



# Detection of stress in tomatoes induced by late blight disease in California, USA, using hyperspectral remote sensing

Minghua Zhang\*, Zhihao Qin, Xue Liu, Susan L. Ustin

*Department of Land, Air and Water Resources, University of California, Davis, CA 95616, USA*

Received 2 October 2002; accepted 14 April 2003

## Abstract

Large-scale farming of agricultural crops requires on-time detection of diseases for pest management. Hyperspectral remote sensing data taken from low-altitude flights usually have high spectral and spatial resolutions, which can be very useful in detecting stress in green vegetation. In this study, we used late blight in tomatoes to illustrate the capability of applying hyperspectral remote sensing to monitor crop disease in the field scale and to develop the methodologies for the purpose. A series of field experiments was conducted to collect the canopy spectral reflectance of tomato plants in a diseased tomato field in Salinas Valley of California. The disease severity varied from stage 1 (the light symptom), to stage 4 (the sever damage). The economic damage of the crop caused by the disease is around the disease stage 3. An airborne visible infrared imaging spectrometer (AVIRIS) image with 224 bands within the wavelength range of 0.4–2.5  $\mu\text{m}$  was acquired during the growing season when the field data were collected. The spectral reflectance of the field samples indicated that the near infrared (NIR) region, especially 0.7–1.3  $\mu\text{m}$ , was much more valuable than the visible range to detect crop disease. The difference of spectral reflectance in visible range between health plants and the infected ones at stage 3 was only 1.19%, while the difference in the NIR region was high, 10%. We developed an approach including the minimum noise fraction (MNF) transformation, multi-dimensional visualization, pure pixels endmember selection and spectral angle mapping (SAM) to process the hyperspectral image for identification of diseased tomato plants. The results of MNF transformation indicated that the first 28 eigenimages contain useful information for classification of the pixels and the rest were mainly noise-dominated due to their low eigenvalues that had few signals. Therefore, the 28 signal eigenimages were used to generate a multi-dimensional visualization space for endmember spectra selection and SAM. Classification with the SAM technique of plants' spectra showed that the late blight diseased tomatoes at stage 3 or above could be separated from the healthy plants while the less infected plants (at stage 1 or 2) were difficult to separate from the healthy plants. The results of the image analysis were consistent with the field spectra. The mapped disease distribution at stage 3 or above from the image showed an accurate conformation of late blight occurrence in the field. This result not only confirmed the capability of hyperspectral remote sensing in detecting crop disease for precision disease management in the real world, but also demonstrated that the spectra-based classification approach is an applicable method to crop disease identification.

© 2003 Elsevier B.V. All rights reserved.

*Keywords:* Tomato; Pest management; Hyperspectral remote sensing; AVIRIS; Late blight disease; Canopy spectra

## 1. Introduction

California accounts for ~75% of US total harvested acreage and >86% of US tomato production (USDA,

\* Corresponding author. Tel.: +1-530-752-4953;

fax: +1-530-752-5262.

*E-mail address:* mhzhang@ucdavis.edu (M. Zhang).

1997). Though climate in California is generally suitable for large-scale tomato farming, disease is still one of the leading factors threatening sustainability of tomato production. Pest management is usually one of the main components in tomato production costs. Each year California tomato farmers use 20–23 million pounds of pesticide active ingredients, with an intensity of 25–30 pounds per acre (CDPR, 2000). The use of large amounts of pesticides not only causes potential risks to the environment and ecosystem but also increases the cost of farming. Large-scale tomato cropping would benefit from a timely and location-specific detection and monitoring system to reduce the cost and risks associated with spraying pesticides. Detection of crop disease stress potentially represents one of the most valuable applications of remote sensing to agriculture (Haboudane et al., 2002; Boegh et al., 2002).

With the rapid development of remote sensing technology in recent decades, hyperspectral remote sensors, such as airborne visible infrared imaging spectrometer (AVIRIS), compact airborne spectrographic imager (CASI), multispectral infrared and visible imaging spectrometer (MIVIS), and hyperspectral mapping (HyMap<sup>TM</sup>) system, are now available to agricultural applications. These sensors can provide quality images with high spatial and spectral resolutions required for precision agriculture (Taranik et al., 1993; Fraser, 1998; Wilson and Felt, 1998; Bianchi et al., 1999; Treitz and Howarth, 1999; Leblanc et al., 1999; Nolin and Dozier, 2000). These high resolution images also have great potential applications in environmental impact assessment (Barducci and Pippi, 1997; Bianchi et al., 1999; Goetz et al., 1985; Deguise and Staenz, 1999; Holden and Ledrew, 1998). Because of the high spectral resolution with a narrow band range of about 10 nm or finer, hyperspectral remote sensing images produce a complete spectrum for each pixel within the scene. These characteristics combined with high signal-to-noise ratio enable us to differentiate various vegetation stresses based upon spectra of small patches of ground surface (Rush, 2002; Christ et al., 2000; Lelong et al., 1998).

When plants are stressed, such as by disease, their absorption of incident light changes in the visible range and in the NIR range (Carter and Knapp, 2001; Adams et al., 1999; Dawson and Curran, 1998; Lichtenthaler et al., 1996; Gitelson and Merzlyak,

1994; Guenther, 1990). This reaction is probably due to the decreased chlorophyll content, changes in other pigments, and foliar internal structure. The change of absorption consequently influences the reflectance of stressed plants. Therefore, in comparing the spectrum difference of stressed and healthy plants, theoretically, we are able to identify the stress severity of green vegetation. Many studies have examined the relationship between chlorophyll content and spectral reflectance in visible and NIR ranges (Carter et al., 1996; Datt, 1998, 1999; Gitelson and Merzlyak, 1996, 1997; Gitelson et al., 1996; Knapp and Carter, 1998; Blackburn and Steele, 1999). However, most of these studies were based on leaf spectral reflectance and very few were on canopy light response (Gitelson and Merzlyak, 1996, 1998).

If a plant is infected, the physiological reaction of a plant caused by disease will result in a change of spectral reflectance due to the decreasing chlorophyll content and changing internal structure. Since the chlorophyll content tends to decrease under disease stress, the incident solar radiation absorption of the green plant generally results in a decrease in the visible region. Consequently, the spectral reflectance generally is higher in the visible green range depending on the infection severity. The strong spectral reflectance of green plants in the NIR range is mainly due to its internal foliar structure. Plants under disease stress also show various degrees of internal structural changes, which lead to a decrease of spectral reflectance in the NIR range. These spectral features of vegetation are the basis for remote sensing of disease-stressed tomato plants.

Late blight, caused by the fungal pathogen *Phytophthora infestans*, is a disease that spreads quickly in tomato fields in suitable weather conditions during growing season. The infection can threaten the sustainability of tomato farming in California (Wisler and Duffus, 2000; Clark et al., 1998). Humid air with high temperature provides favorable weather conditions for the outbreak of the disease. In addition to suitable weather conditions, such factors as irrigation, cultivar and farming system also contribute to the potential risk of the disease in some tomato fields.

Farmers operating large-scale tomato fields required timely and location-precise detecting and monitoring of the infected plants for disease control to ensure their harvest while reducing pollution risks from pesticides.

Conventional ground scouting was not economic to provide the efficient detection and monitoring in a large tomato cropping area. Remote sensing, however, can provide a powerful technology to collect crop canopy data that can be used to analyze the geo-temporal and geo-spatial properties of the biological features of the crop canopies, including the symptoms of late blight.

The objective of this study was to investigate, using tomatoes infected by *P. infestans* as an example, the capability of hyperspectral remote sensing in crop disease detection and to develop an approach to explore this potential. The purpose was to distinguish the infected tomato plants from the healthy ones in the AVIRIS image using an approach developed from the spectral analysis of the tomato canopy spectra collected from the field. We combined the minimum noise fraction (MNF) transformation with multi-dimensional visualization, pure pixels endmember selection, and spectral angle mapping (SAM) to classify the diseased pixels of tomato canopies from the healthy ones in the AVIRIS image. The results from the classification demonstrated the promising application of hyperspectral remote sensing to detect tomato disease when fast image processing is available for proper and easily applied identification algorithms.

## 2. Discrimination analysis using field canopy spectra

### 2.1. Spectra data collection

The canopy spectra data of plants were collected in four fields in Salinas Valley, California (Fig. 1) in the late summer of 1998. The field's size was approximate 40 acres and the tomatoes were at an early to middle fruiting stage. Due to the proximity of the Salinas Valley to the Pacific Ocean, during the summer, the weather is warm and the air relative humidity is high, which are suitable for infection by *P. infestans* and development of late blight disease. During the time of spectra sampling, the tomato fields were identified, by the field plant pathologist, Carla Thomas from the Western Farm Service, as containing late blight diseased plants with various severities. Probably due to the effect of wind direction and moisture levels in the field, the diseased areas often followed a pattern of strips.

We used a spectrometer GER-2600 to measure the spectral reflectance of the canopies at 1 m above the canopy with a field of view (FOV) of 23°. The wavelength of the measurement was configured for the range of 400–2500 nm. A total of 66 canopy spectra

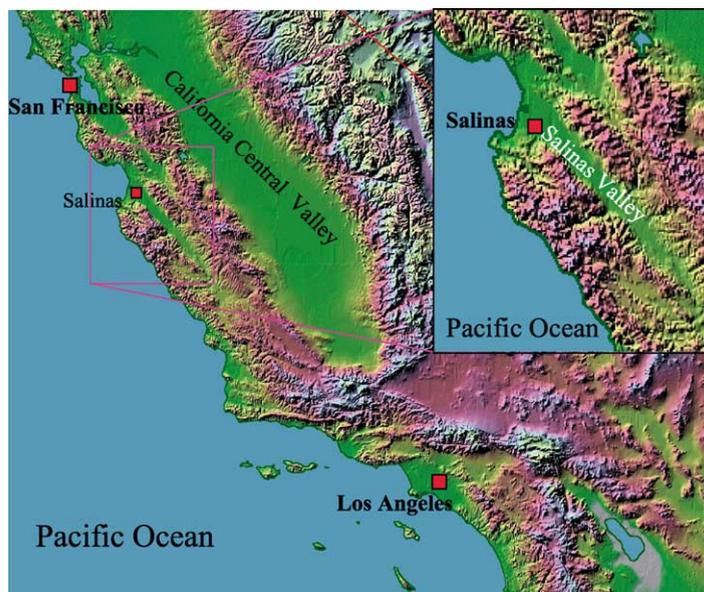


Fig. 1. Relief image of coastal California, showing the location of Salinas Valley.

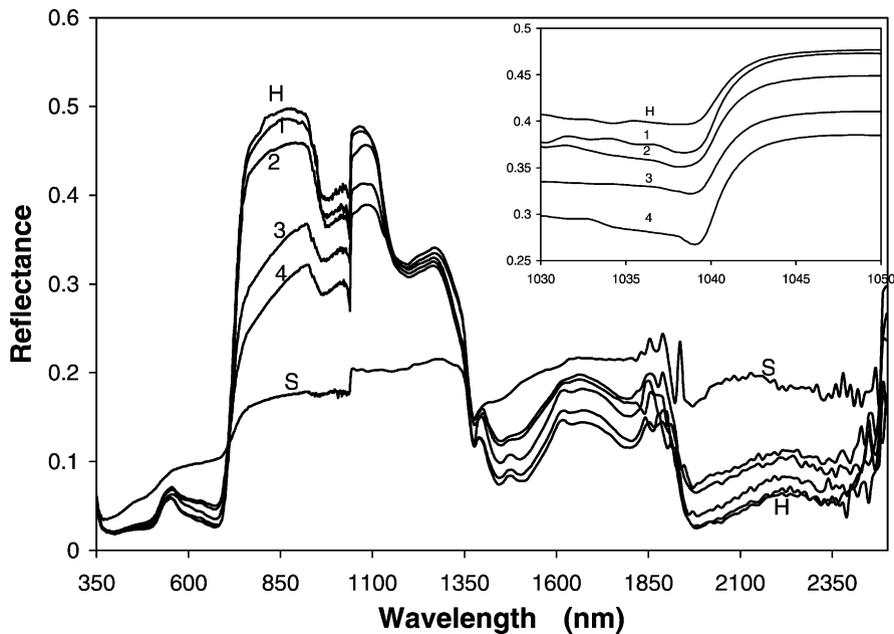


Fig. 2. Field spectral reflectance. Curve H: the average spectra of healthy plants. Curve 1: the average spectra of infected plants at stage 1 and curves 2, 3 and 4: at stages 2, 3 and 4, respectively. Curve S: the average spectra of the soil. The insert is an enlarged view of the abrupt changes at approximate 1040 nm.

data were sampled at various disease infection severities in the field, 22 spectra of healthy plants, 11 spectra of light diseased plants at stage 1, 12 spectra of light diseased plants at stage 2, 17 spectra of diseased plants at stage 3, and 4 spectra of diseased plants at stage 4. We also sampled nine spectra of bare soils to analyze the impact of soil background. Therefore, together with the soil, we had six groups of spectra for the analysis. The statistical differences were assessed by examining the standard deviations of spectra within each group (Zhang et al., 2002). After that, the spectral data were then calculated to obtain a mean spectrum for each group (Fig. 2). Table 1 summarizes the spectral characteristics of the six groups.

To relate the canopy spectra to infection severity, we categorically defined the following five disease severity groups:

- (1) healthy plants in which no symptoms were observed on the leaves;
- (2) infected plants at stage 1 (LB1) in which initial symptoms were seen but only with one lesion on one or two leaves;
- (3) infected plants at stage 2 (LB2) in which symp-

toms were a little more severe than stage 1 (one lesion on more than two leaves);

- (4) infected plants at stage 3 (LB3) in which a minimum of two lesions were on at least one leaf;
- (5) infected plants at stage 4 (LB4) in which two lesions were on over half the canopy leaves.

## 2.2. Spectra analysis for discrimination

The spectral characteristics shown in Fig. 2 provided a theoretical basis for the potential application of remote sensing for the discrimination. The healthy plants had a spectrum that is obviously different from the diseased plants and soil background (Fig. 2). Four sub-ranges were clearly identified for their spectral differences. In the visible range, the soil had the highest reflectance and the healthy plants the lowest. For diseased plants, reflectance was between the reflectance readings of soil and healthy plants. However, a spectral difference between the healthy plants and the infected ones was small in some ranges of the wavelengths.

Table 1 indicates that, in the range of 400–700 nm, average reflectance of healthy plants was 0.0338 and

Table 1  
Average field spectral reflectance of the tomato field

Ranges (nm)	Healthy plants	Diseased plants				Soil
		Stage 1	Stage 2	Stage 3	Stage 4	
Spectral reflectance						
400–700	0.0338	0.0352	0.0406	0.0457	0.0474	0.0791
700–1350	0.4159	0.4051	0.3921	0.3167	0.2790	0.1712
1350–1900	0.1224	0.1300	0.1508	0.1649	0.1720	0.2025
1900–2500	0.0686	0.0625	0.0784	0.0966	0.1124	0.1882
400–500	0.0229	0.0238	0.0257	0.0268	0.0278	0.0525
500–600	0.0456	0.0467	0.0527	0.0570	0.0588	0.0850
600–690	0.0314	0.0339	0.0416	0.0516	0.0544	0.0999
750–930	0.4826	0.4714	0.4466	0.3353	0.2879	0.1713
950–1030	0.4031	0.3842	0.3747	0.3332	0.2954	0.1773
1040–1130	0.4468	0.4427	0.4300	0.3926	0.3661	0.2005
1450–1850	0.0991	0.1049	0.1242	0.1414	0.1495	0.2020
2000–2400	0.0528	0.0540	0.0687	0.0911	0.1012	0.1860
Difference from healthy plants						
400–700		–0.0014	–0.0068	–0.0119	–0.0136	–0.0453
700–1350		0.0108	0.0239	0.0992	0.1369	0.2448
1350–1900		–0.0077	–0.0284	–0.0426	–0.0497	–0.0801
1900–2500		0.0061 <sup>a</sup>	–0.0098	–0.0279	–0.0438	–0.1196
400–500		–0.0009	–0.0029	–0.0039	–0.0049	–0.0297
500–600		–0.0011	–0.0071	–0.0114	–0.0132	–0.0394
600–690		–0.0025	–0.0102	–0.0202	–0.0229	–0.0685
750–930		0.0111	0.0360	0.1473	0.1947	0.3113
950–1030		0.0189	0.0285	0.0700	0.1078	0.2258
1040–1130		0.0041 <sup>a</sup>	0.0168 <sup>a</sup>	0.0542	0.0807	0.2463
1450–1850		–0.0058	–0.0251	–0.0423	–0.0504	–0.1029
2000–2400		–0.0012 <sup>a</sup>	–0.0159	–0.0383	–0.0484	–0.1332

<sup>a</sup> The values are statistically insignificant at 0.05 level. Other values are all statistically significant at 0.01 level.

average reflectance of LB3 and LB4 plants were 0.0457 and 0.0474, respectively. Their standard deviations were minimal in magnitude. Thus, the spectra differences in the visible range between the healthy plants and the most severe infected ones, LB3 and LB4 was around 1.19 and 1.46%, respectively. The spectra difference was even smaller for the LB1 and LB2 plants (0.14 and 0.68%). The spectral difference between the healthy plants and the soil was the highest, 4.53%, among all groups. Thus, except for soils, the potential discrimination between healthy and diseased plants was possible for the LB3 and LB4 plants, but not for the LB1 and LB2 plants (Fig. 2).

The lower half of Table 1 provides more detailed comparisons. In the red range (600–690 nm), a larger difference of spectral reflectance was found for the healthy plants and the plants at higher infection stages. The maximum differences were 2.3 and 2%

for the LB4 and LB3 plants, respectively (Table 1). This was consistent with the result of Carter and Knapp (2001) to loblolly pine. Due the statistical significance, healthy and diseased LB3 and LB4 plants can be separated in this red range, in addition to soils.

In the near infrared (NIR), healthy plants had the highest spectral reflectance, while soils the lowest (Fig. 2). On average, healthy plants had a reflectance of 0.4159, while the LB3 and LB4 plants had a reflectance of 0.3167 and 0.2790 (Table 1). The spectral differences between healthy and diseased LB3 and LB4 plants were, 9.92 and 13.69%, respectively. The spectral difference of healthy plants and soils was 24.48%. Therefore, this range of wavelength seems more valuable than the visible range to detect crop disease.

As shown in Fig. 2, the reflectance of the tomato plants reached the highest within two narrow ranges:

750–930 and 1040–1130 nm. The spectra curves showed a valley shape in the range of 950–1030 nm. Table 1 showed that the range of 750–930 nm had the greatest discrimination among the six groups. In this range, the spectral difference of 19.47% was found between healthy plants and the LB4 plants, and the spectral difference of 31.13% was found between healthy plants and soils. Even for the LB1 and LB2 plants, we observed a difference of 1.11 and 3.60%, respectively. This notable difference of spectra demonstrated the importance of this range in remote sensing of diseased tomatoes.

For the ranges of 950–1030 and 1040–1130 nm, a more than 10%, 5–7% of spectral differences were found for healthy plants and the LB4 and LB3 plants. The spectral differences of healthy and the LB3, and LB4 plants were found to be 4.85 and 7.33% in the 700–750 nm range, and 7.03 and 10.91% in the 950–1040 nm range, respectively (Table 1). These ranges were also valuable for remote sensing of the disease. Even at the ranges of 1450–1850 and 2000–2400 nm, the spectral differences between healthy and disease LB3 and LB4 plants were sufficiently large for the separations.

The above spectral analysis indicated that the range of 750–930 nm was, statistically, the best wavelength interval for remote sensing of the tomato disease, followed by that of 950–1030 and 1040–1130 nm. The spectral reflectance of the LB3 and LB4 plants was significantly different from the healthy plants in these ranges at significance level of  $P = 0.01$ . Some narrow ranges, such as 700–750 nm also have great potential in the discrimination. The generally higher spectral difference in the NIR range than in the visible range demonstrates the importance of the infrared remote sensing in monitoring of tomato disease for precision agricultural practices.

### 3. Approaches to detect plant diseases by AVIRIS imagery

Based on the above optical feature analysis of plants with known disease severity, we used AVIRIS images of the field to identify diseased plants from the healthy plants. The hyperspectral remote sensing image was acquired at low flight altitude during September 1998. The spectra had 224 bands with a band span

of ~10 nm from 400 to 2500 nm. However, due to inadequate quality of some bands in the original image, we conducted a preliminary image process to eliminate those low quality bands, which resulted in a hyperspectral image with 180 bands for the analysis. The image had a spatial resolution of 4 m. The atmospheric removal program ATREM developed by University of Colorado (CSES, 1999) was used to calibrate the image before analyses.

#### 3.1. Minimum noise fraction (MNF) transformation

The MNF transformation proposed by Green et al. (1988) is an effective method to determine the inherent dimensionality of image data and to segregate noise (Boardman, 1993; Boardman and Kruse, 1994; Jia and Richards, 1999). The MNF transformation provided in the image-processing package ENVI, version 3.2 (ENVI, 1999) was essentially a two-cascaded principal component transformation (PCT) procedure. The procedure assumed that each pixel contained both signal and noise. This method, based on the signal-to-noise ratio, allowed us to determine which bands were more useful in the discrimination between the healthy and diseased plants.

Thus, the first step in the MNF transformation was to estimate the noise statistics from the data and create a noise covariance matrix, which was used to decrease correlation among the spectral bands and rescales the noise in the data. Results from this first step were the transformed data in which the noise had unit variance and no band-to-band correlation. The second step in PCT was a standard principal component transformation of the noise-signal data. Using each band as a variable and the pixels as samples, the principal component transformation produced a number of eigenvectors (indicated by their eigenvalues) and their corresponding eigenimages for the image dataset. However, some eigenvectors were more important than others as measured by eigenvalues. Eigenvectors with large eigenvalues constitute the main principle components while the ones with smaller eigenvalues are less important components. The inherent dimensionality of the image data was determined by examination of the results from the PCT in two portions, in which the first was associated with those eigenvectors with large eigenvalues and their coherent eigenimages, and the second with near-unity

eigenvalue and noise-dominated images. By using only the portion with the coherent eigenimages, the noise could be separated from the signal in the image data, hence improving the processing results.

### 3.2. Classification of the image with signal bands

After removing the noise, classification could be performed on the image with signal bands for the separation. Given that we know the different spectral patterns of healthy and diseased plants, supervised classification is a logical choice to analyze the images. Using the field spectra as the trained pattern, we were able to obtain a direct connection of the classification with field spectral measurements.

We selected the SAM technique to classify the image. This SAM technique uses the  $n$ -dimensional angle to match the pixels to the reference (the identified) spectra. It determines the similarity between the two spectra by calculating the angle between them. SAM treats the spectra as vectors in a space with dimensionality equal to the number of bands and compares the angle between the reference spectrum vector and each pixel vector in the  $n$ -dimensional space, the formula as follows:

$$\text{SAM} = a \cos \left( \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}} \right).$$

Smaller vector angle represents closer matches to the reference spectrum, and hence can be grouped into the cluster represented by the reference spectrum. Pixels further away than the specified maximum angle threshold in radians are not classified.

### 3.3. Endmember spectra extraction and potential pattern identification

SAM classification requires the reference spectra as inputs to compare with the observed pixel spectra to identify their similarity. Visualization analysis is another method to identify the potential patterns in the hyperspectral image and to extract the endmember spectra from the image, which can then serve as the reference spectra for SAM classification. As in our case, we first created a subset of the AVIRIS image and select the field with known infected plant distribution as a region of interest (ROI). Then we used

the signal bands to generate an  $n$ -dimension feature space to visualize all the pixels within the ROI subset image, where  $n$  represented the number of signal bands. All the pixels within the ROI were projected into the space for analysis. Then we spun the visualized feature space to observe change of the plot in the space. Continually spinning the space, we were able to identify potential patterns of the pixels in the image.

Pure pixels or the pixels with typical spectral features in the image could be identified with interactive selection in the visualization space (Harsanyi and Chang, 1994). The endmember spectra of the pure pixels could be obtained through computing their mean digital number (DN) value in each band of the image. Finally, we related these pure pixels with original subset image and extracted image spectra of these pure pixels from the 180-band image. This was a crucial step when interpreting the results of endmember spectra extraction because the signal-band MNF image for the visualization only generated some endmember spectra, and was not directly associated to health status of tomato plants. To use the endmember spectra for SAM classification, we first had to understand the meaning of the endmember spectra before using them in SAM classification.

To interpret the actual representation of these extracted image endmember spectra in terms of infection severity, we compared the extracted image spectra with the obtained field spectra. First, we located the disease locations of the field in the image and determined their infection severities from field data. We took 10 pixels for each infection stage and extracted the pixel DN values from the image bands to produce the image spectra. The average DN value for the 10 pixels was used to represent the image spectra of each infection stage. Then we scaled the average DN values into the relative reflectance values for the comparison, using the DN peak for the healthy plants in NIR region as a reference basis. The peak DN value of 4999 for healthy plants at 779 nm on the image corresponds to the reflectance of 0.4639 in the field spectra. The formula used was

$$\text{MDN}_j = \frac{\sum \text{DN}_{ij}}{N},$$

where MDN is the average DN of the 10 pixels for band  $j$ ,  $\text{DN}_{ij}$  the DN of pixel  $i$  for band  $j$ ,  $N$  the total

number of pixels ( $N = 10$ ). The scale formula used was

$$\text{REF}_j = \frac{\text{MDN}_j \times \max(\text{REF})}{\max(\text{MDN})},$$

where  $\text{REF}_j$  is the scaled reflectance for band  $j$ ,  $\text{MDN}_j$  the average DN for  $j$ ,  $\max(\text{REF})$  the maximum reflectance (i.e.  $\max(\text{REF}) = 0.4639$ , REF for the band with wavelength 779 nm),  $\max(\text{MDN})$  the maximum average DN ( $\max(\text{MDN}) = 4999$ , i.e. MDN for the band with wavelength 779 nm). After converting the DN values into reflectance, we calculated the spectra differences between the image and the field spectra data for the specific wavelength regions of interest (the intervals in Table 1). By computing the similarity matrix for each identified group, the SAM classification was able to provide a preliminary view of the capability of separating the identified groups in the image. Rule images were the intermediate classification result, indicating the similarity of each pixel under the classification to the identified clusters. Therefore, for each potential cluster, we had one rule image. The pixel patterns with low correlation to each other could represent the other cluster hidden in the images that need to be highlighted.

#### 3.4. Performance of SAM classification on the subset image of the field

Taking the extracted endmember spectra from the AVIRIS image as the reference, we performed the SAM classification on the subset image to identify the diseased plants. The above visualization analysis provided a result of four endmember spectra representing four clusters. The MNF analysis indicated that the first 28 bands are signal bands. Thus, we aimed at the four clusters on the subset image with 28 bands in the SAM classification. Since our subset image of the field also covered some parts of other fields, we also took the spectra of these fields as reference clusters for classification. There were four fields surrounding the tomato field but the north and the south fields were actually with the same crop. Hence, only three extra reference clusters were identified for these surrounded fields. Therefore, together with the four clusters in the tomato field, we had seven groups for classifying the hyperspectral subset image. The SAM analysis applied a pixel-based compar-

ison between each endmember spectrum and each pixel spectrum in the image for the classification. After trying several options, we found that the angle threshold of 0.18 is the best choice for the classification. This was because small angles of two vectors meant to be similar for the pixels while large angles of two vectors meant to be far apart. With this threshold all image pixels could be properly classified into the seven clusters with minimal misclassification.

## 4. Image analyses

### 4.1. Spectral discrimination in the image space

Field tomato canopy spectra indicated that only certain ranges of the wavelengths were valuable for remote sensing of crop disease. Among the 180 bands of the original AVIRIS image, only 28 eigenvectors had eigenvalues greater than 2 in MNF analysis (Fig. 3a). We used the first 28 eigenimages derived from the 28 eigenvectors with eigenvalues greater than 2 for further analysis and created this 28-band eigenimages from the original AVIRIS image.

To perform the classification on the newly created signal images, we needed to know the relationship between each band and the spectral features of the infected plants relative to the healthy ones. Associated with the MNF analysis was a function to project the field spectra of the tomato plants at various infection severities into the same feature space as the MNF eigenimage. Fig. 3b illustrated the results for the first 28 important bands. Since the MNF transformation was also a normalization process, the resulting MNF reflectance fluctuated between positive and negative values to keep the feature of zero variance. The classification of the field image with the 28 bands could be done through the comparison of the spectral similarity between the pixel and the projected spectra from the obtained field spectral reflectance. As shown in Fig. 3b, the MNF reflectance for the healthy plants (1.714) was very close to the LB1 plants (1.006). Mean MNF reflectance was 4.857 for LB3 plants, 5.750 for LB4 plant, with a difference of 0.893. The difference of the mean MNF reflectance was 3.143 between healthy and LB3 plants, and 4.750 between LB4 and LB1 plants. The largest difference

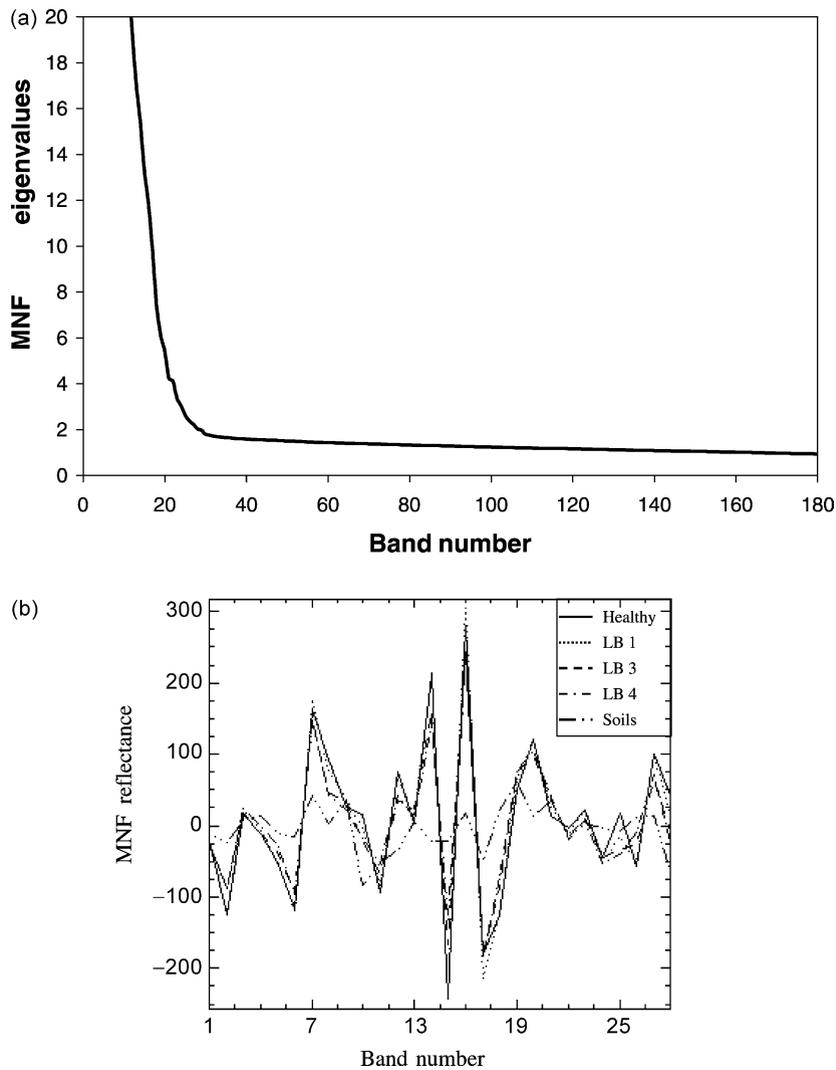


Fig. 3. Results of MNF transformation: (a) MNF eigenvalues for the whole 180 bands; (b) MNF reflectance for the first 28 bands. Healthy represents the healthy plants; LB1, LB3, and LB4 represent the infected plants at stages 1, 3 and 4, respectively; and Soils represents the background soils.

of the mean MNF reflectance was 4.713 between the healthy plants and soils. Moreover, the MNF reflectance of LB3 or LB4 plants was between the two extremes of healthy plants and the soil for each band (Fig. 3b). These spectral features were the same as the features observed from the field spectra. Hence, these features could be used as a basis for the image classification of diseased and healthy plants of tomatoes.

#### 4.2. Visualization analysis

Fig. 3 showed the change of the four endmember spectra in the range of 0.4–2.4  $\mu\text{m}$ . Since the ROI of the subset image was a tomato field without bare soil, we might interpret the endmember spectra as categories of the diseased plants at various severities. In this way, the endmember 1 in Fig. 3 could be characterized as the healthy plants because it had the highest

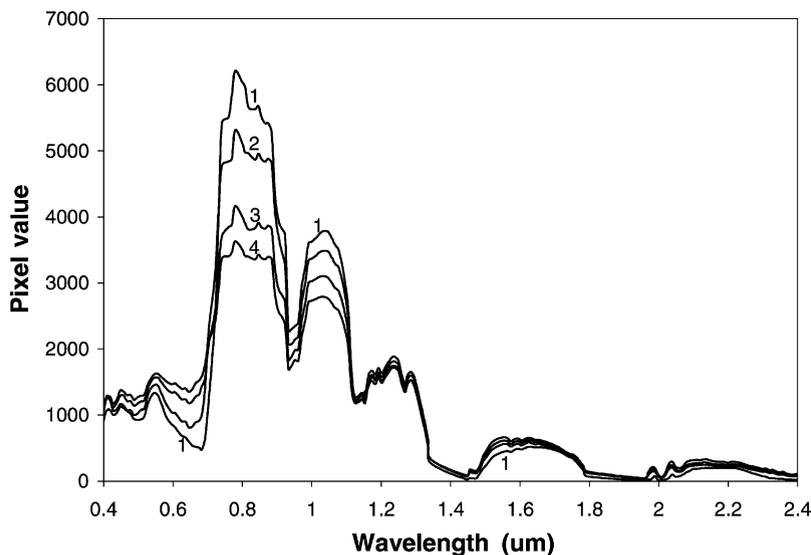


Fig. 4. Endmember spectra of the pure pixels representing each category of the classification, resulted from the visualization analysis of the 28 signal bands. The labels are: 1, the first endmember; 2, the second endmember; 3, the third endmember; 4, the fourth endmember.

DN value in the NIR region (0.7–1.3  $\mu\text{m}$ ) and the lowest DN value in the visible region. In contrast, endmember 4 might be referred to the severe diseased plants because it had the lowest DN value in the NIR region and the highest DN value in the visible region. Since the LB4 was referred to most severe diseased plants in the field, this endmember might represent the characteristics of LB4 plants. Endmember 3 had slightly higher DN value in the NIR region and a relatively lower DN value in visible range than endmember 4, which was referred as LB3 (Fig. 4). Endmember 2 had a mean DN value between endmember 1 and endmember 3 in the NIR region. Its relatively large DN value difference from endmember 1 illustrated that this endmember might represent the characteristics of the LB2 plants.

#### 4.3. Comparison of the image spectra with the field spectra

As a number of factors, especially atmospheric and the environmental factors (mainly soil in our case) may affect remotely sensed data, the image spectra extracted from hyperspectral remote sensing data was usually different from the field collected spectra. Because the influencing factors can be minimized in the field using handheld spectrometer, field spectra gen-

erally were much more pure than the image. Hence, comparison of the image spectra with the field spectra would help us better understand about their similarity and difference in the identification capability of the crop diseases. Fig. 5 illustrates the average image spectra for the tomato plants with various diseased stages and the soil. In spite of magnitude difference, the shape of the image spectra was similar with that of the endmember spectra (Fig. 4). Given that endmember spectra are mixed spectra, this was not surprising.

Though the general shape of the image spectra was similar to the field spectra (Fig. 2) in the visible and NIR regions, the apparent differences could still be observed in the range of 1.4–2.5  $\mu\text{m}$ , where two main domes, with one between 1.5 and 1.8  $\mu\text{m}$  and another one between 2.05 and 2.3  $\mu\text{m}$ . These differences could be mainly ascribed to the effects by the sensor characteristics, the atmospheres, and the environmental factors. The complex atmospheric conditions were not fully known, hence image calibration can only remove the atmospheric effects to some extent. In addition, remote sensor properties and the complicated surface conditions also added the complexity into the remote sensing imagery, which could also contribute to the differences of the image spectra from the field ones.

In the visible region (0.4–0.7  $\mu\text{m}$ ), the average reflectance of the image spectra was relatively higher

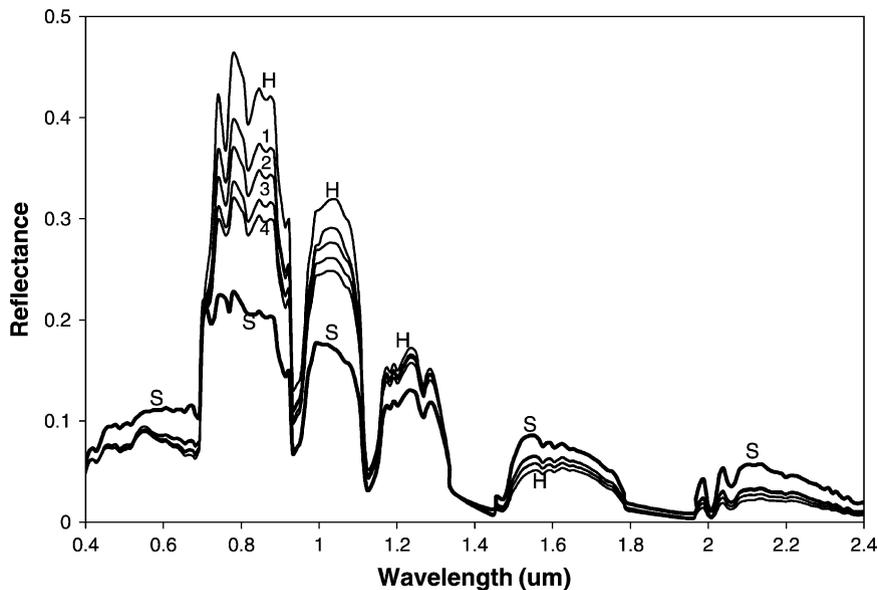


Fig. 5. Image spectra of the field, illustrating the spectral difference of each category in the remote sensing imagery scale. The labels are: H for healthy plants, 1 for LB1, 2 for LB2, 3 for LB3, 4 for LB4, and S for soil.

than that of the field spectra, while in the NIR region of 0.7–2.5  $\mu\text{m}$ , however, the reflectance was generally lower in the image spectra (Table 2). This was probably due to the different contributions of atmospheric and environmental (mainly soil) effects in these two spectral regions. Background soil had higher reflectance in the visible region and much lower reflectance in NIR region than the green crops. The spectral reflection and absorption of the atmosphere was also relatively higher in the visible region than in the NIR region. Thus, in the visible region, the effects of the atmosphere and the background soil might eventually lead to an additive contribution to the spectral reflectance in the image level, while in the NIR region, these effects might lead to a subtraction to the reflectance in the image when compared to the reflectance in the field.

The average reflectance differences of the healthy and diseased plants were low in the range of 0.4–0.7  $\mu\text{m}$  (Table 2). The average reflectance of the image spectra was apparently lower than that of the field spectra in the region of 0.7–1.35  $\mu\text{m}$ . As shown in Fig. 5, there were two deep valleys in this region, while they were very shallow in the field spectra as shown in Fig. 2. Detailed comparison to the narrow ranges such as 0.4–0.5, 0.6–0.69, 0.75–0.93,

1.04–1.130 and 1.45–1.85  $\mu\text{m}$  demonstrated that the separation capability was generally lower in the image spectra than in the field spectra. Though the sensor characteristics, atmosphere and the environmental factors tended to attenuate the general difference of spectral reflectance between the healthy and diseased plants, the separation between them was still high in the NIR region especially in the narrow peak region such as 0.75–0.9 and 0.95–1.0  $\mu\text{m}$ . The relatively high spectral differences in these narrow regions provided the capability of separation for the healthy and diseased plants in the image scale.

#### 4.4. Preliminary view to the relations in the rule image spaces

Using the SAM technique, we generated four rule images to represent the similarity between reference spectrum and each pixel spectrum for a preliminary view to their relations. If two rule images were highly correlated, we could not separate the pixels. Therefore, we should only highlight the rule images with less correlation for their ability to represent the cluster hidden in the images. Fig. 6a illustrated the linear correlation between the healthy and LB2 plants. The determination coefficient was  $R^2 = 0.9638$ , indicating

Table 2  
Average hyperspectral image reflectance of the tomato field

Ranges (nm)	Healthy plants	Diseased plants				Soil
		Stage 1	Stage 2	Stage 3	Stage 4	
Spectral reflectance						
400–700	0.0780	0.0786	0.0771	0.0814	0.0822	0.1030
700–1350	0.2465	0.2211	0.2120	0.2007	0.1922	0.1430
1350–1900	0.0367	0.0409	0.0407	0.0449	0.0455	0.0577
1900–2500	0.0144	0.0175	0.0174	0.0205	0.0218	0.0366
400–500	0.0667	0.0667	0.0662	0.0683	0.0689	0.0842
500–600	0.0847	0.0838	0.0818	0.0856	0.0862	0.1068
600–690	0.0710	0.0748	0.0731	0.0805	0.0819	0.1105
750–930	0.3785	0.3287	0.3087	0.2835	0.2700	0.1883
950–1030	0.2007	0.1847	0.1800	0.1737	0.1665	0.1256
1040–1130	0.2175	0.1950	0.1890	0.1792	0.1706	0.1177
1450–1850	0.0387	0.0432	0.0430	0.0475	0.0482	0.0613
2000–2400	0.0156	0.0189	0.0187	0.0220	0.0233	0.0390
Difference from healthy plants						
400–700		–0.0007 <sup>a</sup>	0.0009 <sup>a</sup>	–0.0034	–0.0042	–0.0250
700–1350		0.0254	0.0345	0.0458	0.0543	0.1035
1350–1900		–0.0042	–0.0040	–0.0083	–0.0089	–0.0211
1900–2500		–0.0031	–0.0030	–0.0061	–0.0073	–0.0222
400–500		0.0000 <sup>a</sup>	0.0005	–0.0016	–0.0023	–0.0176
500–600		0.0009	0.0029	–0.0009 <sup>a</sup>	–0.0014 <sup>a</sup>	–0.0221
600–690		–0.0038	–0.0021	–0.0095	–0.0109	–0.0395
750–930		0.0498	0.0699	0.0950	0.1085	0.1902
950–1030		0.0159	0.0206	0.0269	0.0341	0.0751
1040–1130		0.0226	0.0285	0.0383	0.0469	0.0998
1450–1850		–0.0045	–0.0043	–0.0088	–0.0095	–0.0226
2000–2400		–0.0033	–0.0032	–0.0064	–0.0077	–0.0234

<sup>a</sup> The values are statistically insignificant at the 0.05 level. Other values are all statistically significant at the 0.01 level.

the mutual variation of the pixel's similarity on the two rule images. Similar correlation was also observed in Fig. 6b showing the pixel values of the rule image for LB3 and LB4 plants. This indicated that the separation between the healthy plants and the LB2 plants as well as between the LB3 and LB4 plants was rather weak. Fig. 6c and d display the similarity of pixels in the rule image spaces of the healthy and the LB3 plants as well as LB4 plants. Since the pixels were scattered in the two-dimensional space, there was no significant spectral correlation between healthy and diseased plants at LB3 and LB4 ( $r = 0.2483$  and  $0.1836$ , respectively in Fig. 6c and d). Therefore, using hyperspectral image analysis for late blight on tomatoes, we were able to separate the healthy plants from the LB3 and LB4 plants although it is difficult to discriminate between the LB3 and LB4 plants, and between healthy and LB2

plants. The results from preliminary view to the classification rule images confirmed our conclusion from the field collected spectral analysis. Similar conclusions were found from statistical analyses of the field spectra (Zhang et al., 2002).

#### 4.5. Identification of the disease in the hyperspectral image

Given the above analysis, we focus the following analysis with the hyperspectral image on identification of the healthy and the LB3 and LB4 plants in the field and kept the four clusters as references. Fig. 7a presents the subset image for the classification and Fig. 7b displays the results of clear identification of the LB3 and LB4 plants from the healthy plants or lightly diseased plants. The close spectral features

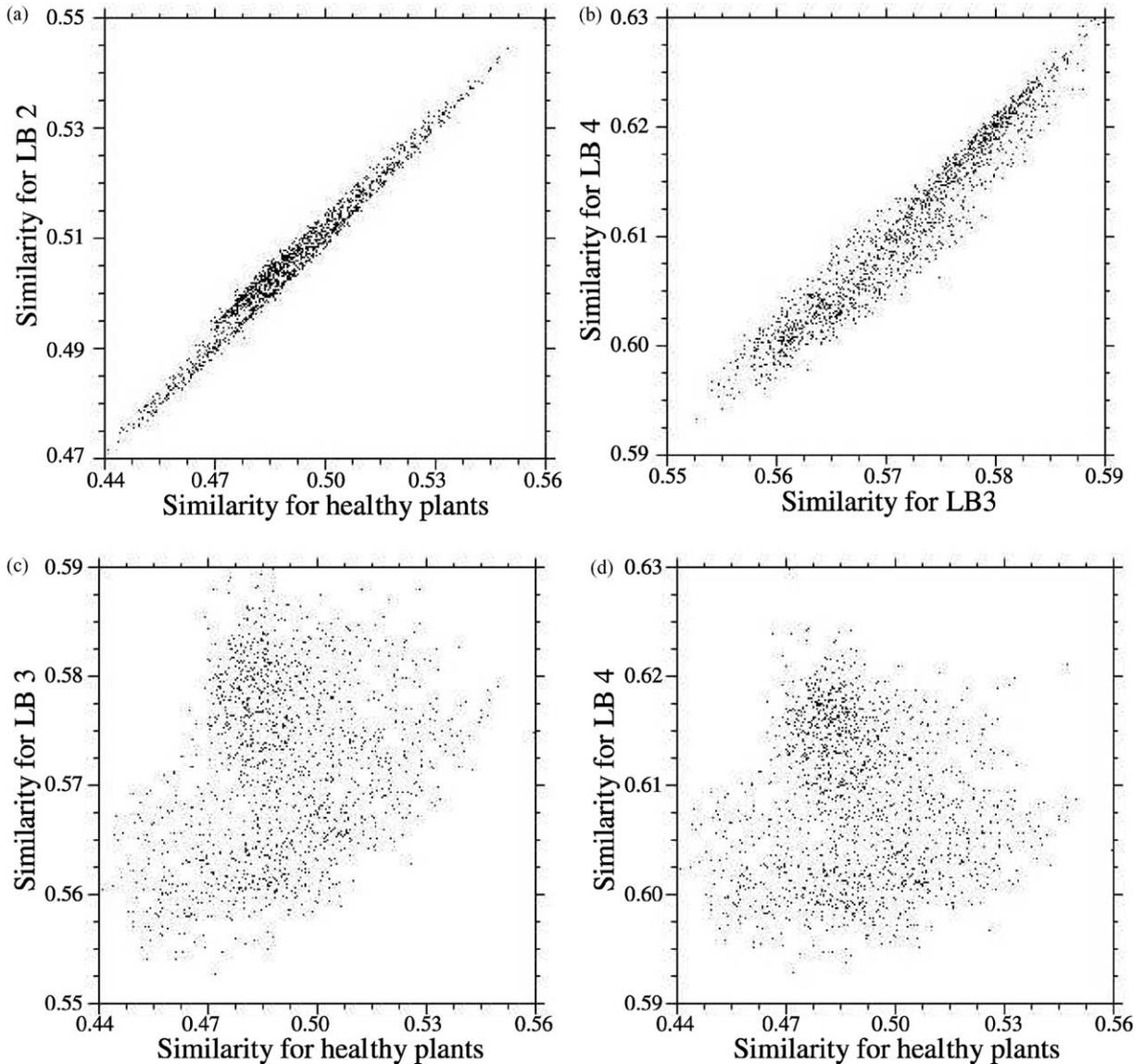


Fig. 6. Similarity between the rule image for each of the categories of tomato plants, illustrating the correlation of the similarity between rule images for (a) healthy plants and infected ones at stage 2, (b) infected plants at stages 3 and 4, (c) healthy plants and infected ones at stage 3, and (d) healthy plants and the infected ones at stage 4.

between healthy and diseased LB2 plants (Fig. 6a), and between diseased LB3 and LB4 plants (Fig. 6b) made it difficult to separate them though some spectral differences were observed in the NIR region 0.7–0.9  $\mu\text{m}$ . Therefore, the image classification only allowed us to clearly identify two categories of the tomato plants: healthy or light diseased plants from diseased LB3

and LB4 plants. Fig. 7b showed the areas with severe infections, corresponding to the original image.

The image analysis also indicated that the southern part of the field had more disease than the northern part. This probably was due to weather-related topography. After the first plant in the field was infected, inoculum would be disseminated by the wind. The

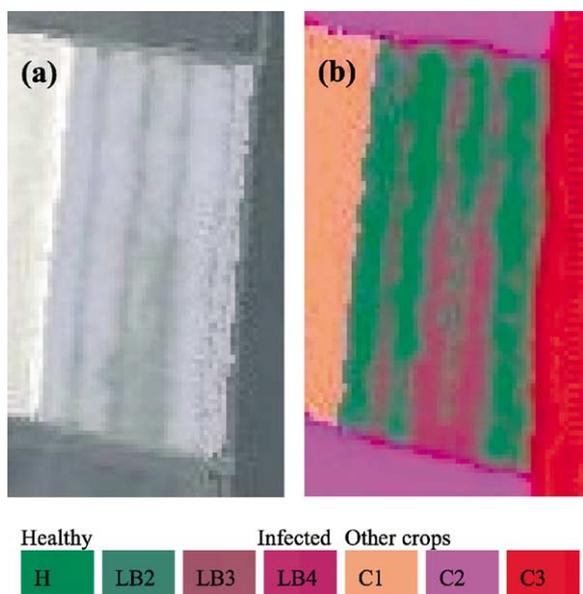


Fig. 7. Results of classification to the hyperspectral image for identification of infected plants from the healthy ones. (a) The pseudo color composite of the image, (b) the classified results of the image for identification of the diseased plants. Note: H is healthy plants; LB2, LB3, and LB4 are infected plants at stages 2, 3, and 4, respectively; and C1, C2 and C3 are other crops respectively in the surrounding fields.

valley topography results in moist winds that usually blow towards the southeast during the day and blow backwards at night with much less intensity. Though some infection could be seen in the eastern strip, it was too faint for image analysis. However, the results from aerial image analysis were consistent with the field investigation. Therefore, we are confident that the aerial hyperspectral images can be used to successfully map the moderate-to-severe diseased plants, and to distinguish them from healthy plants or lightly diseased plants.

Economic loss of tomato production due to late blight often occurs when infection reaches LB3 or above. If farmers could map the disease at the field level using hyperspectral images, they could use an interpreted image as a guide to spray pesticide accordingly to prevent the spread of the disease. This would increase the effectiveness of disease control and decrease the possibility of fungicide contamination in the environment.

## 5. Conclusion

Hyperspectral images have great potential in detecting disease-stressed canopies and thereby allowing precision disease management. This study developed methodologies to detect diseased plants through a spectra-based classification including MNF and SAM. Field spectra analysis indicated that the most valuable wavelengths were between 0.7 and 0.9  $\mu\text{m}$  for remote sensing of the late blight on tomatoes. In the image analysis, after MNF transformation, 28 bands were identified as signal-dominant bands that were valuable for separation of the healthy plants from the diseased plants. High correlations between the healthy plants and the lightly diseased LB1 and LB2 plants and between the LB3 and LB4 plants were found, indicating difficulty in separating the categories. However, healthy plants and lightly diseased plants were successfully separated from the moderately and severely diseased plants. Comparison of plants ranked for disease severity and image analysis also indicated that pathogen-induced lesions could be visualized. A statistical test for significance indicated that both image and field spectra showed consistent results with the informative bands in late blight diseased tomatoes. To conclude, we successfully demonstrated that hyperspectral remote sensing using bands mainly in the range of 0.7–0.9  $\mu\text{m}$  was effective in detecting tomatoes late blight disease. Although we believe the methodologies developed in this study for image analysis can be used to distinguish multiple disease types, further research and field work are required. However, under summer weather conditions in California, late blight is the main disease of concern in the tomato industry. This study will provide the information to tomato farmers and industry for direct applications to late blight precision management. No doubt, hyperspectral remote sensing could be used to monitor disease on large-scale farms, and it would become more common in the future, especially when images are commercially available and appropriate rapid image processing is possible.

## Acknowledgements

The authors would like to acknowledge Professor Alfred Stein, Dr. G. van der Heijde, and the

anonymous reviewer for their editing and the valuable comments to improve the manuscript. We would also like to thank Professor Lynn Epstein of the Department of Plant Pathology at the University of California, Davis for her valuable comments. We would like to thank Carla Thomas, a field plant pathologist from Western Farmer Service for identifying the field disease stages for the study. We would also like to thank Zeneca Ag Product for partial financial support in the earlier stage of the project.

## References

- Adams, M.L., Philpot, W.D., Norvell, W.A., 1999. Yellowness index: an application of spectral second derivatives to estimate chlorosis of leaves in stressed vegetation. *Int. J. Remote Sens.* 20 (18), 3663–3675.
- Barducci, A., Pippi, I., 1997. Environmental monitoring of the Venice algon using MIVIS data. *Int. Geosci. Remote Sens. Symp. (IGARSS)* 2, 888–890.
- Bianchi, R., Cavalli, R.M., Fiumi, L., Marino, C.M., Pignatti, S., 1999. Airborne MIVIS hyperspectral imaging spectrometer over natural and anthropic areas. In: *Proceedings of the Fourth International Airborne Remote Sensing Conference and Exhibition/21st Canadian Symposium on Remote Sensing*, vol. 1, Ottawa, Ont., Canada, pp. 337–344.
- Blackburn, G.A., Steele, C.M., 1999. Towards the remote sensing of matorral vegetation physiology relationships between spectral reflectance, pigment, and biophysical characteristics of semiarid bushland canopies. *Remote Sens. Environ.* 70 (3), 278–292.
- Boardman, J.W., 1993. Automated spectral unmixing of AVIRIS data using convex geometry concepts. *Summaries of Fourth JPL Airborne Geoscience Workshop*, vol. 1. JPL Publication 93-26, JPL, Pasadena, CA, USA, pp. 11–14.
- Boardman, J.W., Kruse, F.A., 1994. Automated spectral analysis: a geologic example using AVIRIS data, north grapevine mountains, Nevada. In: *Proceedings of Tenth Conference on Geologic Remote Sensing*, vol. 1, Environmental Research Institute of Michigan, Ann Arbor, MI, USA, pp. 407–418.
- Boegh, E., Soegaard, H., Broge, N., Hasager, C.B., Jensen, N.O., Schelde, K., Thomsen, A., 2002. Airborne multispectral data for quantifying leaf area index, nitrogen concentration, and photosynthetic efficiency in agriculture. *Remote Sens. Environ.* 81, 179–193.
- California Department of Pesticide Regulation (CDPR), 2000. California Pesticide Use Report (PUR) Database.
- Carter, G.A., Knapp, A.K., 2001. Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. *Am. J. Bot.* 88 (4), 677–684.
- Carter, G.A., Cibula, W.G., Miller, R.L., 1996. Narrow-band reflectance imagery compared with thermal imagery for early detection of plant stress. *J. Plant Physiol.* 148, 516–523.
- Center for the Study of Earth From Space (CSES), 1999. ATmosphere REMoval Program (ATREM) User's Guide, Version 3.1. Center for the Study of Earth From Space (CSES), Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder.
- Christ, B.J., Petersen, G., Ressler, L., Warmer, E., 2000. Early detection of potato late blight using hyperspectral remote sensing. <http://cipm.ncsu.edu/cipmprojects/Reports/00Reports1/CHRISTB2000S.html>.
- Clark, M.S., Ferris, H., Klonsky, K., Lanini, W.T., van Bruggen, A.H.C., Zalom, F.G., 1998. Agronomic, economic, and environmental comparison of pest management in conventional and alternative tomato and corn systems in northern California. *Agric. Ecosyst. Environ.* 68, 51–71.
- Datt, B., 1998. Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a + b and total carotenoid content in eucalyptus leaves. *Remote Sens. Environ.* 66, 111–121.
- Datt, B., 1999. A new reflectance index for remote sensing of chlorophyll content in higher plants: tests using eucalyptus leaves. *J. Plant Physiol.* 154, 30–36.
- Dawson, T.P., Curran, J., 1998. New technique for interpolating the reflectance red edge position. *Int. J. Remote Sens.* 19 (11), 2133–2139.
- Deguise, J.C., Staenz, K., 1999. Agricultural applications of airborne hyperspectral data: weed detection. In: *Proceedings of the Fourth International Airborne Remote Sensing Conference and Exhibition/21st Canadian Symposium on Remote Sensing*, vol. 2, Ottawa, Ont., Canada, pp. 352–358.
- ENVI, 1999. The Environment for Visualizing Images (ENVI): User's Guide, Version 3.2. Research System Institute, USA, 864 pp.
- Fraser, R.N., 1998. Hyperspectral remote sensing of turbidity and chlorophyll a among Nebraska Sand Hills lakes. *Int. J. Remote Sens.* 19 (8), 1579–1589.
- Gitelson, A.A., Merzlyak, M.N., 1994. Spectral reflectance changes associated with autumn senescens of *Aesculus hippocastanum* L. and *Acer Platanoides* L. leaves: spectral features and relation to chlorophyll estimation. *J. Plant Physiol.* 143, 286–292.
- Gitelson, A.A., Merzlyak, M.N., 1996. Signature analysis of leaf reflectance spectra: algorithm development for remote sensing of chlorophyll. *J. Plant Physiol.* 148, 494–500.
- Gitelson, A.A., Merzlyak, M.N., 1997. Remote estimation of chlorophyll content in higher plant leaves. *Int. J. Remote Sens.* 18, 2691–2697.
- Gitelson, A.A., Merzlyak, M.N., 1998. Remote sensing of chlorophyll concentration in higher plant leaves. *Adv. Space Res.* 22 (5), 689–692.
- Gitelson, A.A., Merzlyak, M.N., Lichtenthaler, H.K., 1996. Detection of re edge position and chlorophyll content by reflectance measurements near 700 nm. *J. Plant Physiol.* 148, 501–508.
- 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. *IEEE Trans. Geosci. Remote Sens.* 26 (1), 65–74.
- Goetz, A.F.H., Vane, G., Solomon, J.E., Rock, B.N., 1985. Imaging spectrometer for Earth remote sensing. *Science* 228 (4704), 1147–1153.
- Guenther, K.P., 1990. Vegetation Stress Monitoring by Fluorescence. European Space Agency—Special Publication (ESA SP), ESTEC, Noordwijk, The Netherlands, pp. 135–142.

- Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P.J., Dextraze, L., 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sens. Environ.* 81, 416–426.
- Harsanyi, J.C., Chang, C.I., 1994. Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach. *IEEE Trans. Geosci. Remote Sens.* 32 (4), 779–785.
- Holden, H., Ledrew, E., 1998. Spectral discrimination of healthy and non-healthy corals based on cluster analysis, principal components analysis and derivative spectroscopy. *Remote Sens. Environ.* 65 (2), 217–224.
- Jia, X., Richards, J.A., 1999. Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification. *IEEE Trans. Geosci. Remote Sens.* 37 (1), 538–542.
- Knapp, A.K., Carter, G.A., 1998. Variability in leaf optical properties among 26 species from a broad range of habitats. *Am. J. Bot.* 85 (7), 940–946.
- Lelong, C.C.D., Pinet, P.C., Poilvé, H., 1998. Hyperspectral imaging and stress mapping in agriculture: a case study on wheat in Beauce (France). *Remote Sens. Environ.* 66 (2), 179–191.
- Lichtenthaler, H.K., Lang, M., Sowinska, M., Heisel, F., Mische, J.A., 1996. Detection of vegetation stress via a new high resolution fluorescence imaging system. *J. Plant Physiol.* 148, 599–612.
- Leblanc, S., Chen, M.J., Miller, R.J., Freemantle, J., 1999. Compact airborne spectrographic imager (CASI) used for mapping leave area index (LAI) of cropland. In: Proceedings of the Fourth International Airborne Remote Sensing Conference and Exhibition/21st Canadian Symposium on Remote Sensing, vol. 1, Ottawa, Ont., Canada, pp. 297–303.
- Nolin, A.W., Dozier, J., 2000. Hyperspectral method for remotely sensing the grain size of snow. *Remote Sens. Environ.* 74 (2), 207–216.
- Rush, C., 2002. Prediction, detection and quantification of plant diseases with remote sensing technologies. Precision Agriculture Initiative for Texas High Plains, 2002 Annual Comprehensive Report. <http://precag.tamu.edu/files/rep42/2002PARReportFormandLetter.pdf>.
- Taranik, J.V., Mouat, D.A., Elvidge, C.D., 1993. Hyperspectral technology for geologic applications, better understanding of earth environment. *Int. Geosci. Remote Sens. Symp. (IGARSS)* 2, 917–920.
- Treitz, P.M., Howarth, P.J., 1999. Hyperspectral remote sensing for estimating biophysical parameters of forest ecosystems. *Prog. Phys. Geogr.* 23 (3), 359–390.
- USDA, 1997. Agricultural Statistics, 1997. Washington, DC, USA.
- Wisler, G.C., Duffus, J.E., 2000. A century of plant virus management in the Salinas Valley of California 'East of Eden'. *Virus Res.* 71, 161–169.
- Wilson, T., Felt, R., 1998. Hyperspectral remote sensing technology (HRST) program. *IEEE Aerosp. Appl. Conf. Proc.* 5, 193–200.
- Zhang, M., Liu, X., Oneill, M., 2002. Spectral Discrimination of *Phytophthora infestans* infection on tomatoes based on principal component and cluster analyses. *Int. J. Remote Sens.* 23 (6), 1095–1107.