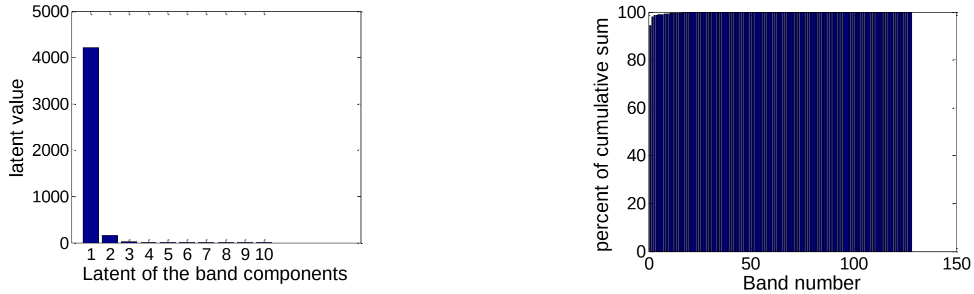


Figure 2. Spectral feature analysis of the HS image data, using mean value and Std shown in bars: mean HS image spectra for calibration dataset of healthy and HLB infected samples in the block.

From Figure 2, we can see that in the visible range (400-700 nm), the mean reflectance of the healthy samples is lower than that of the HLB infected samples, while the mean reflectance of the healthy samples in 700-1000 nm is much higher than that of the HLB infected samples. This result is consistent with the result described by Lee et al. (2008) and Li et al. (2012).

PCA

In Figure 3a, latent is a vector containing the eigenvalues of the covariance matrix of the calibration samples. As shown in Figure 3a, the latent value dropped dramatically after the first band. As we need to take only features containing 98.5% variance, the cumulative sum was calculated, as shown in Figure 3b. The cumulative sum indicates that the first three eigenvectors alone account for over 98.5%. Thus, the first three principal components (PCs) were chosen for PCA dimension reduction.



(a) (b)

Figure 3. The results for PCA using CREC block 8a data: (a) the latent value of the band components, and (b) the percent of cumulative sum of the latent value for the band components.

MNF

MNF transform was performed on the calibration dataset. Every MNF band has an eigenvalue, which rapidly decreases while the number of band increases, as shown as Figure 4. The larger the eigenvalue is, the more useful information it contains. Bands with large eigenvalues (greater than 1) contain data, and bands with eigenvalues near 1 contain noise. In subsequent processing of this data, the subset of the MNF bands only include those bands where the images appear spatially coherent and the eigenvalues are above the break in slope of the MNF eigenvalue plot were chosen. In this case, we chose the first 18 components after MNF transformation.

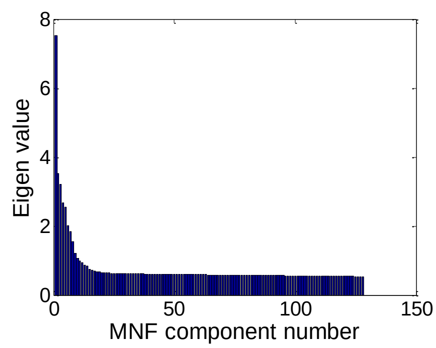
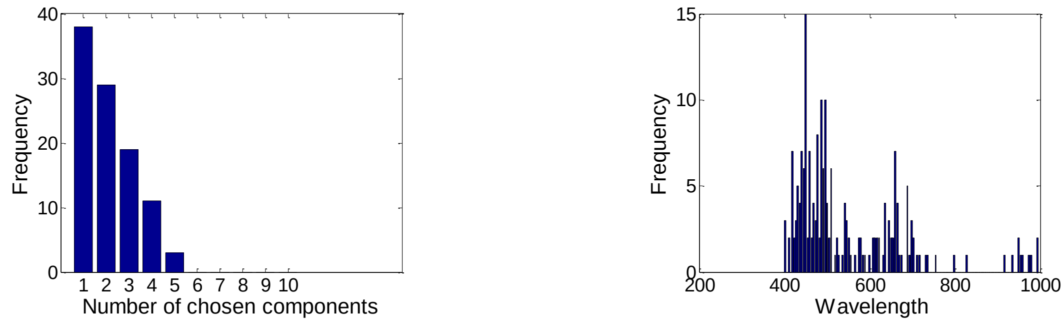


Figure 4. Eigenvalues for each band after MNF transform.

FFSA

FFSA was used to choose the best subset of bands with best performance among the 128 bands of the HS image. For the chosen classifier, which was KNN in this case, the best single color component with the highest accuracy was selected first, and then the best pair was selected where the pair included the best single measurement selected by iterating ten times. A new band was added to the previously selected ones so that the increase in classification accuracy with its addition was the maximum. This procedure was repeated iteratively until the classification accuracy did not improve anymore. In this study, FFSA was repeated 100 times to decide how many bands should be chosen and which combination of bands should be chosen. Figure 5a shows the band number chosen frequency by using FFSA, from which we can see that one band was chosen most frequently. Figure 5b shows the frequency of bands chosen using FFSA, which indicates that the first three most frequently chosen bands were band 13, 19 and 21, and they were chosen 15, 10 and 8 times, respectively. These bands were corresponding to the band 450 nm, 468 nm and 487 nm, respectively.



(a) (b)

Figure 5. (a) The chosen frequency of the first ten components using FFSA, and (b) the frequency of bands chosen using FFSA.

KLD

This method used KL divergence as a distance for the band clustering process. Then hierarchical clustering was performed to find out which bands have the smallest divergence with other bands in the same cluster. Firstly, a hierarchical cluster tree was built using the KL divergence matrix of the 128 bands. Then, the bands were clustered using this hierarchical cluster tree. At last, the bands with the highest average correlation coefficient were chosen, because they had the most similarity with other bands in this cluster. When the cluster number was set to six, six clusters were generated, as shown in Figure 6. After clustering, the bands with smallest average correlation value with other bands in the same cluster were chosen as the representatives. As a result, the band 3, 5, 13, 8, 52, 109 were chosen from 128 bands using the calibration dataset. These bands' wavelengths were 405 nm, 414 nm, 450 nm, 427 nm, 631 nm and 905 nm, respectively.

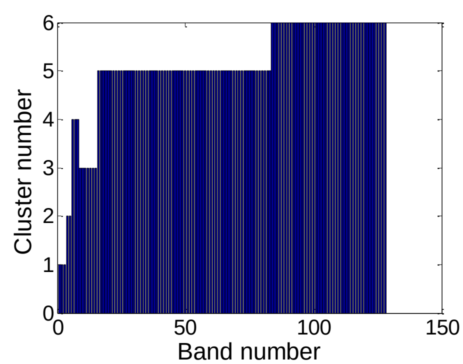


Figure 6. Clusters generated through KLD.

Table 2 summarizes the dimension reduction results of those four band selection methods. The results implied that the PCA and FFSA both recommended three PCs or bands for classification, while KLD has recommended six bands.

Table 2 Band selection results by using different methods

Dimension reduction method	Band selection results
PCA	first 3 PCs after PCA transformation
MNF	first 18 compomemts after MNF transformation
FFSA	450 nm, 468 nm and 487 nm
KLD	405 nm, 414 nm, 427 nm, 450 nm, 631 nm and 905 nm

Pixel based classification

By using the pixel dataset extracted from the HS image, the accuracy analysis results using KNN classifier for bands chosen for these four methods, PCA, MNF, FFSA, and KLD are shown in Table 3. All these four methods, showed 100% identification rate in the calibration set of both hlb and healthy dataset.

For the hlb detection in the validation set, KLD showed the highest accuracy, 63.3%, while the lowest identified rate were resulted from both PCA and FFSA, which was 53.2 %. For healthy pixel detection in the validation set, both PCA and KLD showed the highest detection percentage, which was 74%, while MNF showed the lowest detection rate, which was 60%.

Table 3. Pixel based accuracy analysis using KNN classifier for bands chosen by different methods.

Pixel class	Dimension reduction method	Calibration set (C)		Validation set (V)	
		Detected pixels	Percent (%)	Detected pixels	Percent (%)
hlb (C:30 V:30)	PCA	30	100	16	53.2
	MNF	30	100	18	60.0
	FFSA	30	100	16	53.2
	KLD	30	100	19	63.3
healthy (C:31 V:30)	PCA	31	100	23	74
	MNF	31	100	18	60
	FFSA	31	100	20	65
	KLD	31	100	23	74

Tree based classification

The accuracy results for tree based classification were summarized in Table 4. MNF and KLD showed 93.3% HLB detection accuracy, which was fairly high, while PCA showed 86.7% accuracy and FFSA showed an 83.3% accuracy.

Table 4. Tree based accuracy analysis using MahaDist classifier for bands chosen by different methods.

Tree class	Dimension reduction method	Calibration set (C)		Validation set (V)	
		Detected trees	Percent (%)	Detected trees	Percent (%)
hlb (C:30 V:30)	PCA	30	100	26	86.7
	MNF	30	100	28	93.3
	FFSA	30	100	25	83.3
	KLD	30	100	28	93.3

Figure 7 showed the results of tree based classification for four different dimension reduction methods. Figures 7c - 7f showed the PCA, MNF, FFSA, and KLD result for the area of the red square in Figure 7a. Their mask were created using SVM, respectively. The blue pixels in these four images were background, and the rest of the pixels were trees. The green pixels were healthy ones, while the red pixels were HLB infected ones, after classification.

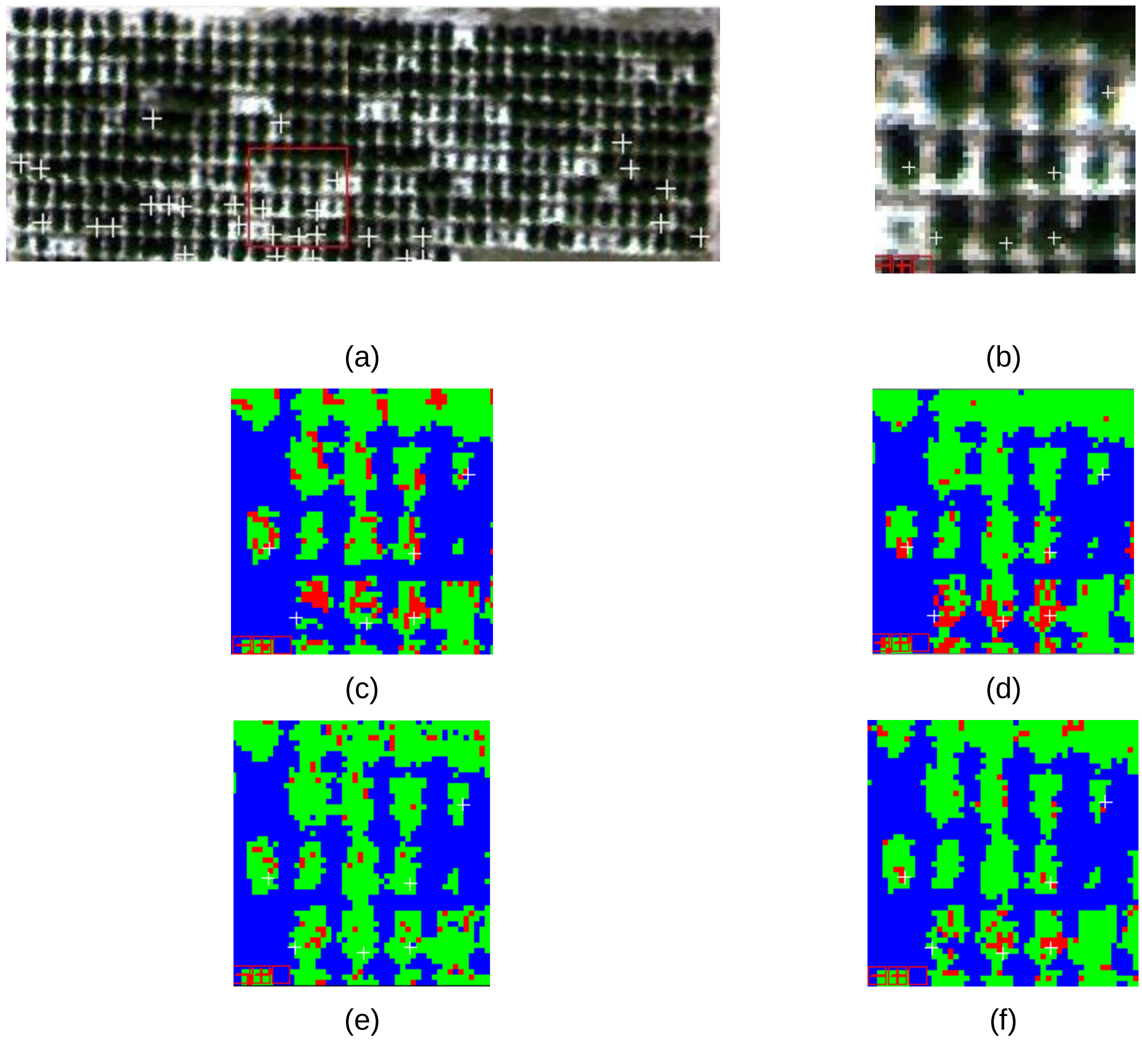


Figure 7. The results of tree based classification for four different dimension reduction methods: (a) the HS RGB image. The white crosshairs were the positions of the hlb validation set, indicating whether the tree is HLB infected or not, (b) zoomed in area of the red square of (a), and (c), (d), (e), (f) showed the PCA, MNF, FFSA, and KLD result for the area of the red square of (a), respectively.

Discussion

It appears that all of these four methods, PCA, MNF, FFSA and KLD, have greatly reduced the bands number, and were able to improve the computation complexity of the classification of HLB infected pixels and healthy pixels. PCA and MNF are feature extraction methods, which means a new and reduced dataset representing the transformed initial information were obtained. While MNF is essentially two cascaded PCA (Green et al., 1988), MNF chose 15 more components than PCA. This is basically because of their distribution of the eigenvalues, which were shown in Figure 3a and Figure 4, respectively.

The bands chosen by FFSA were three bands between blue and green, which usually can't be used independently to deal with complicated classification. However, from Figure 2, the reflectance of these three bands did have major difference between HLB infected and healthy pixel. Among the six bands chosen by KLD, five of them (405 nm, 414 nm, 427 nm, 450 nm, and 631 nm), were visible bands, and one of them (905 nm) was an NIR band. This again proved the importance of the visible bands for HLB detection. The reason why KLD has chosen one NIR band, probably is that the red edge position mentioned by Li et al (2011) also can help with the detection of HLB. From Figure 3, major difference in near infrared (NIR) band between healthy and HLB infected curve can also be observed.

For the pixel based classification results obtained in this paper, the detection accuracy of these four methods were relatively low, which is probably because of the dataset was too small, which caused the over classification of the classifier. The other reason might be the limitation of the ground truth accuracy. However, the results still could give us some information about these four different dimension reduction methods. Among these four methods, the bands chosen by KLD performed the best. The bands chosen by KLD performed better than FFSA, mostly because it has utilized the NIR band. Compared with MNF, KLD used a smaller number of bands, and also yielded better result, which indicated that more bands are not always necessarily better when it comes to this specific problem.

For the tree based classification, both KLD and MNF showed highest accuracy of more than 93% in the validation set. From Figures 7c - 7f, which were corresponding to PCA, MNF, FFSA and KLD, the classification results of different methods could be observed from one subset area of the block. Both Figures7d and 7f showed multiple pixels on the infected trees, with less false positive detections, compared with the other two methods. Figure 7c (PCA) showed good detection result, but more false positives were produced compared with MNF and KLD. Since the status of all trees can not be identified due to the large number of the trees in the HS image,, the rate of the healthy tree detection in the image was not calculated here.

Conclusion

Using the HS image and ground truth data obtained in 2011, the spectral features of HLB infected and healthy citrus groves were analyzed, which indicated the promising application of HLB detection using HS image. Four different dimension reduction methods, PCA, MNF, FFSA and KLD based method were conducted for the determination of band choice for HLB detection using HS image.

a) It was shown that all of these four methods have greatly reduced the dimension, and were able to improve the computation complexity of the classification of HLB infected pixels and healthy pixels. The first three bands were chosen after analysis of PCA transformation and the first 18 bands were chosen after MNF transformation. While FFSA and KLD based method are feature selection, or band selection methods, subsets of bands from the original information were selected. Wavelengths of 450 nm, 468 nm and 487 nm were chosen by FFSA, while 405 nm, 414 nm, 427 nm, 450 nm, 631 nm, and 905 nm were chosen by KLD based method.

b) For pixel based classification using KNN classifier, these four methods all showed 100% accuracy in the calibration dataset for both HLB detection and healthy sample detection. KLD showed the highest HLB detection accuracy of 63.3%, while PCA and KLD based method showed the highest healthy detection accuracy of 74% using the validation dataset.

c) For tree based classification using MahaDist, MNF and KLD showed 93.3% HLB detection accuracy, which was fairly high.

In general, KLD produced the best HLB detection rates for both pixel and tree based classification. Therefore, the bands chosen by KLD were recommended for further research. The promising application of the dimension reduction of the HS image was demonstrated to detect HLB disease. However, since the limitation of the dataset used in this paper, further research with bigger dataset is needed in the future.

Acknowledgement

This project was funded by the Citrus Research and Development Foundation, Inc. The author would like to thank the guidance provided by her advisers, Dr. Won Suk Lee, and Dr. Ku Wang. The author also appreciates the help provided by Ms. Ce Yang, Ms. Sherrie Buchanon, Ms. Lioubov Polonik, Mr. Anurag R. Katti, Mr. Alireza Pourreza at the University of Florida in this study. The authors also would like to thank China Scholarship Council for financial support.

References

California Department of Food and Agriculture and the USDA. 2012. Citrus Disease Huanglongbing Detected in Hacienda Heights Area of Los Angeles County. Available: http://weblogs.nal.usda.gov/invasivespecies/archives/2012/04/citrus_greening.shtml

Duda, R. O., P. E. Hart, and D. G. Stork. 2001. Chapter 4: Nonparametric techniques. *Pattern Classification*, 182-186. John Wiley & Sons, Inc.

FDACS/DPI. (2011). "Sections (TRS) Positive for Huanglongbing (HLB, Citrus Greening) in Florida," Available: http://www.freshfromflorida.com/pi/chrp/greening/StatewidePositiveHLBSections.pdf

Green, A. A., M. Berman, P. Switzer, and M. D.Craig. 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Trans. Geosci. Remote Sensing*, 26(1): 65-74.

Jolliffe, I.T. 2002. *Principal Component Analysis,* Series: Springer Series in Statistics, 2nd ed., Springer, NY.

Kumar, A., W. S. Lee, R. Ehsani, L. G. Albrigo, C. Yang, and R. L. Mangan. 2010. Citrus greening disease detection using airborne multispectral and hyperspectral imaging. In *Proc. 10th International Conference on Precision Agriculture, Denver, Colorado.*

Kumar, A., W. S. Lee, R. Ehsani, L. G. Albrigo, C. Yang, and R. L. Mangan. 2012. Citrus greening disease detection using aerial hyperspectral and multispectral imaging techniques. *J. Appl. Remote Sens*. 6(1), 063542.

Kumar, S., J. Ghosh, and M. M.Crawford, 2001. Best-bases feature extraction algorithms for classification of hyperspectral data. *IEEE Trans. Geosc. Remote Sensing*, 39(7), 1368–1379.

Lee, W. S., R. Chinchuluum,and R. Ehsani. 2009. Airborne hyperspectral imaging for citrus greening disease detection. *Proceedings of the 3rd Asian Conference on Precision Agriculture (ACPA)*, 14-17 October 2009, Beijing, China.

Lee, W. S., Ehsani, R. and Albrigo, L. G. 2008. Citrus greening (Huanglongbing) detection using aerial hyperspectral imaging. *In the Proceedings of the 9th International Conference on Precision Agriculture,* Denver, Colorado.

Li, X., W. S. Lee, M. Li, R. Ehsani, A. R. Mishra, and C. Yang. 2011. Comparison of different detection methods for citrus greening disease based on airborne multispectral and hyperspectral imagery. *Proceedings of 2011 ASABE*, Paper No. 1110570. St. Joseph. Mich.: ASABE.

Li, X., W. S. Lee, M. Li, R. Ehsani, A. R. Mishra, C. Yang, and R. L. Mangan. 2012. Spectral difference analysis and airborne imaging classification for citrus greening infected trees. *Computers and Electronics in Agriculture*, 83: 32–46.

Martinez-Uso, A., F. Pla, J. M. Sotoca, and P. Garcia-Sevilla. 2007. Clustering based hyperspectral band selection using information measures. *IEEE Trans. Geosci. Remote Sens.*, 45(12): 4158–4171.

Singh, A., and A.Harrison, 1985. Standardized principal components. *Int. J. Remote Sensing*, 6 (6), 883-896.

Texas Department of Agriculture and the USDA. 2012. Texas Department of Agriculture and USDA Confirm Detection of Plant Disease That Damages Citrus Trees. Available: http://www.texascitrusgreening.org/

USDA/National Agricultural Statistics Service, "Citrus Fruits 2012 Summary," available: http://usda01.library.cornell.edu/usda/current/CitrFrui/CitrFrui-09-20-2012.pdf, 2012.

Whitney, A. W. 1971. A direct method of nonparametric measurement selection. *IEEE Trans. Comput.*, 20, 1100–1103.

Yang, C., W. S.Lee, and J. G. Williamson. 2012. Classification of blueberry fruit and leaves based on spectral signatures. *Biosystems Engineering*, 113(4): 351-362.