

‘Extended spectral angle mapping (ESAM)’ for citrus greening disease detection using airborne hyperspectral imaging

Han Li · Won Suk Lee · Ku Wang · Reza Ehsani · Chenghai Yang

© Springer Science+Business Media New York 2013

Abstract Hyperspectral (HS) imaging is becoming more important for agricultural applications. Due to its high spectral resolution, it exhibits excellent performance in disease identification of different crops. In this study, a novel method termed ‘extended spectral angle mapping (ESAM)’ was proposed to detect citrus greening disease (Huanglongbing or HLB), which is a very destructive disease of citrus. Firstly, the Savitzky–Golay smoothing filter was used to remove spectral noise within the data. A mask for tree canopy was built using support vector machine, to separate the tree canopies from the background. Pure endmembers of the masked dataset for healthy and HLB infected tree canopies were extracted using vertex component analysis. By utilizing the derived pure endmembers, spectral angle mapping was applied to differentiate between healthy and citrus greening disease infected areas in the image. Finally, most false positive detections were filtered out using red-edge position. An experiment was carried out using an HS image acquired by an airborne HS imaging system, and a multispectral image acquired by the WorldView-2 satellite, from the Citrus Research and Education Center, Lake Alfred, FL, USA. Ground reflectance measurement and coordinates for diseased trees were recorded. The experimental results were compared with another supervised method, Mahalanobis distance, and an unsupervised method, K-means, both of which showed a 63.6 %

H. Li · K. Wang
Department of Information and Electrical Engineering, China Agricultural University,
Beijing 100083, China

W. S. Lee (✉)
Department of Agricultural and Biological Engineering, University of Florida, Gainesville,
FL 32611, USA
e-mail: wslee@ufl.edu

R. Ehsani
Citrus Research and Education Center, University of Florida, 700 Experiment Station Road,
Lake Alfred, FL 33850, USA

C. Yang
USDA-ARS, Southern Plains Agricultural Research Center, College Station, TX, USA

accuracy. The proposed ESAM performed better with a detection accuracy of 86 % than those two methods. These results demonstrated that the detection accuracy using HS image could be enhanced by focusing on the pure endmember extraction and the use of red-edge position, suggesting that there is a great potential of citrus greening disease detection using an HS image.

Keywords Image classification · Precision agriculture · Red-edge position · Spectral angle mapping · Vertex component analysis

Introduction

Citrus greening disease, also known as Huanglongbing (HLB), caused by the 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 the citrus industry in FL, USA (Smith et al. 2005; Huang et al. 2007; Lee et al. 2008; Qin et al. 2009). Once the tree is infected, it tends to die within 3–5 years. Compared with the healthy citrus canopy, HLB infected trees have several symptoms such as blotchy mottle and yellowish leaves, uneven fruits shape, reduced fruit size and severe fruit and leaf drop, as shown in Fig. 1a.

The bacteria associated with HLB, *Candidatus* Liberibacter asiaticus (Cia) was first detected in FL in 2005, which was only 7 years after the discovery of the psyllid vector in 1998. As of August, 2011, 37 counties with over 4 012 square-mile sections were infected in FL, USA, as shown in Fig. 1b (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). Timely and location-specific detection and monitoring of the

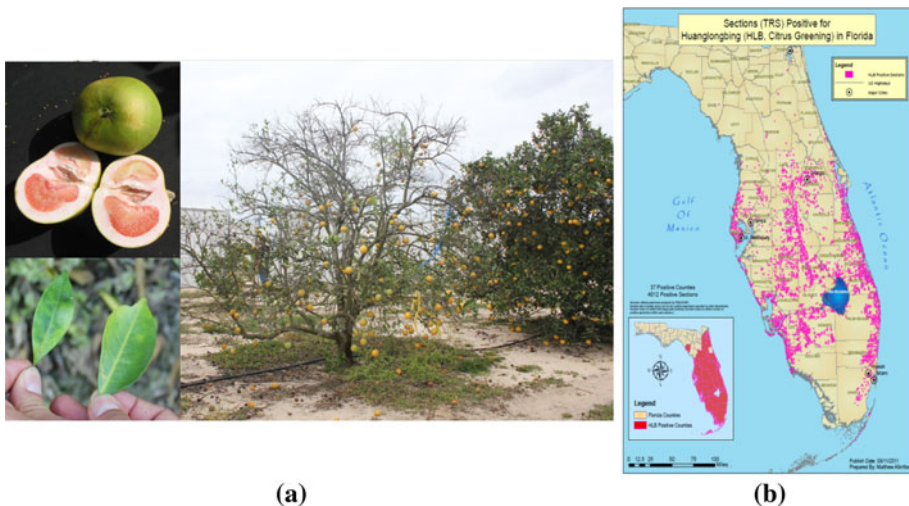


Fig. 1 HLB disease symptom and infected sections: **a** HLB symptom on the fruit, leaves and a whole tree; **b** HLB infected sections (marked in red) in FL, USA (Color figure online)

infected citrus trees are required for efficient disease control, in order to limit further infection.

Different remote sensing techniques, such as hyperspectral (HS) data from spectrometers, airborne and satellite HS and multispectral (MS) images (Ustin et al. 2004; Zhang et al. 2003; Ye et al. 2008; Plaza et al. 2009), are widely used for various agricultural applications. To satisfy the needs of different applications in obtaining land cover information over the past two decades, HS imaging has provided remarkable solutions due to its high spatial and spectral resolutions (Ustin et al. 2004). HS 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 including yield estimation, fruit detection, environmental impact assessment and crop disease detection (Zhang et al. 2003; Yang et al. 2008; Ye et al. 2008; Plaza et al. 2009). Compared with other expensive and time consuming HLB disease detection methods currently used, such as conventional ground scouting (Etcheberria et al. 2007), electron microscopy and bioassay (Chung and Brlansky 2005) and polymerase chain reaction (PCR), remote sensing can quickly collect citrus grove canopy data that can be used to analyze geo-temporal and geo-spatial properties of biological features of crops, including the symptoms of the citrus greening disease (Kumar et al. 2012). Thus, once infected areas and their severity are known from remote sensing data, the growers can concentrate their efforts only on infected areas without inspecting the whole grove. In addition, they don't need to conduct time-consuming and expensive PCR tests, let alone labor intensive and subjective ground inspection, which take a long time (several months) for ground crews to inspect the whole grove. More infection can happen during the inspection period.

A number of image processing techniques and analysis methods have been developed for HS images. Many different methods such as minimum noise fraction (MNF) (Green et al. 1988), spectral angle mapping (SAM) (Kruse et al. 1993), spectral feature fitting (SFF) (Clark et al. 1990), spectral information divergence (SID) (Du et al. 2004) and support vector machines (SVMs) (Nidamanuri and Zbell 2011) have been utilized for HS image classification.

Disease detection of various crops or citrus fruit utilizing HS data has been widely studied, which has become a subject of intensive research. Zhang et al. (2003) investigated the detection of stress in tomatoes induced by late blight disease in California using an HS image. They combined MNF and SAM methods, and reported that the late blight diseased tomatoes at stage three or above could be separated from healthy plants. Smith et al. (2005) found that in the spectral data, the red-edge position (REP) 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 a disease index. The regression equation was validated in a subsequent growing season, and then was applied to HS airborne imagery to discriminate and map the disease index in target fields. Qin et al. (2009) developed a SID based algorithms for HS 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 reported. Mewes et al. (2011) evaluated the suitability of Bhattacharyya distance (BD) (Bhattacharyya 1943) for feature selection, to identify bands within the feature space for efficient classification. The BD measures the similarity of two

discrete or continuous probability distributions. Using a forward feature search strategy, BD was used to select 13 spectral bands. The results exhibited that detection accuracy could be enhanced using a few relevant spectral bands instead of all the bands.

As an earlier study for detecting the citrus greening disease, Lee et al. (2008) used HS images by applying SAM and SFF methods. They reported that it was difficult to obtain good results because of the positioning errors in GPS ground truth and aerial imaging, and the spectral similarity between healthy and the citrus greening disease infected trees. Further, Kumar et al. (2012) used mixture tuned matched filter (MTMF), SAM and linear spectral unmixing (LSU) methods for HLB detection. In this study, a detection accuracy of 80 % was achieved using MTMF on a 2009 HS image, and SAM also yielded an accuracy of 87 % using MS images. However, accuracies of 60 and 66.6 % were obtained using SAM for two experimental sites, which indicated SAM did not perform well on the HS images. Li et al. (2012) used both ground and airborne remote sensing to find the spectral differences between HLB and healthy citrus canopies. Several commonly used classification and spectral mapping methods were implemented in airborne MS and HS images. Their performances and adaptability to detect HLB infected canopy in citrus groves were then compared and evaluated. The SFF showed the best results among all the methods, but the results from the 2007 HS image were not consistent with those from the 2010 HS image. Because of the low spatial quality of the HS image spectrum data obtained, classification using REP characteristics yielded unsatisfactory results.

The overall objective of this study was to develop a new and novel method to detect citrus greening disease using airborne HS and multispectral imaging based on ‘extended spectral angle mapping (ESAM)’, which was achieved from the following specific objectives:

1. to find the spectral differences between healthy and HLB infected canopies from ground measurement and HS images,
2. to develop a new method, termed ‘ESAM’, in order to effectively detect HLB infected trees using an airborne HS image,
3. to evaluate and compare results of the proposed ‘ESAM’ method with two other commonly used methods: K-means and Mahalanobis distance (MahaDist) in order to select the best method for HLB disease detection using HS image in the future and
4. to apply the three methods, SAM, K-means and MahaDist, to a MS image, and compare the performance of these methods for the disease detection, in order to evaluate the feasibility of detecting HLB infected trees using a MS image.

Materials

Airborne image acquisition

On December, 14, 2011, a set of airborne HS image was acquired for three citrus blocks of the Citrus Research and Education Center (CREC) grove, located in Lake Alfred, Central FL, USA.

An airborne hyperspectral camera unit—an AISA EAGLE VNIR Hyperspectral Imaging Sensor (Spectral Imaging, Ltd., Oulu, Finland) was used for acquiring images. The illumination condition when obtaining the image was the following: (1) cloud cover was 0 %, (2) acquisition time was 12:00–13:00 h local time, (3) solar elevation was larger than 35°, (4) flight altitude was 2 100 ft/640 m and (5) flight speed was 65 knots. One 3.6 m × 3.6 m airborne sensor ground reference tarp (type 822 fabrics, moderate weight

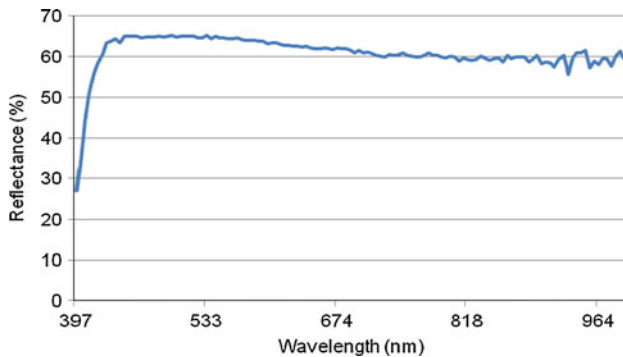


Fig. 2 Reflectance of the reference tarp, made of type 822 fabrics, which is moderate weight woven polyester substrate. The size of the tarp was 3.6 m by 3.6 m

woven polyester substrate) has been placed in the imaging area and spectrally covered during the HS data acquisition. The spectral characteristics of the tarp were well known and were implemented into a spectral library. After the image was obtained, an image spectrum from the airborne sensor was derived reflecting the environmental conditions during the imaging. The image was radiometrically calibrated with the tarp using the empirical line method (Smith and Milton 1999). Fig. 2 shows reflectance of the reference tarp extracted from the HS image. A total of 128 spectral bands between 400 and 1 000 nm were collected, which had the digital number (DN) ranging from 0 to 4 095. The spectral resolution was 5 nm. The HS image was georeferenced to the UTM coordinate system in zone 17 N with the datum of WGS-84. The spatial resolution, also called ground sampling distance, of the final image was 0.5 m.

Also on December 5, 2011, an MS image with eight spectral bands was acquired for the same CREC grove using the WorldView-2 (WV-2) satellite. WV-2 is the first high-resolution 8-band multispectral commercial satellite. Operating at an altitude of 770 km, WV-2 provides 46 cm panchromatic resolution and 1.85 m multispectral resolution. The four primary multispectral bands include traditional blue (450–510 nm), green (510–580 nm), red (630–690 nm), near-infrared (770–895 nm) bands, and four additional bands including coastal (400–450 nm), yellow (585–625 nm), red-edge (705–745 nm) bands and an additional longer wavelength near-infrared band (860–1 040 nm), which is sensitive to atmospheric water vapor.

Ground truth measurement

On December 14 and 15, 2011, ground truth measurements were conducted in the same grove at the CREC. Three blocks, 8a, 2b and 5c, were used as regions of interest (ROIs) for the experiment. In this experiment, two types of ground truth were measured: ground spectral reflectance and location for the measured trees, which were used to determine the infected position in the HS image. Ground spectral reflectance of each tree canopy was measured using a handheld spectrometer (HR-1024, Spectra Vista Corporation, Poughkeepsie, NY, USA). A white Spectralon 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). The tree infection status was determined by experienced ground inspection crews at the

Table 1 Description of different classes of citrus canopies used in this study

Class	Tree growth status	Description	Number of trees
8a-m-hlb	Mature	HLB infected canopy in block 8a	45
8a-m-healthy	Mature	Healthy canopy in block 8a	51
2b-i-hlb	Intermediate	HLB infected canopy in block 2b	66
2b-i-healthy	Intermediate	Healthy canopy in block 2b	66
5c-y-hlb	Young	HLB infected canopy in block 5c	28
5c-y-healthy	Young	Healthy canopy in block 5c	28

CREC grove in June 2011. Their inspection results were comparable to a PCR laboratory test, ~95 % of the time. Because of the large amount of trees in the ROIs, only the pre-determined HLB infected trees and some randomly chosen healthy trees were chosen for the experiment. The status of all the trees measured in the chosen blocks was rechecked by the ground inspection crews. In total, positions of 284 trees were collected in the CREC grove. The measured trees were classified into two classes: HLB infected and healthy trees. Depending on the tree growth status, they were divided into three different classes: mature, intermediate and young, as shown in Table 1.

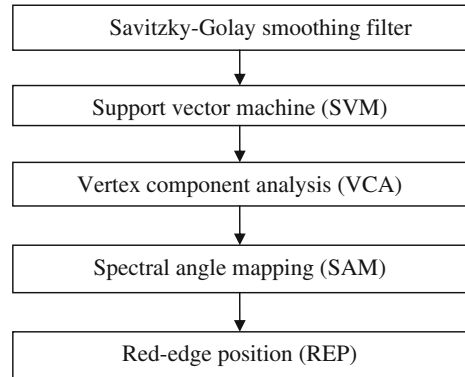
Methods

‘Extended spectral angle mapping (ESAM)’

Many data processing techniques have been developed and used for HS images considering the characteristics of HS datasets. In this study, a new and novel method named ‘ESAM’ was proposed to detect citrus greening disease, and its procedure is shown in Fig. 3. Firstly, the Savitzky–Golay smoothing filter was used to remove spectral noise within the data, yet to keep the shape and absorption features of the spectrum (Savitzky and Golay 1964). A mask for tree canopy was built using support vector machine classification method (SVM), to separate the tree canopies from the other background. Usually, a single pixel contains several substances, which are also called endmembers, in an HS image. The choice of pure endmember is very important for the result of many classification methods. Pure endmembers of the masked dataset for the healthy and the HLB infected classes were chosen using vertex component analysis (VCA), which has better performance compared to other spectral linear unmixing methods. By utilizing the derived pure endmembers from VCA, SAM was applied to classify healthy and citrus greening disease infected areas in the image. Finally, REP was used to filter out most of false positive detections.

The ESAM focused on solving the challenging problem, which is to identify HLB infected citrus trees from other healthy trees. The median detection accuracy for HLB from HS images by Li et al. (2012) ranged 50–70 % using classification methods, such as Parallelepiped, minimum distance (MinDist), MahaDist, SAM, SID, MTMF and SFF. Compared with these methods, which did not perform very efficiently on this problem,

Fig. 3 Diagram of the newly proposed ESAM method



ESAM aimed at building a standard procedure for detecting HLB infected trees. The experimental results suggested that ESAM improved the accuracy greatly compared with the previous work.

The Exelis Visual Information Solutions (ENVI, Inc., Boulder, Colorado, USA) software was used for the HS image analysis. Using the RTK GPS data obtained from the ground truth, the HS image spectrum data were exported from the corresponding position. Each step of ESAM is illustrated below using dataset in block 8a. The citrus variety of this block was Hamlin. The statuses of all the trees measured were determined by an experienced inspection crew of CREC. From a sample set including 51 healthy and 45 HLB samples, a subset of 26 healthy and 23 HLB infected pixel spectra was randomly chosen to form a calibration set. The rest of the samples, including 25 healthy and 22 HLB infected pixel spectra formed a validation set. To evaluate the proposed ESAM method, MahaDist and K-means method were also tested along with the HS image of all three blocks. To evaluate the feasibility of detecting HLB infected trees using a MS image, SAM, MahaDist and K-means method were tested also with the MS image of block 8a.

Savitzky–Golay smoothing filter

The Savitzky–Golay smoothing filter was chosen to smooth the acquired HS image (Savitzky and Golay 1964). The filter is defined as a weighted moving average with weighting given as a polynomial of a certain degree. When applied to a spectrum, the returned coefficients perform a polynomial least-squares fit within the filter window. It keeps the shape of reflectance and absorption features of the spectrum, but removes spectral noise within the data. In this study, a second order polynomial and five data points were used as a Savitzky–Golay smoothing filter and degree of smoothing polynomial was chosen as two. These parameters were determined by a trial and error using the calibration set. The ENVI software was used for Savitzky–Golay smoothing filter.

Support vector machine (SVM)

Instead of substance based classification of the land cover, the objective of this study was to classify healthy and infected parts of the same substance, citrus trees. This is more challenging, as it requires the classification methods to be stricter. To simplify the problem, a background mask is needed, avoiding any influence from background objects such as grass, shadow and bare ground. Li et al. (2012) observed that SVM was an effective and

quick method to build a mask for the trees. This method was performed in ENVI. Firstly, five calibration ROIs were randomly chosen from the smoothed HS calibration image for five different objects, which were tree, grass, shadows, white sand and reddish soil. Using these ROIs as inputs, SVM method was performed and pixels were classified. Based on the classification result, a mask for the tree class was built and other background pixels were removed. In this study, each ROI of the five classes included 100 pixels. Radial basis function was chosen as kernel type for SVM, 0.008 was chosen as a gamma value, and 100 was chosen as a penalty parameter in this function based on preliminary trials.

Spectral angle mapping (SAM)

The SAM determines spectral similarity between two spectra by treating them as vectors in a space with dimensionality equal to the number of bands (nb) (Kruse et al. 1993). A simplified explanation of this can be given by considering a reference spectrum and a test spectrum from two-band data represented on a two dimensional plot as two points, as shown in Fig. 4.

The angle between a test spectrum a , and a reference spectrum b , can be calculated by Eq. (1), which can also be written as Eq. (2). The angle (α) between the vectors is the same regardless of their length, which can eliminate illumination effects in the different environments.

$$\alpha = \cos^{-1} \left(\frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \cdot \|\vec{b}\|} \right) \quad (1)$$

$$\alpha = \cos^{-1} \left(\frac{\sum_{i=1}^{nb} a_i b_i}{\left(\sum_{i=1}^{nb} a_i^2 \right)^{1/2} \left(\sum_{i=1}^{nb} b_i^2 \right)^{1/2}} \right), \quad (2)$$

where nb equals to the number of bands.

Vertex component analysis (VCA) endmember extraction

The VCA is a linear unmixing method, aiming at estimating the number of reference endmembers, their spectral signatures and their abundance fractions. Fig. 5 illustrates the

Fig. 4 Schematic plot of SAM algorithm for a two-band image: the angle (α) between a test spectrum a , and a reference spectrum b , was used to determine the similarity of the two spectrum (adopted from Kruse et al. 1993)

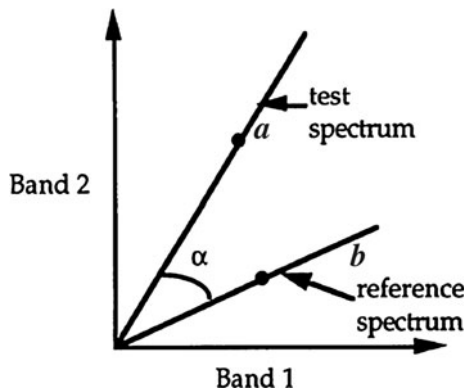
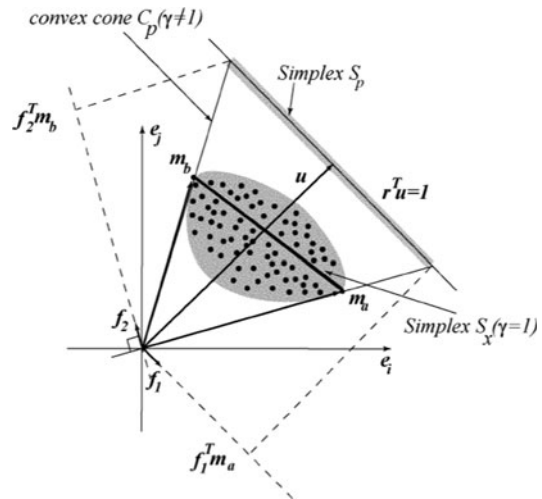


Fig. 5 Illustration of VCA algorithm: the algorithm iteratively projects data onto a direction orthogonal to a subspace spanned by the endmembers already determined. The new endmember signature corresponds to the extreme of the projection (adopted from Nascimento and Dias 2005)



principle of VCA algorithm. The algorithm iteratively projects data onto a direction orthogonal to a subspace spanned by the endmembers already determined. The new endmember signature corresponds to the extreme of the projection. As shown in Fig. 5, in the first iteration, data are projected onto the first direction f_1 . The extreme of the projection corresponds to endmember m_a . In the next iteration, endmember m_b is found by projecting data onto direction f_2 , which is orthogonal to m_a . The algorithm iterates until all endmembers are exhausted.

Pixel purity index (PPI) (Boardman 1993) and the N-finder (N-FINDR) algorithm (Winter 1999) are two of the most popular approaches which have been developed. The PPI is based on the geometry of convex sets. The N-FINDR method is an automated approach that finds a set of pixels which defines the simplex with the maximum volume, potentially inscribed within the dataset. Nascimento and Dias (2005) reported that VCA performed better than PPI and better than or similarly to N-FINDR. The VCA also had the lowest computational complexity among these three methods. Therefore, VCA was chosen in this study to extract pure endmembers of the dataset.

Red-edge position (REP)

The red-edge represents the abrupt reflectance change in 680–740 nm range of vegetation spectra, which is caused by the combined effects of strong chlorophyll absorption in the red wavelengths and high reflectance in the near-infrared (NIR) wavelengths due to the leaf internal scattering. The point of maximum slope is termed the REP, which is also the maximum FDR of the reflectance spectrum (Collins et al. 1977; Collins 1978). Due to the fact that double-peak feature exists in the FDR of the reflectance spectrum, several techniques have been developed to determine REP, including linear four-point interpolation, high-order polynomial fitting, inverted Gaussian fitting methods, Lagrangian technique and the linear extrapolation method (Dawson and Curran 1998; Cho and Skidmore 2006). Based on the fact that linear extrapolation method is as simple as the linear four-point interpolation, yet performs better than the other methods (Cho and Skidmore 2006), it was chosen for calculating REP of healthy and HLB infected dataset.

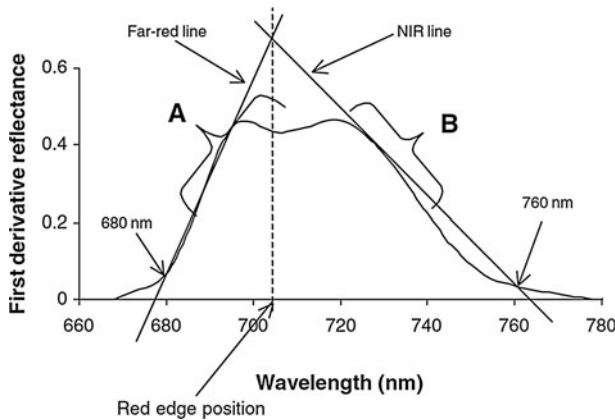


Fig. 6 Schematic representation of the linear extrapolation technique for extracting the red-edge position (REP), wavelength of the meeting point between two *straight lines* extrapolated on the far-red and NIR flanks of the first derivative spectrum (adopted from Cho and Skidmore 2006)

The FDR was calculated using a first-difference transformation of the reflectance spectrum as shown in Eq. (3).

$$FDR_{(\lambda,i)} = (R_{\lambda(j+1)} - R_{\lambda(j)}) / \Delta\lambda, \quad (3)$$

where FDR is the FDR reflectance at wavelength i , which is midpoint between wavebands j and $j + 1$, $R_{\lambda(j)}$ is the reflectance at the j waveband, λ is wavelength and $\Delta\lambda$ is the difference in the wavelength between j and $j + 1$.

Four coordinate points were used to calculate the REP by the linear extrapolation method, including two points on the far-red (680–700 nm) band and two points on the NIR (725–760 nm) region. The method is based on the linear extrapolation method of the two straight lines through the four points. The REP is defined by the wavelength value at the intersection of the two straight lines in Fig. 6.

$$\text{Far-red line : } FDR = m_1\lambda + c_1 \quad (4)$$

$$\text{NIR line : } FDR = m_2\lambda + c_2, \quad (5)$$

where m and c represent the slope and intercept of the straight lines.

At the intersection, the two lines have equal λ wavelength, and FDR values. Therefore, REP is given by Eq. (6), which is defined as the wavelength at the intersection

$$REP = \frac{-(c_1 - c_2)}{(m_1 - m_2)}. \quad (6)$$

Results

Spectral feature analysis

For spectral feature analysis, the ground truth and HS data for HLB infected and healthy trees from block 8a were used. Although the ground hyperspectral measurements had a spectral range of 348–2 505 nm, only the spectra ranging from 400 nm to 1 000 nm were

used in this study for a better comparison with the HS image spectral data having the same wavelength range. Since brightness conditions of each leaf were different due to illumination change when the ground measurements were conducted, the mean and standard deviation (Std) spectra of different classes were calculated to identify characteristics of the two classes. The Std shows how much variation exists from the mean value.

Spectral feature from ground truth and the HS image

From the ground measurements, two average class spectra (healthy and HLB) from 9 healthy and 17 HLB infected samples in 400–1 000 nm and their first derivatives are plotted in Fig. 7. The numbers in parentheses indicate the number of samples used for averages and standard deviations are shown in bars. The analysis results of spectral data from 51 healthy and 45 HLB infected samples, extracted from the HS image are shown in Fig. 8.

From Figs. 7a and 8a, obvious reflectance difference is shown in both ground truth data and the HS image. In Fig. 7a, 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. 8a, 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 700–1 000 nm is much higher than that of the HLB infected samples. The difference between Figs. 7 and 8 was because they were from different instruments and the graphs were from the average of a small number of samples. The result from Fig. 7 is consistent with the result described by Lee et al. (2008). From Figs. 7b and 8b, the obvious FDR difference is shown in both ground truth data and the HS image, especially in the red-edge from 680 to 740 nm. A zoomed-in subplot in the 680–760 nm is shown in the upper right corner in Figs. 7b and 8b. The mean FDR of the healthy samples in this range is much higher than that of the HLB infected samples.

Preprocessing of the MS image

The MS image taken by the WV-2 satellite had a 1.85 m spatial resolution. Part of the image is shown in Fig. 9a, and the zoomed-in image of the red square area in Fig. 9a is shown in Fig. 9b. The different classes of substances on the ground cannot be separated even with human eyes and it would be difficult to apply SVM or other classification methods on the image. With the 46-cm panchromatic resolution image also acquired using WV-2, pan-sharpening was conducted to get a better spatial resolution image by merging a high-resolution panchromatic and a lower resolution multispectral images. The bilinear resampling method (Parker et al. 1983) was chosen in ENVI to complete this procedure. The output pixel value in the bilinear resampling method is the weighted average of the four closest input pixel values. Then, the spatial resolution of multispectral image was improved to 0.5 meter, as shown in Fig. 9c. The zoomed-in image of the red square area in Fig. 9c is shown in Fig. 9d. The crosshairs in the image are the ground truth locations of HLB infected trees. To some extent, this is an exploration to see if some cost could be saved if the results of satellite MS image and the airborne image have comparable results.

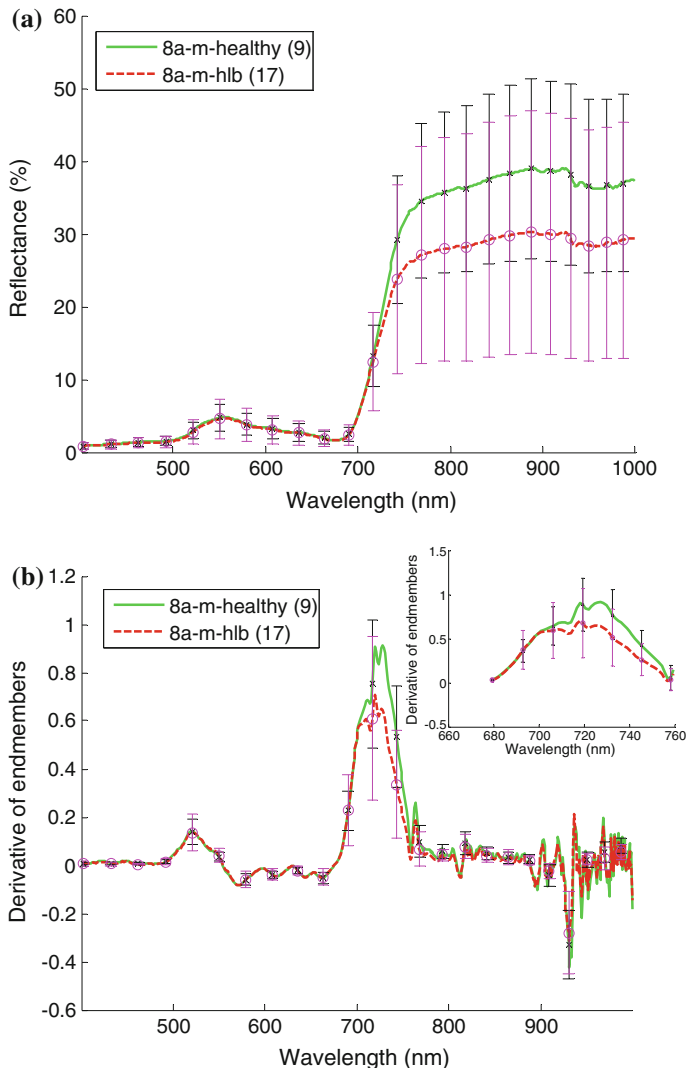


Fig. 7 Spectral feature analysis of ground truth data, using mean value and standard deviation shown in bars: **a** mean field spectra for samples in block 8a; **b** mean value of first derivative of field spectra for samples in block 8a. The numbers in parentheses indicate the number of spectra used for calculating averages

Spectral feature analysis for the MS image

In Fig. 10, the solid green line is an average spectra of 51 healthy samples, and the red dotted line is an average spectra of 45 HLB infected samples from the MS image, including Std shown in bars.

Band 2 (blue, 450–510 nm), band 3 (green, 510–580 nm), band 4 (yellow, 585–625 nm), band 5 (red, 630–690 nm), band 6 (red-edge, 705–745 nm) and band 7 (near-infrared, 770–895 nm) all have obvious difference between the two classes. The

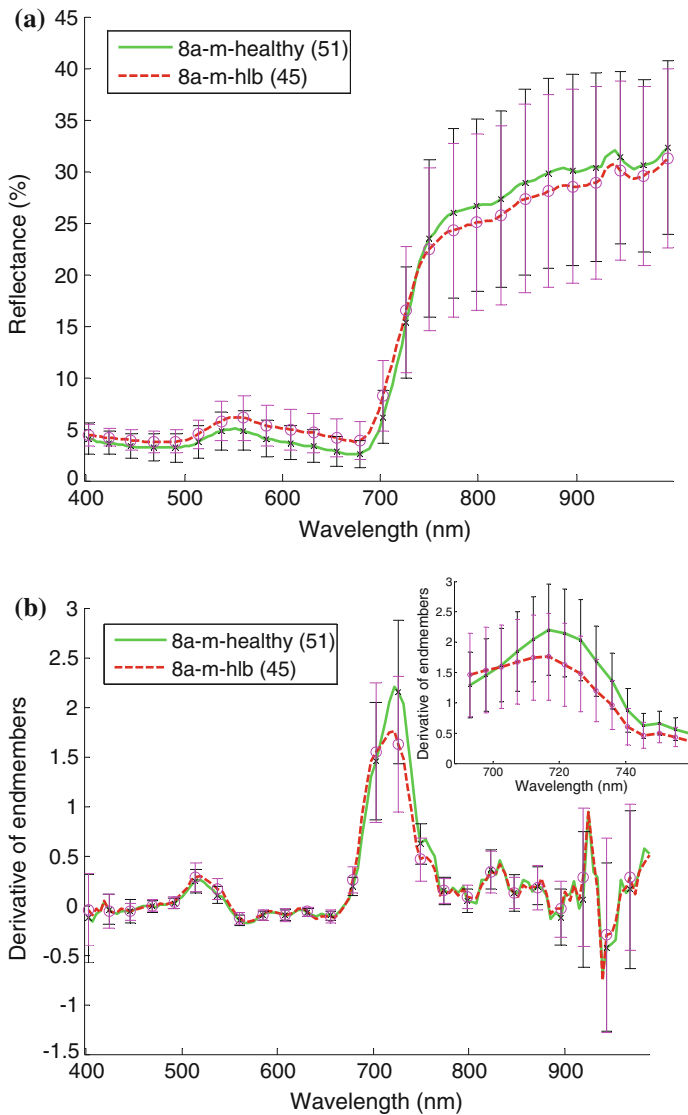


Fig. 8 Spectral feature analysis of the airborne HS image data, using mean value and Std shown in bars: **a** mean HS image spectra for healthy and HLB infected samples in block 8a; **b** mean value of first derivative of the HS image for samples in block 8a. The numbers in parentheses indicate the number of spectra used for calculating averages

difference in the visible range (bands 1 through 6) was mainly caused by the decrease of chlorophylls because of the HLB disease. Meanwhile, due to the disease, the inner cellular structure was damaged, which caused the lower reflectance in bands 7 and 8 (Li et al. 2012).

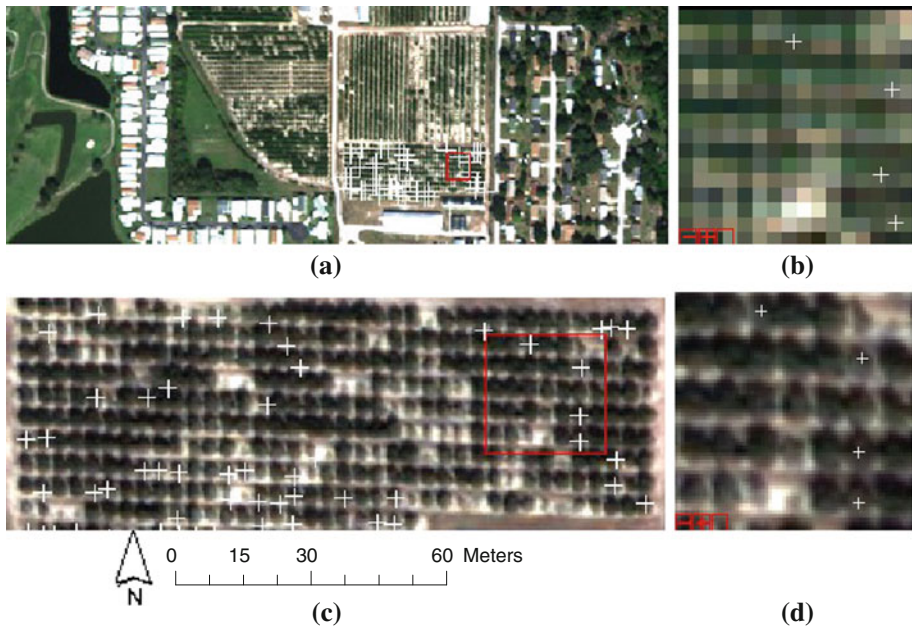
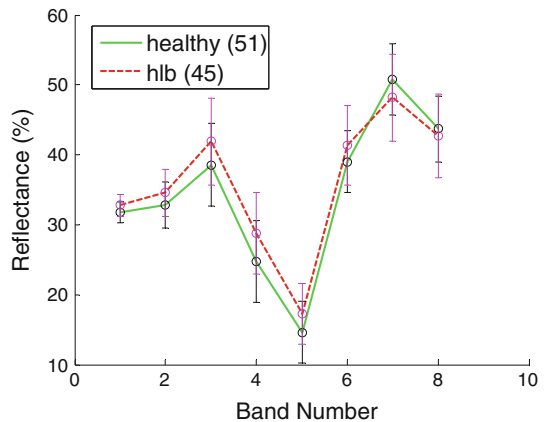


Fig. 9 Preprocessing of the MS image: **a** part of the MS image of CREC taken by WorldView-2; **b** zoomed-in image of the red square area in (a); **c** MS image after pan-sharpening; **d** zoomed-in image of the red square area in (c)

Fig. 10 Spectral feature analysis of the MS image data for healthy and HLB infected samples in block 8a, using mean value and standard deviation



Results of ESAM for HS image

After the Savitzky–Golay smoothing filter was applied, SVM was performed on the block and a mask was obtained based on the tree class, including HLB infected and healthy trees. A mask for the tree canopy was built and applied to the image. The result is shown in Fig. 11. The red line in Fig. 11a shows the boundary of the calibration and validation sets.

The calibration set was used to find pure pixels for the two classes using VCA. Two pure endmembers were selected successfully, and are shown in Fig. 12. Their spectral

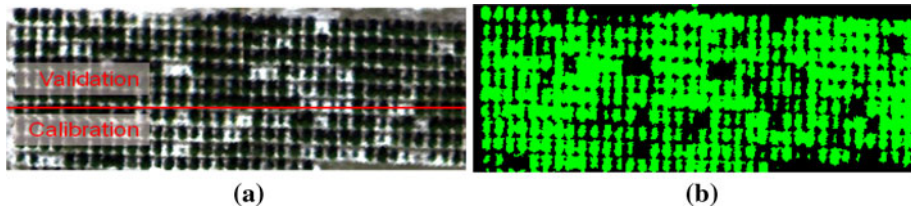


Fig. 11 SVM classification and masked results: **a** original HS true color image of the block. The validation and the calibration regions in the image are separated by a red line; **b** image with tree class only after masking (Color figure online)

features were consistent with those analyzed previously. The solid green line is the 5th sample selected among the 26 healthy samples in the calibration set. The dotted red line is the 13th sample selected from the 23 HLB infected samples in the calibration set.

In the visible range, the reflectance of the healthy sample is lower than that of the HLB infected sample, while the reflectance of the healthy sample in 700–1 000 nm is much higher than that of the HLB infected sample. These pure pixel spectra were used as a spectral library to carry out SAM. For any unknown input pixel, two spectral angles were calculated, including one with the healthy pure endmember and the other with the HLB infected pure endmember. Experimental studies have shown that low leaf chlorophyll concentration is associated with REP values near 700 nm, while high chlorophyll concentration in combination with leaf internal scattering influence REP values near 725 nm. The difference between HLB and healthy samples before and after 700 nm was mainly caused by the decrease of chlorophylls with the presence of the HLB disease, which is why reflectance properties of HLB and healthy samples change near 700 nm.

When SAM was applied to the masked image, a threshold was needed as an input parameter, which would be very important for classification. If the value was too high, false positives would be introduced. If it was too low, the image would be over-classified. To choose a proper threshold, spectral angles between each data in the calibration set and the pure endmembers chosen by VCA were calculated as shown in Fig. 13a (with healthy pure endmember) and 13b (with HLB infected endmember). From these figures, spectral angles of each sample are very different when the healthy pure sample is used and when the HLB infected pure sample is used. Thus, two separate thresholds should be chosen for the identification of HLB infected and healthy samples, based on the spectral angles between the training dataset and the chosen endmembers. The higher the threshold, the higher the accuracy for classification as observed in Fig. 13. However, the higher the

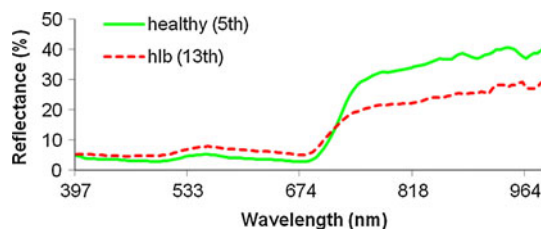


Fig. 12 Pure endmember spectra chosen by VCA. The solid green line is the 5th sample selected among the 26 healthy samples in the calibration set. The dotted red line is the 13th sample selected from the 23 HLB infected samples in the calibration set (Color figure online)

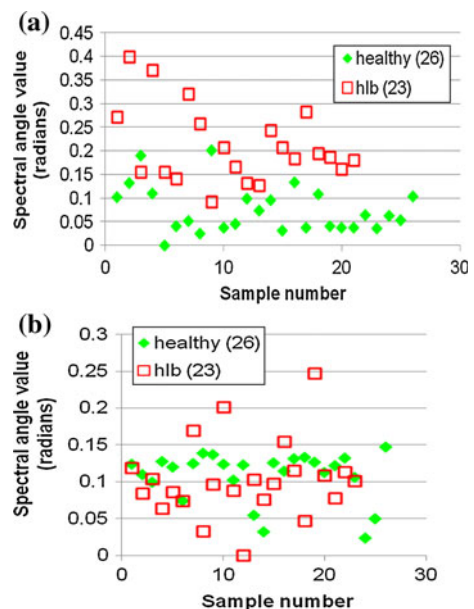
threshold, the higher the misclassified sample number is too. Thus, multiple maximum spectral angles were chosen based on the processed results. A trade-off should be made to get better detection result, yet not induce too many false positives. A spectral angle of 0.15 was chosen for the healthy class. A spectral angle of 0.1 was chosen for the HLB infected class for the HS image analysis, since a spectral angle of 0.15 produced many false positives based on preliminary testing.

Using the spectral library chosen by VCA and the chosen spectral angle based on the dataset, SAM was applied to the block and the results are shown in Fig. 14. Red and green pixels are infected and healthy areas, respectively. The white crosshairs indicate HLB infected tree canopy locations identified by the ground truth. Since there were still too many false positives, which means a lot of healthy pixels in the image, especially the edge points of the trees, were classified as HLB infected, a further analysis was needed.

Based on the above feature analysis, an REP value was calculated for the calibration and validation sets of the block, as shown in Fig. 15. Based on the training dataset (Fig. 15a), 720 nm was tested to yield the best classification results for HLB infected and healthy samples. Thus, a wavelength of 720 nm was determined to be a threshold to filter out the false positives. Table 2 shows the classification accuracy for the block using REP.

The processed results are shown in Fig. 16, after filtering the false positives using an REP of 720 nm. The accuracy was calculated from the ground truth and the detected results for healthy and HLB infected pixels. The RMSE for geo-accuracy of the image acquisition system after geometric correction was 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. If there are any detected pixels (red pixels) inside the buffer, it is considered as a correct identification of the disease. The proposed method was applied to different blocks, the trees of which were in different maturity status, and the detection results are shown in Table 3.

Fig. 13 Spectral angle value between each data and the pure endmembers: **a** spectral angle value between the training dataset and the healthy pure endmember chosen by VCA and **b** spectral angle value between the training set and the HLB infected pure endmember chosen by VCA



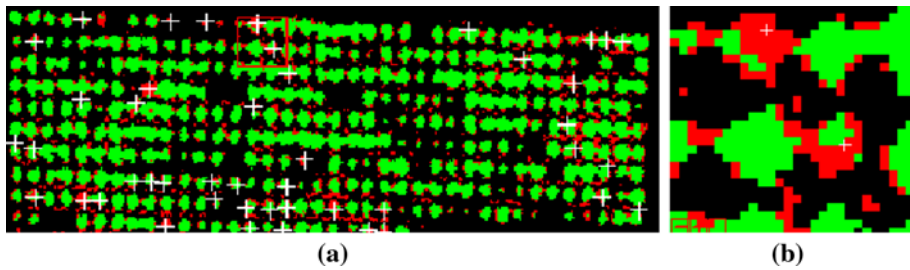


Fig. 14 SAM results applied to the block, *red pixels* are infected areas and *green pixels* are healthy areas: **a** SAM results with spectral angle of 0.1 for the HLB infected pixels and 0.15 for healthy pixels. The *white crosshairs* indicate HLB infected tree canopy locations identified by the ground truth; **b** zoomed-in image of the area marked using a *red square* in (a) (Color figure online)

Fig. 15 REP value from: **a** calibration set; and **b** validation set in the block

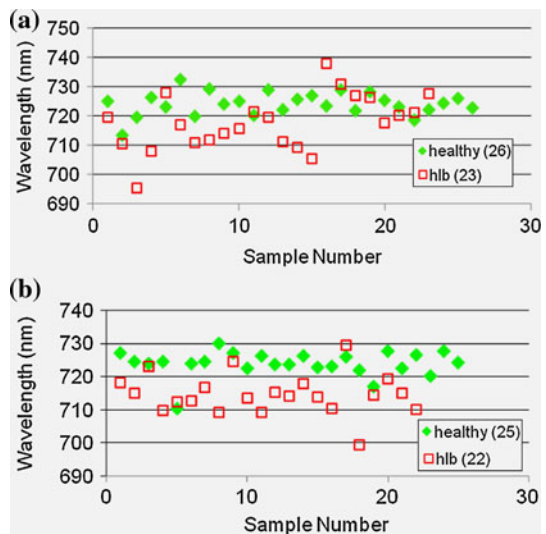


Table 2 Classification accuracy using REP

Tree class	Number of calibration set	Number of validation set	Calibration set (C)		Validation set (V)	
			Detected trees	Percent (%)	Detected trees	Percent (%)
HLB	23	22	15	65.2	19	86.3
Healthy	26	25	24	92.3	23	92.0

Comparison of results of HS image using different methods

The classification results after applying different methods to block 8a are shown in Table 4. The proposed ESAM method yielded the highest detection accuracy of 82.6 % for the calibration samples and 86.3 % for the validation samples. For another supervised method, MahaDist had lower accuracy both in the calibration and validation sets. The

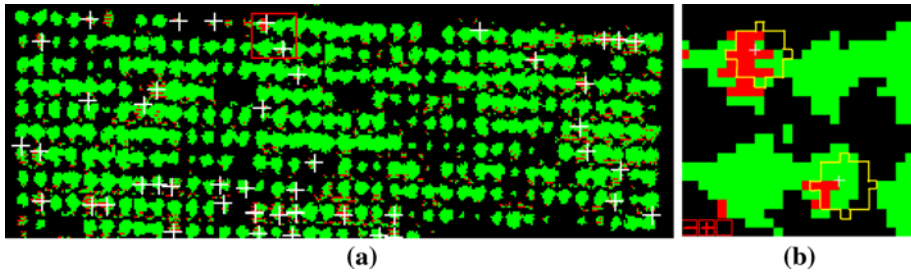


Fig. 16 Results after using REP to filter out false positive pixels on the block: **a** REP was applied to Fig. 14a. The crosshairs indicate HLB infected tree canopy locations identified by the ground truth; **b** zoomed-in image of the area marked with a red square in **a**. Region around the white crosshairs with yellow boundaries is the buffer region. If there are any detected pixels (red pixels) inside the buffer, it is considered as a correct identification (Color figure online)

Table 3 Classification accuracy of ESAM for different blocks

Classification method	Block	Number of trees (pixel)	Calibration set (C)		Validation set (V)	
			Detected trees (pixel)	Percent (%)	Detected trees (pixel)	Percent (%)
ESAM	8a (mature trees)	Infected trees (C:23, V:22)	19	82.6	19	86.3
	2b (intermediate trees)	Infected trees (C:14, V:14)	9	64.2	10	71.4
	5c (young trees)	Infected trees (C:33, V:33)	24	72.7	21	63.6

Table 4 Classification accuracy comparison after applying different methods to block 8a

Classification method	Number of trees (pixel)	Calibration set (C)		Validation set (V)	
		Detected trees (pixel)	Percent (%)	Detected trees (pixel)	Percent (%)
ESAM	Infected trees (C:23, V:22)	19	82.6	19	86.3
MahaDist		15	65.2	14	63.6
K-means		N/A	N/A	14	63.6

unsupervised method K-means had the same detection accuracy with MahaDist in the validation set.

Results of the disease detection using MS image

Using the calibration and validation dataset in the block 8a, three methods, including SAM, MahaDist and K-means were applied to the MS image. The results are shown in Table 5. Because the original spatial resolution of the MS image was 1.85 m and the resolution of panchromatic image was 0.46 m, four points in the pan-sharpened MS image were one pixel in the original MS image. Thus, a 9 by 9 ($= (4 + 1 + 4)$ by $(4 + 1 + 4)$) pixel buffer

Table 5 Classification accuracy comparison after different methods applied to the MS image on block 8a (number of infected trees: 23 for calibration and 22 for validation)

Classification methods	Calibration set (C)		Validation set (V)	
	Detected trees (pixel)	Percent (%)	Detected trees (pixel)	Percent (%)
SAM	18	78.2	14	63.6
MahaDist	15	65.2	13	59.1
K-means	N/A	N/A	9	40.9

window was chosen for the validation set to get the classification accuracy for the MS image.

The results using the MS image were lower than those from the HS image, since this study was targeted on the image segmentation for the same vegetation with stricter classification criteria, and also more information of the vegetation was needed.

Discussion

Spectral features of the original reflectance and FDR for both healthy and HLB infected samples obtained in 2011 were analyzed. Kumar et al. (2012) found that HLB showed higher reflectance in the visible region than healthy trees using field measurement data, which was consistent from the conclusions drawn by Li et al. (2012). And in this study, a similar conclusion was drawn: in the visible range (400–700 nm), the mean reflectance of the healthy samples was lower than that of the HLB infected samples, while the mean reflectance of the healthy samples in 700–1 000 nm was much higher than that of the HLB infected samples.

Along with the results from these research, if airborne images or ground spectral data were acquired under ideal environmental conditions, HLB infected canopy would have higher reflectance in the visible range, and lower reflectance in the NIR range than healthy canopy. The ideal environment conditions usually means that image acquisition should occur between 12 and 2 pm to guarantee enough sunshine along with no cloud cover, and little or no wind. But the spectral value could be easily affected by the environmental conditions such as the time of measurement, solar radiation, or cloud cover. Especially when a passive spectrometer is used rather than an active one, notable error could be introduced if spectral calibration frequency did not catch the radiation change or the angle between solar incident light and object surface varied greatly. For example, in our study, three measurements were made for one leaf sample when the ground spectral data were collected. The data obtained could be affected if there were cloud cover in between the three measurements, which may reduce the incident energy on leaf surface and thereby reduce signal to noise ratio. Then different conclusions could be drawn.

The proposed ESAM had a pixel based accuracy of 86.3 % for block 8a in the HS image in CREC. This result was much better than those by Kumar et al. (2012), the highest accuracy of which for HS image was 66.6 %, and those by Li et al. (2012), the highest accuracy of which for HS image was 61.9 %. The major improvement by the proposed ESAM method was resulted from the extension of the traditional SAM by utilizing the Savitzky–Golay smoothing filter, SVM, VCA for the pre-processing and utilizing REP for post-processing. As mentioned in the methods section, the Savitzky–Golay smoothing filter

could remove the spectral noise, while keeping absorption features. This makes sure that the spectrum noise was removed and was helpful for the subsequent steps. In addition, SVM was used as an effective alternative to build a mask for separation of tree canopy and the background.

The biggest advantage of ESAM compared with traditional SAM was the application of VCA and REP. Since selection of pure endmember of the two classes was very important as this was the reference of SAM, using the average of the samples as the endmember was not the best choice. The VCA was chosen as a spectral linear unmixing method because it worked better than PPI and N-FINDR. The use of REP also helped greatly for improving the results because it utilized the inherent characteristics of the vegetation. The choice of the parameters was also important: 0.1 and 0.15 were selected as the angles for HLB and healthy classes for SAM classification, respectively. A wavelength of 720 nm was chosen to be the threshold of REP to filter out false positives after applying SAM classification.

The accuracy results with citrus groves of different growth stages are shown in Table 3. The results of block 5c with young trees and block 2b with trees of intermediate growth stages were poorer compared with that in block 8a with mature trees. This reveals that the difference of tree growth stages could affect the performance of the proposed method. This is probably because the younger the trees, the smaller the HLB infected areas. In this case, each tree occupied fewer pixels in blocks 5c and 2b in the HS image than block 8a, less than five pixels per tree in some cases, which indicated that one single pixel represented more than one-fifth of a tree. Thus the probability of a single pixel having mixed information of both healthy and HLB infected canopies could increase, which would affect the result tremendously. In addition, some spectral differences among mature, young, and intermediate citrus canopies were also found in the preliminary processing results. This difference might be caused by the difference of the canopy volume of the citrus trees, because the reflectance could be influenced by the leaf density of one pixel or neighboring pixels. The influence of the citrus maturity on HLB detection will need to be studied in the future.

Compared with the results obtained by other methods, the proposed ESAM method yielded the highest detection accuracy of more than 80 % in the calibration set and 86.3 % in the validation set. Compared with the study by Li et al. (2012), the results using the proposed ESAM method had a good improvement in detection accuracy. This reveals the great potential of citrus greening disease detection using the ESAM method. Since it was not possible to conduct PCR confirmation for all the trees in the image, the false positive detection rate could not be calculated in this study. If pixel based detection was needed, more accurate ground work needs to be done.

For the MS image, most of the tree canopies under shadow were misclassified since too few bands' information was available. The HS image was proved to be more suitable to detect citrus greening disease compared with the MS image.

For implementing the findings for commercial use, a more automated software system would be needed, so that it can be easily implemented by end users or growers. Every grove has different conditions, and so a more adaptable system would be necessary also. The findings from this study could be used as a basis of such system.

Conclusion

Spectral features of healthy and HLB infected citrus trees were analyzed based on the ground truth data, HS and MS images of the corresponding area. The reflectance difference

and the REP characteristic demonstrated the promising application of the HS image to detect HLB infected trees in a grove. Based on the investigation conducted in this study, some major findings were summarized as follows.

- (1) A new method, 'ESAM' was developed to detect HLB disease. A fairly high detection accuracy of 82.6 % was achieved in the calibration set and 86.3 % in the validation set was achieved using the proposed ESAM method for the HS image.
- (2) For pre-processing, pure endmembers were extracted using VCA, which was vital to the result of SAM classification. For post-processing, an REP of 720 nm was chosen to filter out false positives based on the analysis of the dataset, which was proved to be efficient.
- (3) The results from this study revealed that the difference of tree maturity status could affect the outcome of the classification method. The results were compared with two other methods, including a supervised method MahaDist and an unsupervised method K-means. The proposed ESAM method yielded better results than both of these methods.
- (4) SAM, MahaDist and K-means methods were also applied to the MS image, resulting in poorer results than using the HS image. The promising application of the HS image was demonstrated to detect HLB disease.

Acknowledgments This project was funded by the Citrus Research and Development Foundation, Inc. The authors would like to thank Ms. Ce Yang, Ms. Sherrie Buchanon, Ms. Liubov Polonik, 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 the China Scholarship Council for financial support.

References

- Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by probability distributions. *Bulletin of the Calcutta Mathematical Society*, 35, 99–109.
- Boardman, J. W. (1993). Automating spectral unmixing of AVIRIS data using convex geometry concepts (vol. 1). In *Summaries 4th Annual JPL Airborne Geoscience Workshop*, pp. 11–14.
- 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. Retrieved March 21, 2013.
- Cho, M. A., & Skidmore, A. K. (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.
- Chung, K. R., & Brlansky, R. H. (2005). *Citrus diseases exotic to Florida: Huanglongbing (citrus greening)* (p. 210). Gainesville: IFAS Extension Publication, University of Florida.
- Clark, R. N., Gallagher, A. J., & Swayze, G. A. (1990). Material absorption band depth mapping of imaging spectrometer data using the complete band shape least-squares algorithm simultaneously fit to multiple spectral features from multiple materials (pp. 176–186). In *Proceedings of the Third Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Workshop*, JPL Publication 90-54.
- Collins, W. (1978). Remote sensing of crop type and maturity. *Photogrammetric Engineering and Remote Sensing*, 44, 43–55.
- Collins, W., Raines, G. L., & Canney, F. C. (1977). Airborne spectroradiometer 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.
- Dawson, T. P., & Curran, P. J. (1998). A new technique for interpolating the reflectance red edge position. *International Journal of Remote Sensing*, 19(11), 2133–2139.
- Du, Y., Ives, R., Chang, C.-I., Ren, H., Chang, C.-C., Jensen, J., et al. (2004). A new hyperspectral discrimination measure for spectral similarity. *Optical Engineering*, 43(8), 1777–1786.

- Etcheberria, E., Gonzalez, P., Dawson, W., & Spann, T. (2007). An iodine-based starch test to assist in selecting leaves for HLB testing. Gainesville: IFAS Extension Publication, University of Florida, HS-1122.
- FDACS/DPI. (2011). Sections (TRS) Positive for Huanglongbing (HLB, Citrus Greening) in Florida. Available: <http://www.freshfromflorida.com/pi/chrp/greening/StatewidePositiveHLBSections.pdf>. Retrieved March 21, 2013.
- Green, A. A., Berman, M., Switzer, P., & Craig, M. D. (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *Geoscience and Remote Sensing, IEEE Transactions on*, 26(1), 65–74.
- Huang, W., Lamb, D. W., Niu, Z., Zhang, Y., Liu, L., & Wang, J. (2007). Identification of yellow rust in wheat using in situ spectral reflectance measurements and airborne hyperspectral imaging. *Precision Agriculture*, 8(4–5), 187–197.
- Kruse, F. A., Lefkoff, A. B., Boardman, J. W., Heidebrecht, K. B., Shapiro, A. T., Barloon, P. J., et al. (1993). The spectral image-processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, 44(2–3), 145–163.
- Kumar, A., Lee, W. S., Ehsani, R., Albrigo, L. G., Yang, C. R., & Mangan, L. (2012). Citrus greening disease detection using aerial hyperspectral and multispectral imaging techniques. *Journal of Applied Remote Sensing*, 6(1), 063542.
- Lee, W. S., Ehsani, R., & Albrigo, L. G. (2008). Citrus greening (Huanglongbing) detection using aerial hyperspectral imaging. In *Proceedings of the 9th International Conference on Precision Agriculture*, Denver, CO.
- Li, X., Lee, W. S., Li, M., Ehsani, R., Mishra, A. R., Yang, C., et al. (2012). Spectral difference analysis and airborne imaging classification for citrus greening infected trees. *Computers and Electronics in Agriculture*, 83, 32–46.
- Mewes, T., Franke, J., & Menz, G. (2011). Spectral requirements on airborne hyperspectral remote sensing data for wheat disease detection. *Precision Agriculture*, 12(6), 795–812.
- Nascimento, J. M. P., & Dias, J. M. B. (2005). Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Transactions on Geoscience and Remote*, 43(4), 898–910.
- Nidamanuri, R. R., & Zbell, B. (2011). Transferring spectral libraries of canopy reflectance for crop classification using hyperspectral remote sensing. *Biosystem Engineering*, 110(3), 231–246.
- Parker, J. A., Kenyon, R. V., & Troxel, D. E. (1983). Comparison of interpolating methods for image resampling. *IEEE Transactions on Medical Imaging MI*, 2(1), 31–39.
- Plaza, A., Benediktsson, J. A., Boardman, J. W., Brazile, J., Bruzzone, L., Valls, G. C., et al. (2009). Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment*, 113(Supplement 1), S110–S122.
- Qin, J., Burks, T. F., Ritenour, M. A., & Bonn, W. G. (2009). Detection of citrus canker using hyperspectral reflectance imaging with spectral information divergence. *Journal of Food Engineering*, 93(2), 183–191.
- Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8), 1627–1639.
- Smith, G. M., & Milton, E. J. (1999). The use of the empirical line method to calibrate remotely sensed data to reflectance. *International Journal of Remote Sensing*, 20(13), 2653–2662.
- Smith, K. L., Steven, M. D., & Colls, J. J. (2005). Plant spectral responses to gas leaks and other stresses. *International Journal of Remote Sensing*, 26(18), 4067–4081.
- 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/news/tda-and-usda-confirm-citrus-greening/>. Retrieved March 21, 2013.
- Ustin, S. L., Roberts, D. A., Gamon, J. A., Asener, G. P., & Green, R. O. (2004). Using imaging spectroscopy to study ecosystem processes and properties. *BioScience*, 54(6), 523–534.
- Winter, M. E. (1999). N-findr: An algorithm for fast autonomous spectral endmember determination in hyperspectral data (pp. 266–277). In *Proceedings of SPIE Conference Image Spectrometry V*.
- Yang, C., Everitt, J. H., & Bradford, J. M. (2008). Yield estimation from hyperspectral imagery using spectral angle mapper (SAM). *Transactions of the ASABE*, 51(2), 729–737.
- Ye, X., Sakai, K., Sasaoa, A., & Asadac, S. (2008). Potential of airborne hyperspectral imagery to estimate fruit yield in citrus. *Chemometrics and Intelligent Laboratory Systems*, 90(2), 132–144.
- Zhang, M., Qin, Z., Liu, X., & Ustin, S. L. (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.