SPECTRAL ANGLE MAPPER (SAM) BASED CITRUS GREENING DISEASE DETECTION USING AIRBORNE HYPERSPECTRAL IMAGING

H. Li

China Agricultural University
University of Florida
Gainesville, Florida

W. S. Lee

Agricultural and Biological Engineering
University of Florida
Gainesville, Florida

K. Wang

China Agricultural University
Beijing, China

R. Ehsani

Citrus Research and Education Center
University of Florida
Lake Alfred, Florida

C. Yang

Kika de la Garza Subtropical Agricultural Research Center
USDA ARS
Weslaco, Texas

ABSTRACT

Over the past two decades, hyperspectral (HS) imaging has provided remarkable performance in ground objects classification and disease identification, due to its high spectral resolution. In this paper, a novel method named ‘extended spectral angle mapping (ESAM)’ is proposed to detect citrus greening disease (Huanglongbing or HLB), which is a destructive disease of citrus. Firstly, Savitzky-Golay smoothing filter was applied to the raw image to remove spectral noise within the data, yet keep the shape, reflectance and absorption features of the spectrum. Then support vector machine (SVM) was used to build a mask to segment tree canopy from the other background. Vertex component analysis (VCA) was chosen to extract the pure endmembers of the masked dataset, due to its better performance compared to other spectral linear unmixing methods.

Spectral angle mapping (SAM) was applied to classify healthy and citrus greening disease infected areas in the image using the pure endmembers as an input. Finally, red edge position (REP) was used to filter out most of false positive detections. The experiment was carried out with the image acquired by an airborne hyperspectral imaging system from the Citrus Research and Education Center (CREC) in Florida, USA. Ground truth including ground reflectance measurement and diseased tree confirmation was conducted. The experimental results were compared with another supervised method, Mahalanobis distance, and an unsupervised method, K-means. The ESAM performed better than those two methods.

Keywords: Citrus greening, ESAM, K-means, Mahalanobis distance, REP, SAM, SVM, VCA.

INTRODUCTION

Over the past two decades, hyperspectral (HS) imaging has provided remarkable solutions to the needs of a lot of applications in obtaining land cover information, due to its high spatial and spectral resolution (Ustin et al., 2004). Hyperspectral remote sensors, such as airborne visible infrared imaging spectrometer (AVIRIS), and multispectral infrared and visible imaging spectrometer (MIVIS), are now available for precision agriculture applications, such as yield estimation, target detection, environmental impact assessment, etc. (Plaza et al., 2009; Zhang et al., 2003; Yea et al., 2008).

Disease detection of vegetable or tree crops using hyperspectral data has become a subject of intensive research. Many researchers have evaluated the usefulness of HS data for disease detection of various crops or citrus fruit. Zhang et al. (2003) investigated the detection of stress in tomatoes induced by late blight disease in California using HS image. They combined minimum noise fraction (MNF) and spectral angle mapping (SAM) methods. Results showed that the late blight diseased tomatoes at stage three or above could be separated from the healthy plants. Smith et al. (2005) found that in the spectral data, the red edge position was strongly correlated with chlorophyll content across all treatments. Stress due to extreme shade could be distinguished from the stress caused by natural gas and herbicide from the change in spectrum. Huang et al. (2007) used in-situ spectral reflectance measurements of crop plants infected with yellow rust to develop a regression equation to characterize the disease index. This was validated in the subsequent growing season, and then was applied to hyperspectral airborne imagery to discriminate and map the disease index in target fields. Lee et al. (2008) used HS images to detect the citrus greening disease by applying SAM and spectral feature fitting (SFF) methods. They reported that it was difficult to obtain good results because of the positioning errors of GPS ground truth and aerial imaging, and the spectral similarity between healthy and the citrus greening disease infected trees. Qin et al. (2009) developed a spectral information divergence (SID) based algorithms for hyperspectral image processing and

classification to differentiate citrus canker lesions from normal and other diseased peel conditions. The SID based classifier could differentiate canker from normal fruit peels and other citrus diseases, and it also could avoid the negative effects of stem-ends and calyxes. The overall classification accuracy of 96.2% was achieved. Li et al. (2012) used both ground and airborne remote sensing to find the spectral differences between HLB and healthy citrus canopies. Several classification and spectral mapping methods were implemented in airborne multispectral (MS) and HS images and their performances and adaptability to detect HLB infected canopy in citrus groves were then compared and evaluated.

Citrus greening, also known as Huanglongbing (HLB), caused by Asian citrus psyllids, is a disease which has no cure reported yet. 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 (Smith et al., 2005; Huang et al., 2007; Lee et al., 2008; Qin et al., 2009). Timely and location-specific detection and monitoring of the infected citrus trees are required for efficient disease control while reducing pollution risks. The disease detection methods currently used, such as conventional ground scouting, electron microscopy and bioassay, and polymerase chain reaction (PCR), are expensive and time consuming. Remote sensing, on the other hand, can quickly collect citrus grove canopy data that can be used to analyze geo-temporal and geo-special properties of the biological features of the tree canopies, including the symptoms of the citrus greening.

The overall objectives of this study were to develop a method to classify citrus greening infected trees from healthy trees using HS image, based on the analysis of spectral features of HLB infected and healthy canopies from both ground truth and HS image. The performance and adaptability of the proposed method was evaluated and compared with two other methods: K-means and Mahalanobis distance. The promising application of HS image was demonstrated to detect HLB disease.

MATERIALS

Image acquisition in 2011

In December 2011, a set of aerial hyperspectral images was acquired for three blocks of the Citrus Research and Education Center (CREC) grove along with ground truth data, which was located in Lake Alfred, Central Florida, USA.

A reference tarp was used for calibration of the reflectance value of HS data. Fig.1 is the reflectance curve of the reference tarp measured using a handheld spectrometer (HR-1024, Spectra Vista Corporation, Poughkeepsie, NY, USA), which had a spectral range of 348-2505 nm with an interval of 3 nm.

The HS image was georeferenced to the UTM coordinate system in zone 17 N with the datum of WGS-84, and the ground sampling distance (GSD) of the final image was 0.5 m. A total of 128 spectral bands in 400-1,000 nm were collected, which had the digital number (DN) ranging from 0 to 4095. The spectral resolution was 5 nm.

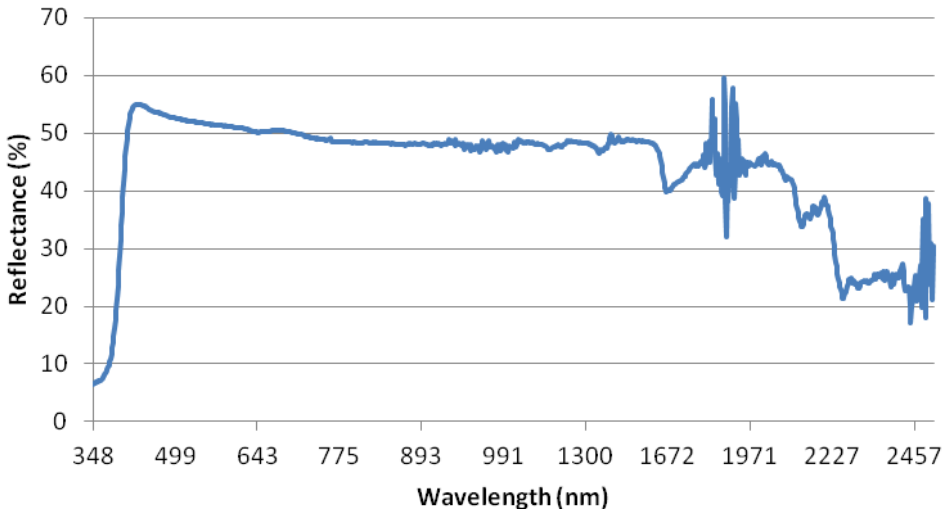


Fig. 1. Reflectance of the reference tarp, made of type 822 fabrics, which is moderate weight woven polyester substrate with long-term durability. The size of the tarp was 3.6 m by 3.6 m, and the average reflectance of the tarp was 56% in 420-1050 nm.

Ground truth measurement

In the 2011 experiment, two types of ground truth were measured: ground spectral reflectance and location data for the measured trees. Ground spectral reflectance of each tree canopy was measured using the handheld spectrometer. A white reference panel was used for calibration. For each measured leaf, three scans were conducted consecutively. Locations of all the measured trees were recorded with an RTK GPS receiver (HiPer XT, Topcon, Livermore, CA, USA).

In total, the positions of 96 trees were collected in a block in the CREC grove. The measured trees were classified into two classes, which are 45 HLB infected trees and 51 healthy trees, as shown in Table 1. The tree status was determined by experienced ground inspection crews at the CREC grove.

Table 1. Brief description of tree canopy classes used in this study.

Class	Description	Number of trees
hlb	HLB infected canopy	45
healthy	Healthy canopy	51

METHODS

Taking into account both the spectral and spatial characteristics of hyperspectral datasets, many data processing techniques have been developed and used in HS images. In this paper, a novel method, named ‘extended spectral angle mapping (ESAM)’, is proposed to detect citrus greening disease using HS image. In the proposed ‘ESAM’ method, different hyperspectral image processing

techniques, such as Savitzky-Golay smoothing filter, support vector machine (SVM), vertex component analysis (VCA), spectral angle mapping (SAM) and red edge position (REP), were combined together to obtain the best results in this study. Firstly, Savitzky-Golay smoothing filter was applied to the raw image to remove spectral noise within the data, yet keep the shape and absorption features of the spectrum (Savitzky and Golay, 1964). Then SVM was used to build a mask to segment tree canopy from the other background (Li et al., 2012). VCA was chosen to extract the pure endmember of the masked dataset, due to its better performance compared to other spectral linear unmixing methods (Nascimento and Dias, 2005). SAM was applied to classify healthy and the citrus greening disease infected areas in the image using the pure endmember chosen by VCA. Finally, REP was used to filter out most of the false positive detections (Collins et al., 1977; Collins, 1978; Cho et al., 2006; Dawson et al., 1998).

Two other methods were also performed on the 2011 HS image. A supervised method, Mahalanobis distance (MahaDist), was chosen because it showed more balanced results according to the work by Li et al. (2012). An unsupervised method K-means was also tested in this study.

ENVI (Exelis Visual Information Solutions, Inc., Boulder, Colorado, USA) was used for HS image analysis. Using the RTK data obtained from the ground truth, HS image data were exported from the corresponding position. Block 8ab in the HS image was chosen to be an example grove to implement the proposed ESAM, MahaDist, and K-means methods mentioned above. Among the sample set including 51 healthy samples and 45 HLB samples, a subset of 26 healthy pixel spectra and 23 HLB infected spectra was chosen to form a calibration set. The rest of the samples, including 25 healthy pixel spectra and 22 HLB infected spectra were chosen to form a validation set.

RESULTS AND DISCUSSION

Spectral feature analysis

Ground truth and HS image based hyperspectral data from block 8ab in the CREC grove obtained in 2011 were used for spectral feature analysis. Although the ground hyperspectral measurements had a spectral range of 348 - 2505 nm, only data ranging from 400 nm to 1000 nm were used in this study for a better comparison with the HS image data having the wavelength range of 400-1000 nm. Although brightness conditions of each leaf were different due to illumination change when the experiments were conducted, the mean spectra of different classes can imply some different characteristics. Standard deviation (Std) is a widely used measure of variability or diversity used in statistics and probability theory. It shows how much variation exists from the mean value. Both mean and Std were used in the analysis of the feature of the dataset.

From the ground measurements, two sample class spectra (healthy and HLB), ranging from 348 nm to 1000 nm are shown in Fig. 2, and their spectral data from the HS image are shown in Fig. 3. In Figs. 2 and 3, the solid green lines are the average spectra of healthy samples, and the red solid lines are the average spectra of HLB infected samples, respectively. The Std and mean values for these two classes are marked in the figures.

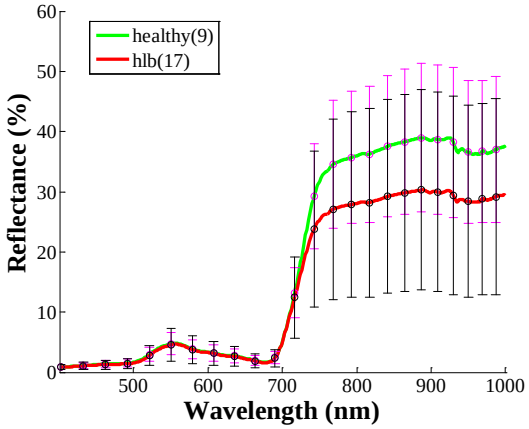


Fig. 2. Average reflectance spectra of healthy (green line) and HLB infected (red line) canopies from the ground measurements. The vertical lines are Std at selected wavelengths. The number in a parenthesis indicates the number of samples for calculating an average.

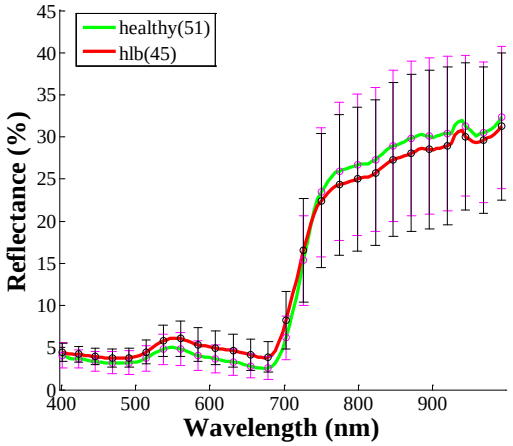


Fig. 3. Spectral feature analysis of HS image data. The solid lines are mean HS image spectra for healthy and HLB infected samples, marked with mean value and Std.

From Figs. 2 and 3, the obvious reflectance difference can be seen in both ground truth data and HS image. In Fig. 2, below 700 nm, the mean reflectance difference of the two classes is very little. Nevertheless, after 700 nm, the mean reflectance difference is very obvious. The mean reflectance of the healthy samples is much higher than that of the HLB infected samples. In Fig. 3, in the visible range (400-730 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 730-1000 nm is much higher than that of the HLB infected samples. This result is consistent with the result described by Lee et al. (2008).

Results of ESAM

After the Savitzky-Golay smoothing filter was applied, the training set containing the two classes was used to find pure pixels for the two classes using VCA without assigning the category of each sample. Two pure endmembers were selected successfully, and are shown in Fig. 4. Their spectral features were consistent with those analyzed above. The solid green line is the 5th sample selected among the 26 samples of healthy training set. The solid red line is the 13th sample selected in the 23 samples of HLB infected training set. The pure pixel spectra were used as the spectral library to carry out SAM.

SVM was performed on the block and a mask was obtained based on the tree class. A mask for the tree canopy was built and applied on the image. The result is shown in Fig. 5.

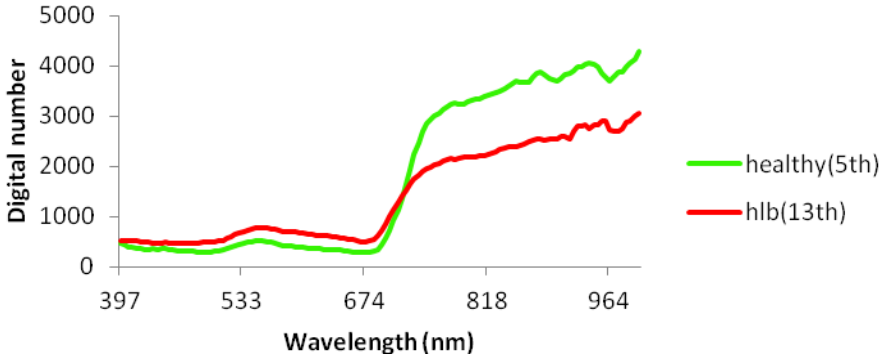
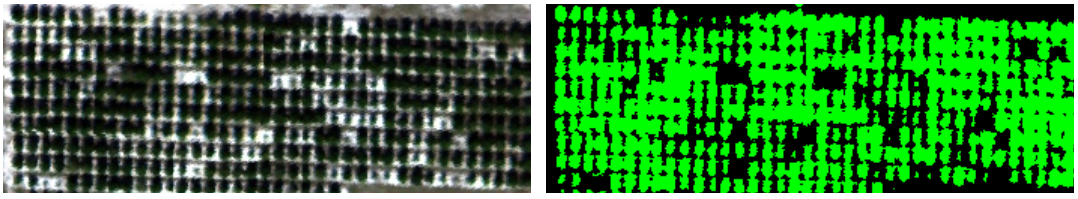


Fig. 4. Pure endmember spectra chosen by VCA. The solid green line is the 5th sample selected among the 26 samples of healthy training set. The solid red line is the 13th sample selected in the 23 samples of HLB infected training set.



(a) (b)

Fig. 5. SVM classification and masked results: (a) Original HS RGB image of the block. (b) Mask for tree class applied on the HS RGB image.

A threshold was needed as an input parameter when SAM was applied on the masked image. It was very important for the classification result. If the value was too high, false positives would be introduced. If it was too low, the image will be over-classified. To choose a proper threshold, the spectral angle between each

data and the target endmember chosen by VCA, were calculated, as shown in Fig. 6.

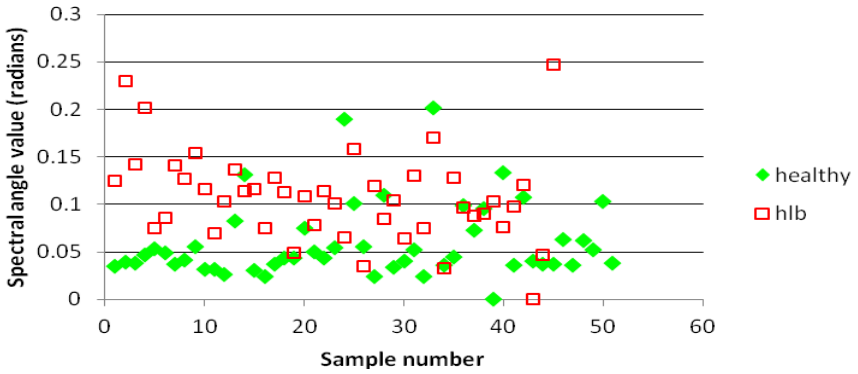


Fig. 6. Spectral angle value between each data and the target endmember.

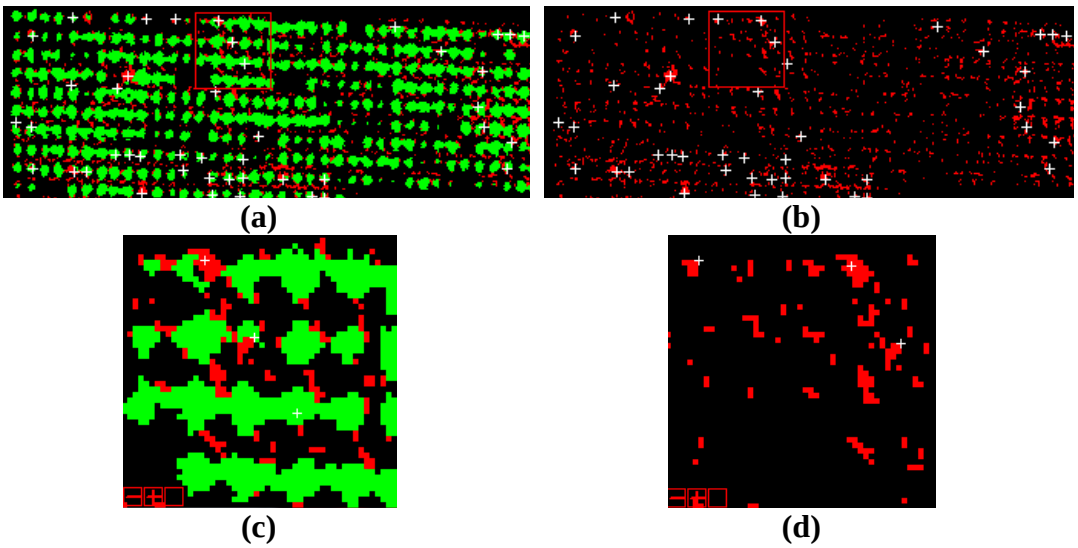
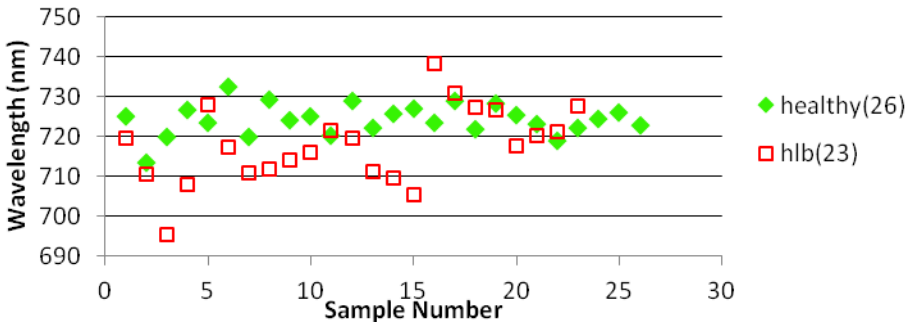


Fig. 7. SAM results applied on the block, red pixels are infected area, and green pixels are healthy area: (a) SAM results with spectral angle 0.1 for the HLB infected pixels and 0.15 for healthy pixels. (b) Results of HLB infected pixels. (c) and (d) are the zoomed-in image of the area marked using a red square in (a) and (b), with white crosshairs showing HLB infected positions.

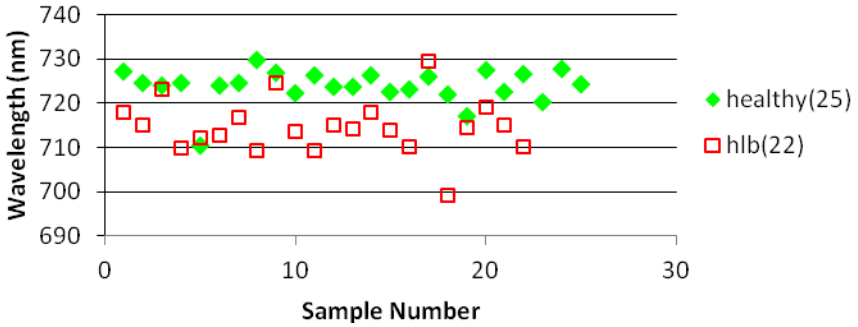
Multiple maximum spectral angles were chosen based on the processed results. The detection accuracy for each category is shown in Table 2. The higher the threshold is, the higher the accuracy for classification is. A trade off should be made to get better detection result, yet not induce too many false positives. The spectral angle of 0.1 was chosen for the healthy category and spectral angle of 0.15 was chosen for the HLB infected category for the HS image analysis.

Table 2. Detection accuracy for the total data set using different thresholds for SAM method.

Data set	Threshold (0.05)		Threshold (0.1)		Threshold (0.15)	
	Samples (pixel)	Percent (%)	Samples (pixel)	Percent (%)	Samples (pixel)	Percent (%)
hlb (45 samples)	5	11.1	22	48.9	39	86.7
healthy (51 samples)	30	58.8	45	88.2	49	96.1



(a)



(b)

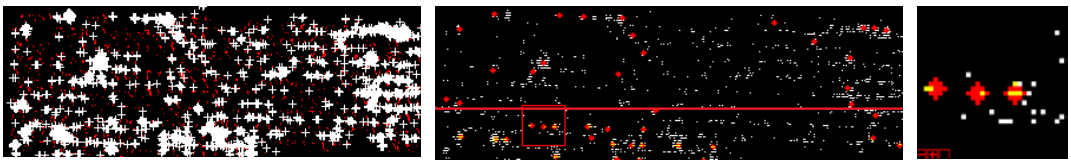
Fig. 8. REP value from (a) the training set, and (b) the validation set in the block.

Using the spectral library chosen by VCA, and the chosen angle based on the dataset, SAM was applied on the block, and the results are shown in Fig. 7. Since there were still too many false positives, which means a lot of healthy points in the image, especially the edge points of the trees, were classified as HLB infected, and a further analysis was needed. Based on the above feature analysis, the REP value was calculated for both the training and validation sets of the block, which can be seen in Fig. 8. 720 nm was chosen as the REP to filter out the false positive pixels. Table 3 shows the classification accuracy for the block data using REP.

Table 3. Classification accuracy by using REP technique.

Tree Category	Numbers of Training set (pixel)	Numbers of Validation set (pixel)	Training set (T)		Validation set(V)	
			Detected trees (pixel)	Percent (%)	Detected trees (pixel)	Percent (%)
hlb	23	22	15	65.2	19	86.3
healthy	26	25	24	92.3	23	92.0

The processed results can be seen in Fig. 9, after filtering the false positives using an REP of 720 nm. The accuracy was calculated from the ground truth and the detected results for HLB infected pixels and healthy pixels. The RMSE for geo-accuracy of the image acquisition system after geometrical calibration is 2 pixels, therefore a 5×5 pixel buffer window was chosen for the validation set, using positions of the validation set as the center of the window. The results are shown in Table 4.



(a) (b) (c)

Fig. 9. Results after using REP to filter out false positive pixels on the block HS image subset: (a) REP technique was applied to Fig. 8b. The white crosshairs are the HLB infected pixels left after using REP. (b) The validation and the training sets are marked using red points, which are separated by a red line. (c) The zoomed-in image of the area marked using a red square in (b). The yellow points are the intersection of the validation set and the classification results.

Results comparison of different methods

The classification results after applying different methods on the filtered 2011 HS image are shown in Table 4. The proposed ESAM method showed the highest detection accuracy of more than 80% in the training set, and 86.3% in the validation set. For another supervised method, MahaDist had lower accuracy both in the training and validation sets. And the unsupervised method K-means had the worst accuracy in the training set, and the same detection accuracy with MahaDist. Compared with the results by Li et al. (2012), the results using the proposed method had a great improvement in detection accuracy.

Table 4. Classification accuracy comparison after applying different methods on the block.

Classification method	Number of Trees	Training set (T)		Validation set (V)	
		Detected trees (pixel)	Percent (%)	Detected trees (pixel)	Percent (%)
Proposed ESAM method	Infected trees (T:23,V:22)	19	82.6	19	86.3
MahaDist		15	65.2	14	63.6
K-means		12	52.1	14	63.6

CONCLUSION

Using the HS image obtained in 2011, a SAM based method was developed to detect HLB disease, named 'extended spectral angle mapping (ESAM)'. The spectral feature of the healthy and the HLB infected citrus trees were analyzed based on the ground truth data and the HS image of the corresponding area. The reflectance difference and the REP characteristic demonstrated the promising application of HS image to detect HLB infected trees from the healthy ones.

The choice of spectral library was vital to the result of SAM classification. To build the spectral library needed in SAM algorithm, instead of using the average of all the training set, pure pixels were found using VCA. As the higher spatial quality of HS image data obtained in 2011, the REP characteristic, which was better than the one in the 2010 image (Li et al., 2012), was utilized in HS image data analysis.

A fairly high detection accuracy of 82.6% was achieved in the training set, and 86.3% in the validation set was achieved using the proposed ESAM method. The results were compared with two other methods, including one supervised method MahaDist, which was recommended by Li et al. (2012), and one unsupervised method K-means. Both of these methods yielded poorer results.

ACKNOWLEDGEMENT

This project was funded by the Citrus Research and Development Foundation, Inc. The authors would like to thank Ms. Ce Yang, Ms. Sherrie Buchanon, Mr. Anurag R. Katti, Mr. Alireza Pourreza, and Mr. Junsu Shin at the University of Florida for their assistance in this study. The authors also would like to thank China Scholarship Council for financial support.

REFERENCES

Cho, M. A., and A. K. Skidmore. 2006. A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method. Remote Sensing of Environment, 101 (2):181–193.

Collins, W., G. L. Raines, and F. C. Canney. 1977. Airborne spectrora diometer discrimination of vegetation anomalies over sulphide mineralisation—a remote sensing technique. Geological Society of America, 90th Annual Meeting Programs and Abstracts, 9 (7): 932 –933.

Collins, W. 1978. Remote sensing of crop type and maturity. Photogrammetric Engineering and Remote Sensing, 44:43–55.

Dawson, T. P., and P. J. Curran. 1998. A new technique for interpolating the reflectance red edge position. International Journal of Remote Sensing, 19 (11): 2133–2139.

Huang, W., D. W. Lamb, Z. Niu, Y. Zhang, L. Liu, and J. Wang. 2007. Identification of yellow rust in wheat using in-situ spectral reflectance measurements and airborne hyperspectral imaging. Precision Agriculture, 8 (4-5): 187–197.

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

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.

Nascimento, J. M. P., and J. M. B. Dias. 2005. Vertex component analysis: A fast algorithm to unmix hyperspectral data. IEEE Transactions on Geoscience and Remote, 43(4): 898–910.

Plaza, A., J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. C. Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni. 2009. Recent advances in techniques for hyperspectral image processing. Remote Sensing of Environment 113 (Supplement 1):110–122.

Savitzky, A., and M. J. E. Golay.1964.Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal. Chem., 36 (8): 1627–1639.

Smith, K. L., M. D. Steven and J. J. Collis. 2005. Plant spectral responses to gas leaks and other stresses. International Journal of Remote Sensing, 26 (18): 4067–4081.

Ustin, S. L., D. A. Roberts, J. A. Gamon, G. P.Asener and R. O. Green. 2004. Using Imaging Spectroscopy to Study Ecosystem Processes and Properties. BioScience, 54 (6): 523–534.

Qin, J., T. F. Burks, M. A. Ritenour, and W. G. Bonn. 2009. Detection of citrus canker using hyperspectral reflectance imaging with spectral information divergence. Journal of Food Engineering, 93 (2): 183–191.

Yang, C., J. H. Everitt, and J. M. Bradford. 2008. Yield Estimation from Hyperspectral Imagery Using Spectral Angle Mapper (SAM),Transactions of the ASABE, 51 (2): 729–737.

Yea, X., K. Sakaia, A. Sasaoa, and S. Asadac. 2008. Potential of airborne hyperspectral imagery to estimate fruit yield in citrus. Chemometrics and Intelligent Laboratory Systems, 90 (2):132–144.

Zhang, M., Z. Qin, X. Liu, and S. L. Ustin. 2003. Detection of stress in tomatoes induced by late blight disease in California. International Journal of Applied Earth Observation and Geoinformation, 4 (4):295–310.