Precision Agriculture, 6, 489–508, 2005

Remote Sensed Spectral Imagery to Detect Late Blight in Field Tomatoes

MINGHUA ZHANG mhzhang@ucdavis.edu
ZHIHAO QIN
XUE LIU

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

Abstract. Late blight, caused by the fungal pathogen *Phytophthora infestans*, is a disease that quickly spreads in tomato fields under suitable weather conditions and can threaten the sustainability of tomato farming in California, USA. This paper explores the applicability of remotely sensed images to detect disease spectral anomalies for precision disease management. We used the indices approach and generated a 5-index image that we used to identify the disease in tomato fields based on information from field-collected spectra and linear combinations of the spectral indices. Field results indicated that we were able to identify five clusters in the image space with small overlaps of a few clusters. Using the identified 5-cluster scheme to classify the tomato field images, we were able to successfully separate the diseased tomatoes from the healthy ones before economic damage was caused. Hence, the method based on a 5-index image may significantly enhance the capability of multispectral remote sensing for disease discrimination at the field level.

Keywords: remote sensing, late blight, plant disease, near infrared, image feature space, visualization analysis

Introduction

The concept of precision agriculture is widely accepted for site-specific farm management to effectively control pests and reduce farm costs/environmental impacts (Evans *et al.*, 2003; Fitzgerald *et al.*, 2004). Late blight, caused by the fungal pathogen *Phytophthora infestans*, is a disease that quickly spreads in tomato fields under suitable weather conditions and can threaten the sustainability of tomato farming in California. This paper explores the applicability of remotely sensed images to detect disease spectral anomalies for precision disease management.

California produces 94% of the United States' (USA) processing tomatoes (USDA, 2002) and close to 74% of fresh market tomatoes. To assure sustainable large-scale crop production, California tomato growers depend heavily on pesticide applications, especially for fungicides, because currently there is no curative fungicide available. The treatment window for controlling fungal diseases such as late

blight, *Phytophthora infestans*, is one week. Therefore, the majority of growers use calendar-scheduled pesticide sprays to protect the fields from diseases. However, the large amount of pesticide use not only causes potential risks to the environment and ecosystem but also increases the farming cost (CDPR, 2002). This large-scale farming requires timely detection of diseases for precision pest management to avoid overuse of pesticides. We used the indices approach and generated a 5-index image to identify the disease in tomato fields based on information from field-collected spectra and linear combinations of the spectral indices.

Spectral measurements of crop diseases offer valuable information on the state of the crop for precision agriculture. When repetitive large-scale estimates are required, remote sensing is probably the only feasible method for obtaining this data (Steven and Clark, 1990; McDonald *et al*., 1998; Fitzgerald *et al*., 2004). Literature indicated that the spectral reflectance of green vegetation in the red band (0.6–0.7 μm) is most sensitive to leaf chlorophyll and pigment contents while the near infrared (NIR) band (0.7–0.9 μm) is most sensitive to biomass (Thomas and Oerther, 1972; Toler *et al*., 1981; Blazquez and Edwards, 1983; Kurschner *et al*., 1984; Blakeman, 1990). Moreover, the quantity of chlorophyll content of green plants directly correlates to the healthiness of the plants. Biologically, plants are in various stress stages when under unfavorable growing conditions. When tomato plants are infected by *Phytophthora infestans*, the chlorophyll content will decrease. Generally, the stressed plants have lower absorption of red light and higher absorption of NIR radiation (Lillesand and Kiefer, 1994; Guyot, 1990; Hatfield and Pinter, 1993). These spectral characteristics of green plants have been used to evaluate the stresses of various crops (Chapelle and Kim, 1992; Shibayama *et al*., 1993; Zhang *et al*., 2002; Fitzgerald *et al*., 2004).

Phytophthora infestans (Agrios, 1997) is an aggressive fungus that spreads quickly in tomato fields. Once the tomatoes are infected, the symptoms of late blight first appear on leaves, which will gradually change color from green to yellow, then the fungus will infect the stems and fruits. Growers will face economic loss if the early infection is not properly treated. Developing a remote sensing method to detect late blight at field level would be of great value to the tomato industry and the state economy.

Due to the difference in light response to various biological features, scientists have used an indicator approach for identifying plant stresses (Tucker, 1979; Wiegand *et al*., 1991; Zhang *et al*., 1997; Gitelson and Merzlyak, 1998). The existing indicators are mostly derived from various combinations of red and NIR spectral bands to evaluate biomass and detect water deficit and diseases (Tucker, 1979; Blazquez and Edwards, 1983; Huete, 1988; Shibayama *et al*., 1993; Bausch, 1993; Huete *et al*., 1994). However, remote sensing indices have not been evaluated for monitoring and evaluating tomato late blight.

The objective of the study was to develop a method for application of multispectral remote sensing to detect the diseased tomato plants. Several possible indices derived from red and NIR bands of the multispectral remote sensing data were evaluated for discriminating the diseased tomato plants. The method based on the effective indices was developed to explore the potential applications of the multispectral remote sensing to detect the diseased tomato plants. Therefore, our analysis in the study mainly involved the following three parts: (1) field spectral

analysis to determine whether spectral features of diseased plants can be differentiated, (2) numeric analysis to determine the best combination of red and NIR bands for effective separation of diseased plants and (3) validation analysis to test the developed image identification method for late blight on tomatoes with available ground truth information.

Materials and methods

Field spectral data collection

Field spectra of tomato plants were collected from four fields near King City in Salinas Valley, California. A handheld spectrometer GER 2600 (Geophysical & Environmental Research Corporation, Milbrook, New York USA) with 10 nm spectral sensitivity was used to collect the spectral data from 1 m above tomato canopies at a field of view of 23° at various infection stages. Altogether, 66 spectral samples were obtained in September 1998 under clear sky conditions.

Four infection stages were assessed (Agrios, 1997): Stage 1 (LB1)—one lesion on one or two canopy leaves, Stage 2 (LB2)—one lesion on more than two canopy leaves (Figure 1a), Stage 3 (LB3)—two lesions on one to many canopy leaves (Figure 1b), Stage 4 (LB4)—two lesions on over half the canopy leaves. The selected rating for the disease infection was the common rating for the disease assessment in California. Economic impact of the disease started at stage 3 of infections. Among the 66 collected samples, 22 were for healthy tomato plants, 11 for LB1, 12 for LB2, 17 for LB3, and 4 for LB4 (Zhang *et al.*, 2002).

Multispectral data acquisition and tomato fields

ADAR (Airborne Data Acquisition and Registration, using system 5500 airborne sensor from Positive System, Inc. Idaho, USA, http://www.possys.com) broadband system was used to acquire the multispectral image of four broad TM bands (blue, green, red and near infrared) and 1 m pixel size. Two images were obtained for two fields: a late blight diseased tomato field in Yolo County, California on August 15, 1999 and a late blight diseased tomato field in Salinas Valley on September 17, 1997.

Healthy plants were dominant in both fields. The disease infection stages varied from LB1 to LB4 at the time of image acquisition for the two fields. These selected tomato fields were irrigated and tomatoes were planted in rows with identical spacing. Usually, late blight occurred on tomatoes when the crop was at early fruiting phenological stage. The soil types of the fields ranged from loam to clay loams in Yolo field to clay loams to sandy loams in Salinas field. Because there were almost full canopy covers at the time of sampling, soil effects on each image pixel can be minimal. In addition, the soil covers in the ground were similar from each sampling site, the soil contribution to each pixel can be viewed as almost equal share.

With multispectral remote sensed images, one may think that the discrimination of diseased plants can be done through comparison of image indices or pixel DN values

Figure 1. (a) Tomato leaves with late blight (Courtesy R.E. Stall) at infection stage 2. (b) Tomato leaves with late blight infection stage 3.

with the known class (infection status). In fact, this represents to directly identify the diseased plants in images and this direct identification can only be done if the spectral features of diseased plants are distinctly different from the healthy plants. However, the real world is much more complicated than this. In many cases, the

spectral features of diseased plants are usually not so obvious in the images for direct identification because of their morphological similarities. For example, it is very difficult to visually identify the diseased plants on the image shown in Figure 2a and b. Figure 2a shows the false color composite image of the diseased tomato field in Yolo County. Redder color associated with healthier plants appears at the right upper part of the field, while brighter light yellow green color associated with less dense crop canopy appears at the left lower corner of the field (Figure 2a). Moreover, different colors within the strips show that some lines of tomato plants within the same strip are denser biomass than others. This might be attributed to the slight difference in water and nutrient distribution in the field. The black line in the middle of the field indicates an irrigation ditch. The dark spots in the field were the diseased LB3 or LB4 plants. Some dots were the bare soil or dead plants caused by the disease as observed in the field during sampling. The diseased plant pixels were sparsely distributed in the field (Figure 2a). The size of these spots on the canopies generally ranged over a few meters, covering a few pixels in the image.

Figure 2b shows the false color composite image of the diseased tomato field in Salinas Valley. The circled locations in the field were identified as LB3 or LB4 in the ground observation during the imaging. The darker lines with black spots were the water ditches for irrigation of the field. The black spots on the ditches were the bare areas. The bright strip in upper part of the image was where a disease resistant variety was planted. Since direct identification of diseased plants is very difficult on the image such as those shown in Figure 2a and b, we have to develop more applicable method for remote sensing analysis to explore potential precision disease management. In the following, we present the work in developing such remote sensing analysis methods for pest management in tomato farming.

Image analysis methods

Five feature vector indices development. Spectral reflectance of vegetation canopy in some wavelength ranges such as infrared is obviously related to the health of plants (Yang *et al.*, 1988). In our case, the healthy plants have the highest digital number (DN) in the near infrared (NIR) band while the LB4 plants have the lowest DN. The LB1 and 2 plants should have DN values in between the two extremes. Several combinations of red and NIR bands have been commonly developed to relate the spectral reflectance of vegetation canopy to their biological features (Tucker, 1979; Kimes *et al.*, 1981; Huete 1988; Huete *et al.*, 1994).

Based on previous research, a numeric analysis was employed to select the best index from the combinations of red and NIR bands for discriminating healthy and late blight diseased plants. Hence, the difference in DN values of healthy and diseased plants in the NIR band was used as a standard for assessing the effectiveness of these combinations of red and NIR bands. The rule for selecting the most effective indices was that, for a combination of red and NIR bands, if the difference in DN values after transformation of the red and NIR bands was larger than the difference in DN values in the NIR band imagery, this combination of red and NIR bands was selected for enhancing separation between the healthy and diseased tomato plants.

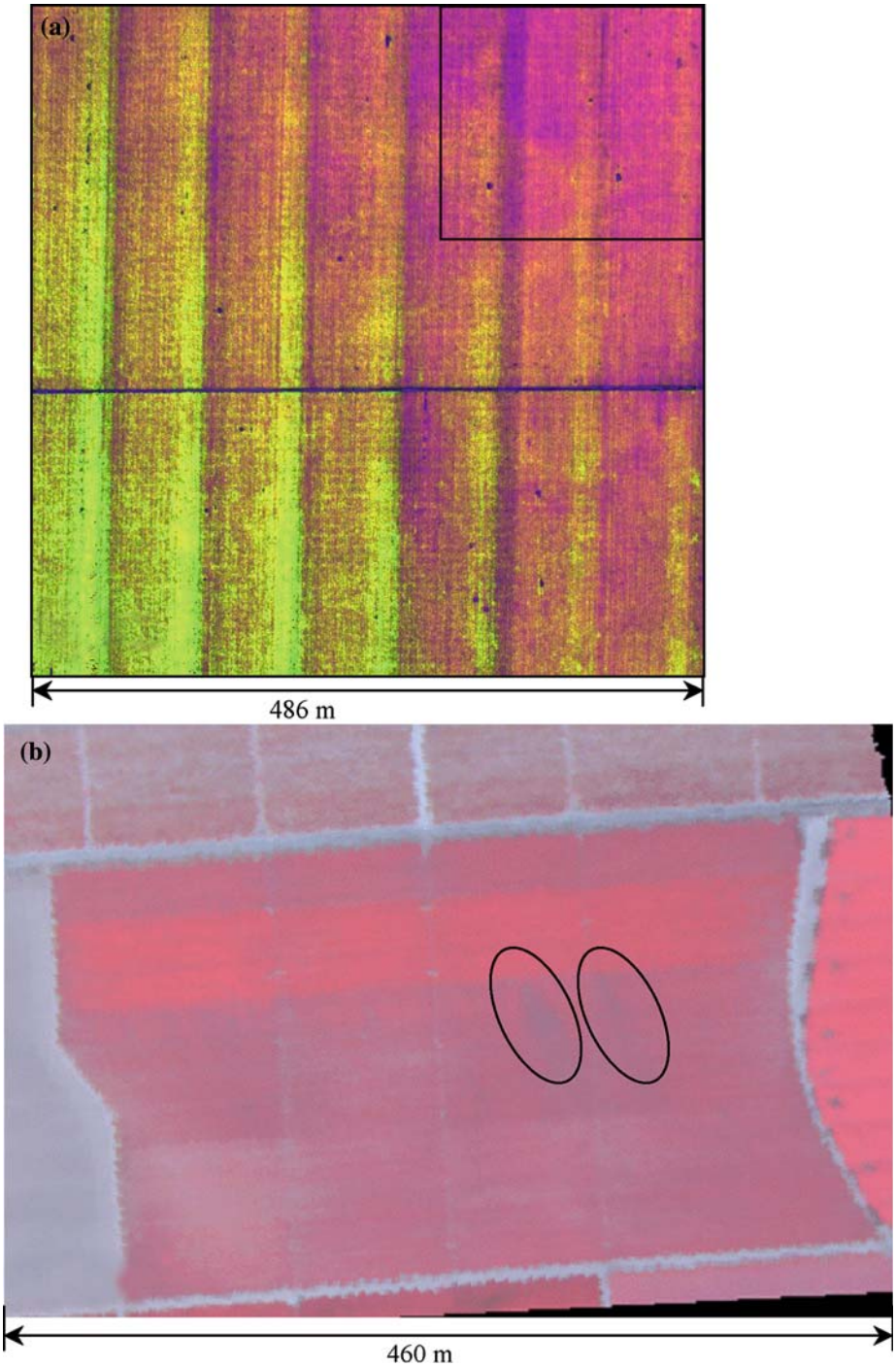

Figure 2. (a) Diseased tomato field (24 ha) in Yolo County, CA imaged in August 1999. (b) Diseased tomato field (16 ha) in Salinas Valley, CA imaged in September 1997.

The symbols used in this numeric analysis are as follows: Y_2, X_2 represent the pixel DNs of healthy plants in the red and NIR bands, and Y_1, X_1 represent the pixel DNs of diseased plants in the red and NIR bands, respectively. The spectral separation ability (SP) between the healthy and diseased plants in the NIR band can be defined as

$$\mathrm{SP} = X_2/X_1. \tag{1}$$

Since several combinations of the red and NIR bands are used for the analysis, $\mathrm{SP}_{\mathrm{new}}$ is referred to as the separation ability after transformation. Therefore, using the combination NIR/R to transform the red and NIR images, the pixel DNs in the output image will be X_2/Y_2 for pixel value of healthy plants and X_1/Y_1 for late blight diseased plants, which results in:

$$\mathrm{SP}_{\mathrm{new}} = \frac{X_2/Y_2}{X_1/Y_1} = (X_2/X_1)^*(Y_1/Y_2). \tag{2}$$

Since $(Y_1/Y_2) > 1$, we can logically expect that

$$\mathrm{SP}_{\mathrm{new}} > SP. \tag{3}$$

Therefore, the ability to separate healthy plants from diseased ones improves after the transformation.

Similarly, for the combination , the product is

$$\mathrm{SP}_{\mathrm{new}} = \frac{\sqrt{X_2/Y_2}}{\sqrt{X_1/Y_1}} = \sqrt{(X_2/X_1)^*(Y_1/Y_2)}. \tag{4}$$

Since, $\sqrt{(Y_1/Y_2)} > 1$ and $\sqrt{(X_2/X_1)} < (X_2/X_1)$, we cannot determine whether $\mathrm{SP}_{\mathrm{new}} > \mathrm{SP}$. Thus, this combination does not provide conclusive results about the effectiveness of the transformation.

For the combination NIR−R,

$$\mathrm{SP}_{\mathrm{new}} = \frac{(X_2 - Y_2)}{(X_1 - Y_1)}. \tag{5}$$

Since $X_1Y_2 < X_2Y_1$, then $(X_2-\mathrm{Y}_2)/(X_1-\mathrm{Y}_1) > X_2/X_1$. This leads to $\mathrm{SP}_{\mathrm{new}} > \mathrm{SP}$, indicating that the transformation is able to improve the separation ability.

For the combination NIR + R, the $\mathrm{SP}_{\mathrm{new}}$ can be computed as

$$\mathrm{SP}_{\mathrm{new}} = \frac{(X_2 + Y_2)}{(X_1 + Y_1)}. \tag{6}$$

If it is expected $\mathrm{Sp}_{\mathrm{new}} > \mathrm{SP}$, then $(X_2+Y_2)/(X_1+Y_1) > X_2/X_1$, which leads to $X_1Y_2 > X_2Y_1$. However, this relation is clearly incorrect. Therefore, the assumption of $\mathrm{SP}_{\mathrm{new}} > \mathrm{SP}$ is not always true. In other words, the separation ability after this transformation cannot be further improved.

For the combination (NIR−R)/(NIR + R), new separation ability is as follows:

$$\mathrm{SP}_{\mathrm{new}} = \frac{(X_2 - Y_2)/(X_2 + Y_2)}{(X_1 - Y_1)/(X_1 + Y_1)}.$$

Since all the variables X_i and Y_i are larger than 0, $Y_1 > Y_2$ and $X_2 > X_1$, then $X_1X_2(Y_1 - Y_2) + Y_1Y_2(X_2 - X_1) + Y_1X_2^2 - Y_2X_1^2 > 0$. Thus, $((X_2 - Y_2)/(X_2 + Y_2))/((X_1 - Y_1)/(X_1 + Y_1)) > X_2/X_1$, which is $\mathrm{SP}_{\mathrm{new}} > \mathrm{SP}$. The ability to separate after this transformation can further be enhanced.

Transformation with the combination (NIR + R)/(NIR−R) will lead to

$$\mathrm{SP}_{\mathrm{new}} = \frac{(X_2 + Y_2)/(X_2 - Y_2)}{(X_1 + Y_1)/(X_1 - Y_1)}. \tag{8}$$

Provided that variables X_i and Y_i are larger than 0, $Y_1 > Y_2$ and $X_2 > X_1$, one can derive that $X_1(X_2 + Y_2) < X_2(X_1 + Y_1)$ and $(X_1 - Y_1) < (X_2 - Y_2)$. Thus, $\mathrm{SP}_{\mathrm{new}} < \mathrm{SP}$, indicating that this combination is not able to improve the separation ability after this transformation. Using the combination $\sqrt{(\mathrm{NIR} + \mathrm{R})/(\mathrm{NIR\text{-}R}) + 0.5}$ to transform, the ability to separate the difference in the output image will be

$$\mathrm{SP}_{\mathrm{new}} = \frac{\sqrt{0.5 + (X_2 - Y_2)/(X_2 + Y_2)}}{\sqrt{0.5 + (X_1 - Y_1)/(X_1 + Y_1)}}. \tag{9}$$

Under the conditions $X_i > 0$, $Y_i > 0$, $Y_1 > Y_2$ and $X_2 > X_1$, one is not able to prove that X_1Y_1 must be greater or less than X_2Y_2. Therefore, it is not certain if the $\mathrm{SP}_{\mathrm{new}}$ is greater than SP after the transformation.

The above analysis indicates that, for the tomato disease DN values, only the indices of the simple ratio NIR/R, the simple difference NIR−R, and the normalized difference vegetation index NDVI are able to enhance the differentiation between healthy and diseased plants after the transformation. Therefore, together with the red and near infrared bands (R, NIR) and the three selected indices, a new feature space is formed with the five indices expressed as a feature vector:

$$\text{Feature Vector} = (\mathrm{R},\ \mathrm{NIR},\ \mathrm{NIR/R},\ \mathrm{NIR\text{-}R},\ (\mathrm{NIR\text{-}R})/(\mathrm{NIR} + \mathrm{R})). \tag{10}$$

to be utilized in the image analyses to demonstrate the usefulness of multispectral remote sensing images in monitoring tomato disease.

Visualization, minimum noise fraction transformation and pixel purity index. In order to identify the diseased plants from the healthy ones in the image, one has to determine the criterion for comparison of the pixels with various health levels. Visualization analysis is one of the methods in remote sensing to identify the pure pixels (endmembers) representing the degrees of the plant health. Therefore, after the above feature vectors were computed, we then used the *n*-dimensional visualization

function embedded in the Environment for Visualization Images (ENVI, 1999) software together with Minimum Noise Fraction (MNF) transform and Pixel Purity Index (PPI) to identify spectra classes of diseased tomato plants.

The principle of the *n*-dimensional visualization was that the spectrum of every pixel in the multi-band images could be thought of as points in an *n*-dimensional scatter-plot, where n was the number of indices. The coordinates of the points in the *n*-space consisted of "*n*" values that were simply the DN values in each index for a given pixel. MNF transform is essentially two cascaded principal components transformations. The first transformation, based on an estimated noise covariance matrix, de-correlates and rescales the noise in the data. The second step is a standard principal components transformation of the noise-whitened data. By using only the coherent portions, the noise is separated from the data, thus improving spectral processing results. PPI, however, finds the most "spectrally pure" pixels in multi-spectral images. The most spectrally pure pixels typically correspond to mixing endmembers. The PPI is calculated by repeatedly projecting *n*-dimensional scatter-plots onto a random unit vector. The procedure is iterative. For each step, the extreme pixels in each projection and those falling into the ends of the unit vector are recorded in the analysis. At the same time, total number of iterations is also noted. The iteration will be stopped when the pure pixels are identified for the clusters. Using the pure pixels' spectra in the remote sensing bands considered, we can perform a comparison of all the pixels in the image for classification to separate the diseased plants from healthy ones.

After projecting into a two-dimensional plane for visualization, the distribution of these points in *n*-space could be used to estimate the number of spectral endmembers and their pure spectral signatures (ENVI, 1999). Therefore, along with using the *n*-dimensional visualization function of ENVI, MNF and PPI were also applied to locate, identify and cluster the purest pixels and most extreme spectral responses in a dataset.

The classification thresholds were determined interactively through visually representing and dynamically observing all the pixels in a feature space. The corresponding pixels of each cluster generated by these thresholds were then used as regions of interest (ROI) to calculate the mean image spectrum for each cluster.

Although the soils could be treated as a uniform background for the images, the selected images may still be influenced by soil because the field may or may not be fully covered by the tomato canopies. Therefore, soil background was included as a candidate class during image analysis and classification. After conducting a number of different rotations in the visualization, at the end, five candidate classes were considered in the image analysis: C1 to C5. This 5-cluster classification scheme was the best result in our case. Accordingly, the diseased plants were detected from the imagery with the assistance of ground truth. For the effective interpretation of each cluster in connection with the health status of the plants, the mean DN value of each cluster in each band of the new images must be calculated and compared. The key to the interpretation was that the lower mean DN value in index 3, 4, and 5 (Table 1) indicated severe infection. A high mean DN value in these three indices represented the healthy level of the plants. As a contrast, the high mean DN value in index 1

Table 1. The mean DN value order in each index for the 5 clusters for the image of Yolo County, California

Index	Mean DN value order for the 5 clusters
Index 1 (R)	C1 > C4 > C5 > C2 > C3
Index 2 (NIR)	C1 < C4 < C5 < C3 < C2
Index 3 (NIR/R)	C1 < C5 < C4 < C3 < C2
Index 4 (NIR−R)	C1 < C4 < C5 < C2 < C3
Index 5 (NDVI)	C1 < C4 < C5 < C2 < C3

(Table 1) implied severe infection while a low DN value meant healthy plants. Therefore, combining the visualization analysis with the spectral variations of tomato plants in relation to their infection severity, we were able to identify the pixel distribution patterns in the *n*-dimension feature space and interpret the patterns for classifying the images.

Results

Figure 3a shows the average spectral curves for each infection stage. Figure 3b displays the confidence bands for each infection stage in the red and NIR ranges. The largest spectral difference was observed in the NIR range. Figure 3b shows that approximately 20% of the spectral variations for LB3 plants overlapped with that for the healthy plants in the NIR range. Approximately 80% of spectral variations for healthy plants overlapped with that for LB1 plants, and 50% of spectral variations of healthy plants partly covered with LB2 plants. Therefore, the small overlap of spectral variations between LB3 plants and healthy plants indicated that they could be effectively separated in their spectra, while the high percentage of overlap between healthy plants and LB1 and LB2 plants indicated a very low separation.

Severely diseased plants were associated with much higher spectral reflectance in the red range and lower spectral reflectance in the NIR band. Figure 4a and b show the changes of spectral reflectance between visible and NIR ranges for each infection stage. Consequently, these field spectral characteristics might be used for crop disease detection in remote sensing. Late blight diseased plants can be differentiated from healthy tomato plants when the infection stage reaches LB3 and LB4. This observation laid the foundation for our subsequent analyses.

As indicated in Figure 4b, the healthy tomato plants had the highest reflectance, while diseased plants had lower reflectance within the NIR range (0.7–0.9 μm wavelength). In this range, the spectral reflectance of the diseased plants rapidly decreased with the severity level. The lowest reflectance, apart from soils, was found for the LB4 plants. However, healthy tomatoes had the lowest reflectance within the red range (0.6 m to 0.7 μm), while the LB4 plants had the highest (Figure 4a) apart from soils.

The numeric analysis of selecting the better indices indicated that, for the tomato disease spectra, only the indices of the simple ratio NIR/R, the simple difference NIR−R, and the normalized difference vegetation index NDVI were able to enhance

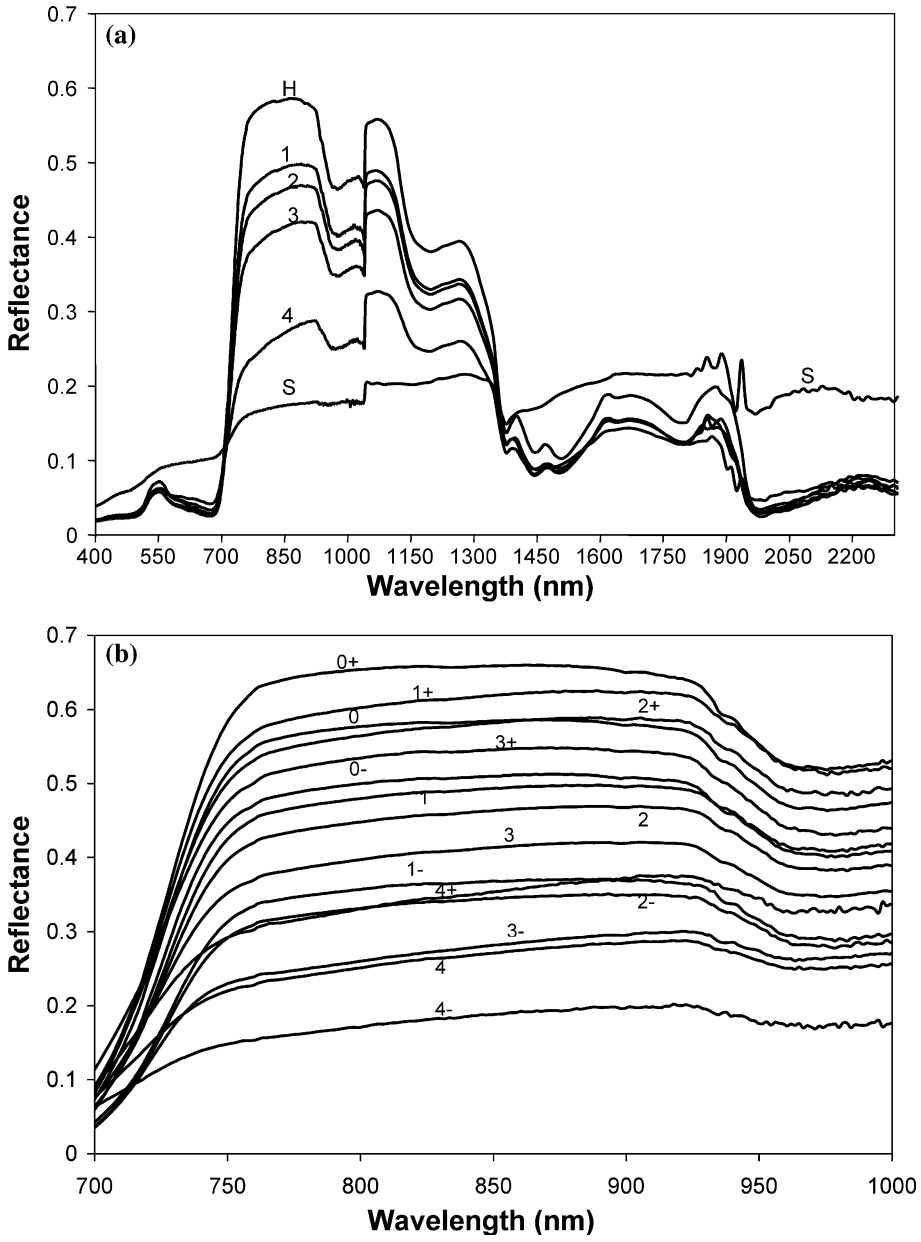

Figure 3. (a) Mean spectra of tomato plant canopies and soil background. (b) The spectral fluctuation ranges of the canopies with ±2 standard deviations. H-healthy plants, 1-infection stage 1(LB1), 2-infection stage 2 (LB2), 3-infection stage 3 (LB3), 4-infection stage 4 (LB4), s-soils.

the differentiation between healthy and diseased plants after the transformation. Therefore, these indices together with red and near infrared bands were used in the image analysis to monitor tomato disease.

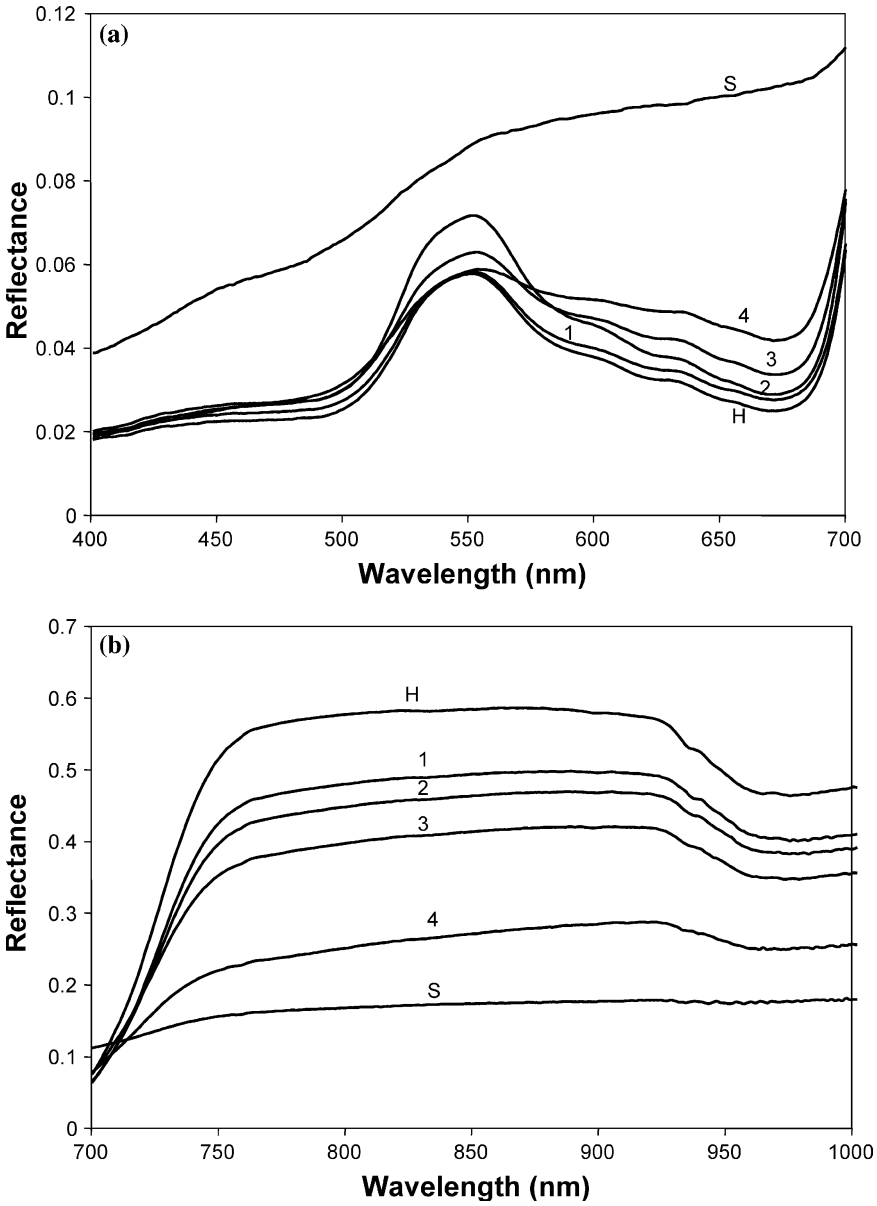

Figure 4. Comparison of healthy canopy spectra with infected ones within (a) the visible range and (b) the near infrared range. H- healthy plants, 1-infection stage 1(LB1), 2-infection stage 2 (LB2), 3-infection stage 3 (LB3), 4-infection stage 4 (LB4), s-soils.

Using the approach described above for the image with a 5-index image (R, NIR, NIR/R, NIR−R, and NDVI), the diseased tomato field in Yolo County was investigated and analyzed. Each index was used as a vector to create a 5-dimensional

feature space. Then we performed visualization analysis of all pixels projected into the space for potential cluster identification**.** Rotating the space to a proper position to give the best cluster identification of the depicted pixels, the 2-dimensional planes were obtained as shown in Figure 5a and b, which indicated that three clusters could be clearly separated: (1) cluster C1; (2) cluster C2; (3) unclassified pixels (the rest, not annotated). Cluster C1 only has a few pixels, whose positions were relatively far from all the other pixels in these projected 2-dimensional planes (Figure 5a). As shown in Figure 5a, after cluster C1 has been identified, the rest pixels were implicitly classified as the other cluster, which might involve several sub-clusters hence need to further rotate (Figure 5b and c) to identify these sub-clusters. Cluster C2 consisted of numerous pixels occupying the upper center of the 2-dimensional plane (Figure 5b). The values of the pixels spread out densely in a triangular style in the plane. Although the pixels of cluster C2 had some overlaps with those scattering in the lower left part of the plane (Figure 5b), they were still very clearly clustering and hence could be distinguished from those designated as the unclassified pixels.

Figure 5c shows the resultant position that provided the best classification of the pixels. Thus, three sub-clusters can be identified in this plane (Figure 5c). Since the general distribution of these unclassified pixels in Figure 5c appeared as a tilted bar, the classification was based on their relative density in the space. According to the density of pixel distribution, three patterns could be identified in Figure 5c in spite of heavy overlay. The left upper part of the bar was identified as a cluster, dubbed as cluster C3. The right lower part of the bar was another cluster, dubbed as cluster C5. Finally, the pixels between the two extremes could be distinguished as another cluster dubbed cluster C4. Although there were some overlaps among the three clusters, the identification of these pixels could be helpful in interpreting the image for disease detection.

After classifying all the pixels in the feature space into their separate clusters, we continued the analysis to interpret the clusters. The purpose was to identify which cluster represented the healthy plants and which cluster represented the infected plants. Using the five clusters, the next step was to differentiate the clusters representing the healthy and diseased plants. This could be done through calculating the mean DN vector of each cluster with reference to the five-index image using the technique of Region of Interest (ROI) (ENVI, 1999). Figure 5d shows the mean DN curves corresponding to the five clusters in the image indices.

The clue for the cluster interpretation could be obtained from analyzing the changes of their relative mean DN value among the five indices of the image (Table 1). The three new indices (NIR−R, NIR/R, NDVI) maintained the same characteristics of the spectra of the healthy and diseased plants as they were in the NIR band, i.e., the pixel DN values in these new indices remaining in the spectral magnitude order: healthy plants > LB3 plants > LB4 plants > soils. Therefore, the five clusters could be in three groups: healthy plants, diseased plants, and soils. Since cluster C1 had the lowest mean DN value in indices 3, 4, and 5, and the highest DN value in index 1 (Table 1), this cluster may be the soils or LB5 plants. As a contrast, cluster C2 had the highest mean DN value in indices 3, 4, and 5, and the lowest mean DN value in index 1 (Table 1). This must be the healthy plants cluster. Cluster C3 had similar spectral features as cluster C2, very high mean DN values in indices 3, 4

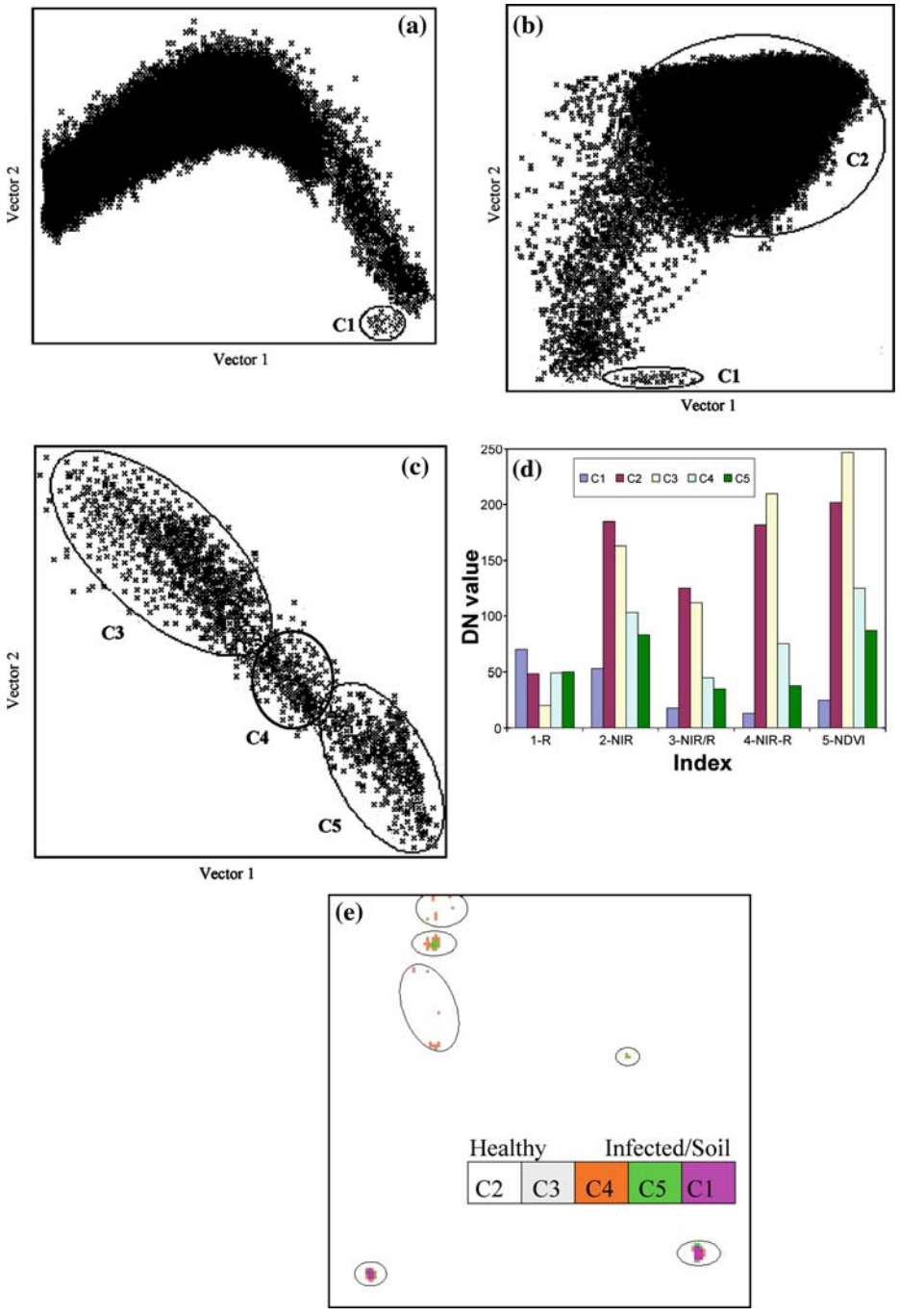

Figure 5. Validation of the method to the tomato field in Yolo County of California, (a) identification of cluster 1, (b) clusters 1 and 2, and (c) sub-clusters 3, 4, and 5 in the 5-dimensional feature space, (d) mean digital number values of the 5 clusters in the 5 index images (indices 1 through 5 refers to R, NIR, NIR/R, NIR−R, and NDVI, respectively, Table 1), (e) results of applying the 5 clusters to classify the subset images marked in (a).

and 5 and low mean DN value in index 1. However, the mean DN values of this cluster in indices 3, 4, and 5 were slightly lower than that of cluster C2. The mean DN value in index 1 of cluster 3 is also slightly higher than that of cluster C2. Thus, cluster C3 could be classified with the LB1 plants. Since the mean DN value of C3 was so close to cluster C2, some healthy plants might also be included in this cluster. Clusters C4 and C5 had a mean DN value much lower in indices 3, 4, and 5 than clusters C2 and C3. Thus, these pixels should be LB3 and LB4 plants. Since the differences of mean DN value in indices 3, 4, and 5 between clusters C4 or C5 and C2 or C3 were large (Figure 5d), it was clear that the approach of using a five feature space in image analysis was able to separate the infected plants from the healthy ones when the disease was stage LB3 or above.

The original image can be classified by exporting the pixels within each cluster into it. Figure 5e shows the results of the image classification to distinguish the diseased plants from healthy ones in the field image. The results in Figure 5e for the diseased plants were consistent with field ground observations at the time of imaging. The LB4 plants were found around the two locations in the field (Figure 5e). Figure 5e also illustrates that the LB3 plants were mainly distributed around the LB4 plants. The healthy plants distributed around the LB3 plants. This identified infection pattern followed the developmental stages of late blight in the field under natural conditions. Since the disease develops along the direction from light to severe damage, it is relatively easy to explain why the observed light diseased plants locate around the severely diseased plants.

Following the same procedure used in the previous image analysis, Figure 6a, b, c, and d illustrate the process of the analysis for the diseased field in Salinas Valley, which indicated that five clusters were the best classification in this newly generated feature space. Figure 6d showed the mean DN curves of the five clusters in the new image that was the basis for the classification of infection severity in this field. The detection results of late blight diseased plants are shown in Figure 6e.

Some pixels in the upper right part of Figure 6a were clearly apart from other pixels. These were assigned as a cluster, C1. Then the space was rotated until other pixels were all assigned. Figure 6b shows a position in which a cluster could be identified: the cluster C2. Although this cluster had some pixels overlapping with those scattering in the upper left part of Figure 6b, the dense distribution of most pixels within the oval space become a cluster. After two clusters were extracted, the remaining scattering pixels in the upper left part of Figure 6b could be identified as an unclassified cluster. After separating the classified pixels of the two clusters out from the space, further rotation of the space was necessary to identify the potential distribution patterns within these unclassified pixels (Figure 6c). Based on the relative density of the pixel distribution, three sub-clusters could be identified even though their boundaries were not as clear as expected. The cluster C3 was composed of the pixels scattering in the upper left part and cluster C5 with the pixels in the lower right part (Figure 6c). The pixels locating between the two extremities along the bar belonged to cluster C4. Together with the two main clusters (C1 and C2), five clusters were ultimately obtained for the classification of the pixels of the 5-index image.

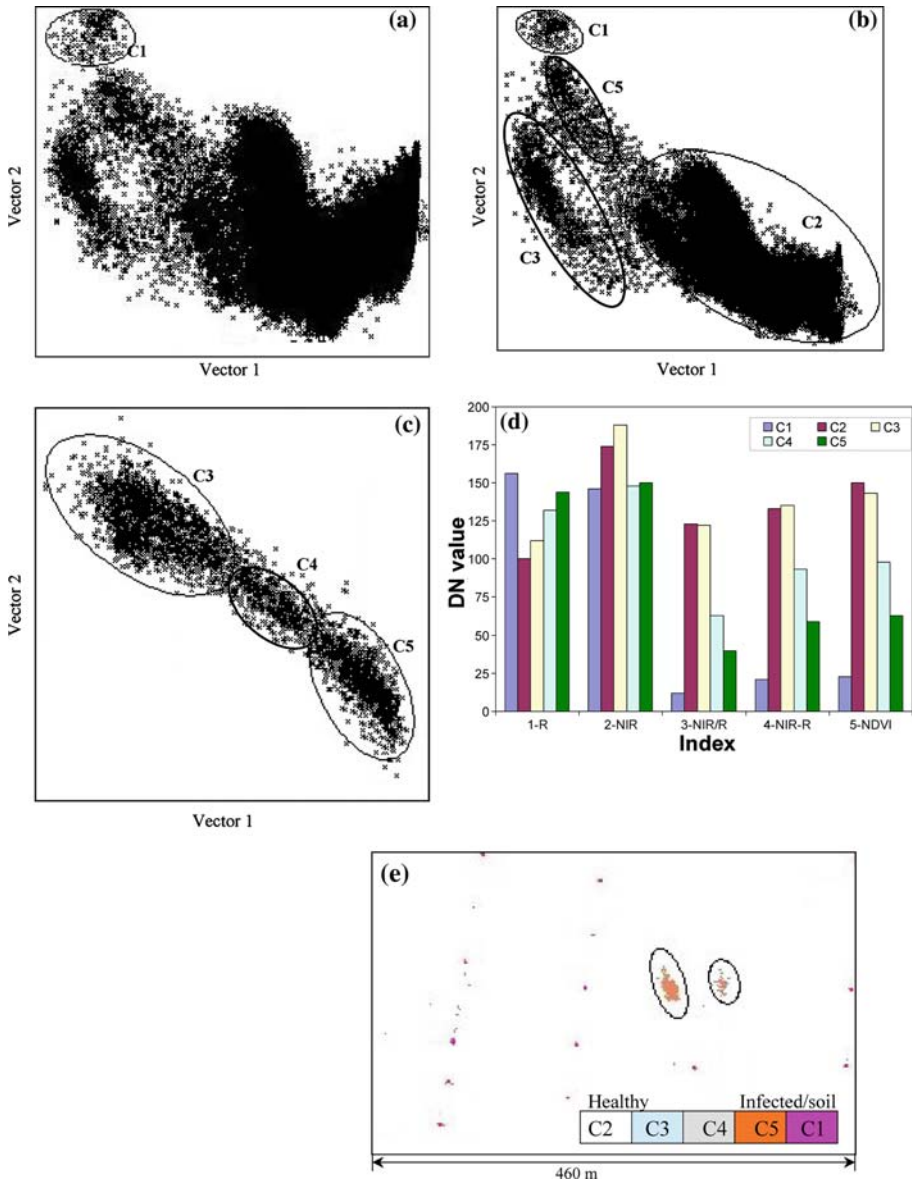

Figure 6. Validation of the method to the tomato field in Salina Valley of California, (a) identification of cluster 1, (b) clusters 1 and 2, and (c) sub-clusters 3, 4, and 5 in the 5-dimensional feature space, (d) mean DN values of the 5 clusters in the 5 indices (indices 1 through 5 refers to R, NIR, NIR/R, NIR-R, and NDVI, respectively, Table 1), (e) results of applying the 5 clusters to classify the images.

After the classification, in relation to the mean DN value of each cluster for the 5 indices (Figure 6d), the clusters were associated with the various disease severities. The cluster C1 should be soil background because the cluster had the lowest mean DN value in the indices 3 (NIR/R), 4 (NIR−R) and 5 (NDVI) and the highest mean

Table 2. Accuracy of classification of the selected 5 locations with infection above stage 3

Selected location	Total #of pixels	Correct classification	Misclassification	Accuracy (%)
1	8	8	0	100.00
2	62	54	8	87.10
3	54	45	9	83.33
4	98	87	11	88.78
5	84	72	12	85.71
Sum	306	266	40	86.93

DN value in index 1 (R) (Figure 6d). The cluster C2 should be the healthy plants and the cluster C3 the light infected plants such as LB1 plants or mixed with healthy plants. This was because clusters C2 and C3 had very close mean DN values in the indices 3, 4 and 5, but cluster C3 had a slightly higher mean DN value than C2 in index 1. Since most plants in the field were not infected, and cluster C2 contained most pixels of the image, it was logical to state that this C2 cluster represented the healthy plants. The clusters C4 and C5 had a mean DN value much lower than the clusters C2 and C3 (Figure 6d). The mean DN values of these two clusters were higher than the cluster C1 in the indices 3, 4 and 5 (Figure 6d). Thus, clusters C4 and C5 were the LB3 and LB4 plants, respectively (Figure 6d).

Exporting the above classification results from visualization analysis to the original image, we obtained the final result of image classification shown in Figure 6e to differentiate the healthy plants from diseased plants in the field image. As expected, most pixels were classified as healthy plants (cluster C2). The only two patches of infected plants identified in Figure 6e have been classified as clusters C4 and C5. Although the ground truth observation identified both patches as infected canopies at LB3, it was still possible to distinguish the slight differences between them in the image. The left patch was slightly more severely infected than the right one (Figure 6e). The spots along the straight lines in Figure 6e were the locations of sprinklers for irrigation, surrounded by bare soils. These spots were grouped as soil background. With the limited ground truth to the background and the infected locations, the classification of pixels with infection stages 4 and 5 as well as the background soil is very accurate. For the selected locations with infection at stage 3, the accuracy can be above 85%. Table 2 presents the accuracy of the classification for the 5 selected locations with medium to severe infections (above stage 3). We had a total of 303 pixels with infection above stage 3. Our analysis procedure correctly identified 266 of them, with an average accuracy of 86.93% (Table 2). Thus, we can conclude that the procedure is effective to identify the diseased plants from healthy ones when the infection reaches to stage 3. For light infection at stages 1 and 2, it may be difficult to obtain an accurate identification due to its similar spectral reflectance to that of the healthy plants.

The analysis results from the two images revealed that late blight infected tomatoes can be successfully differentiated from the healthy tomatoes using multispectral remote sensing image when the infection reached stage 3 or above. However, under

the field condition, we also recognized that it is difficult to separate the lightly infected tomato plants from the healthy plants.

Discussion

As early as 1950s, the possible connection between crop stress and remote sensing measurement has been examined (Colwell, 1956; Keegan *et al.,* 1956; Lathrop and Pennypacker, 1980; Toler *et al.,* 1981; Kurschner *et al.,* 1984; Blakeman, 1990). Our field spectral analysis revealed the possibility of spectrally discriminating the infected plants from healthy plants when the infection reached stage 3 or above. There were two major spectral reflectance characteristics of the infected plants: (1) spectral reflectance is relatively higher than the healthy tomatoes in the red band and is much lower in the NIR region; and (2) the spectral reflectance of infected plants increases in the red band and decreases in the NIR region with infection severity (Zhang *et al.,* 2002).

The red and NIR bands are much more sensitive to plant disease infections and are more valuable for detecting disease infections (McDonald *et al.,* 1998). The numeric analysis of the vegetation indices based on red and NIR bands indicated that three transformations could effectively enhance the separation ability of the healthy plants from the infected tomatoes. These applicable indices included (1) the simple ratio NIR/R, (2) the simple difference NIR−R, and (3) the normalized difference vegetation index (NDVI). Using each band of the new 5-index image as a vector, the distribution patterns of the pixels were identified to distinguish healthy tomatoes from diseased tomatoes and then to further classify the infection stages. This new 5-index image integrated the best spectral discrimination for infections on plants and increased the spectral differences between healthy and diseased tomatoes. Although any of the indices alone can discriminate healthy and infected plants to a certain degree, it is difficult to generate a threshold for the classification based on the pixel values and the distributions of each index. This new 5-index image is able to enhance the discrimination capability of crop disease infections because it represents the features relating to chlorophyll and biomass of the canopies. Pest damage and disease pressures significantly change the pigment contents and biomass in the canopy. These changes can be spectrally detected in the visible and infrared regions. The 5-index image has been shown to successfully discriminate late blight infected tomato canopies from healthy tomatoes.

The inference of the numeric analysis and validation of the method through image analyses on two diseased fields indicated that we were able to successfully classify the images into clusters representing healthy and diseased tomato plants using this 5-index image when the infections reached stage 3 or above. Although our hope was to detect the infection at an earlier stage, the results achieved were significant (Zhang *et al.,* 2002, 2003) for farmers and crop consultants to protect the disease damage to the production. The infection at stage 3 represents the critical point at which the late blight starts to create serious economic effects to tomato production. According to crop consultants and farmers in Yolo County in California, economic loss of tomato farming is usually obvious ($>50\%$) when the infection reaches stage 4 or above. However, the loss is much lower ($<20\%$) if the infection is at stage 3 or less. This

indicates that some treatments can still be done. Thus, working together with researchers and crop consultants, farmers will be able to use the classified disease result from multispectral remote sensing to prevent the spread of disease before significant economic loss occurs. The benefits of using multispectral imagery for precision disease management were clear, as also noted by Toler *et al*. (1981) and Blakeman (1990). Our study provided a means for precise disease management for tomatoes.

Conclusions

Late blight infections can be identified by remotely sensed images at resolution of 1 m when infection reached stage 3 or above. Earlier detection of the disease can be difficult due to its similar spectral response to that of healthy plants. The developed 5-index feature vector method can enhance the separation ability for diseased plants from healthy ones. This study provided a method of identifying late blight infection on tomato fields in California and demonstrated the capability of utilizing multi-spectral images in monitoring crop growth and precisely managing diseases in fields.

Acknowledgments

We want to thank the editors and the anonymous reviewers for the valuables comments to make this a better paper. We want to thank Carla Thomas, a plant pathologist, from Western Farm Services for providing field disease identifications, and Thomas Christensen from Zeneca Ag Products for valuable discussions and comments on earlier work of this paper. The authors also want to thank Zeneca Ag Products for partial financial support of this study.

References

Agrios G. N. 1997. Plant Pathology 4th edn (Academic Press. London, UK), p. 653.

Bausch, W. C. 1993. Soil background effects on reflectance-based crop coefficients for corn. Remote Sensing of Environment **46**, 213–222.

Blakeman, R. H. 1990. The identification of crop disease and stress by aerial photography. In: *Application of Remote Sensing in Agriculture*, edited by M. D. Steven and J. A. Clark (Butterworths, London, UK), p. 229–254.

Blazquez, C. H. and Edwards, G. J. 1983. Infrared color photography and spectral reflectance of tomato and potato diseases. Journal of Applied Photographic Engineering **9**, 33–37.

California Department of Pesticide Regulation (CDPR). 2002. Pesticide use in California indexed by chemicals and by commodities. (Sacramento, CA, USA).

Chapelle, E. W. and Kim, M. S. 1992. Ratio analysis of reflectance spectra (RARS): an algorithm for the remote estimation of concentration of chlorophyll a, chlorophyll b and carotenoids in soybean leaves. Remote Sensing of Environment **18**, 255–267.

Colwell, R. N. 1956. Determining the prevalence of certain cereal crop diseases by means of aerial photography. Hilgardia **26**, 223–286.

ENVI, 1999. Environment for Visualization Images User's Guide. (Research System Institute, Denver, CO, USA).

Evens, K, Webster, R., Barker, A., Halford, P., Stafford, J. and Griffin, S. 2003. Mapping infestations of potato cyst nematodes and the potential for spatially varying application of nematicides. Precision Agriculture **4**, 149–162.

Fitzgerald, G. J., Maas, S. J. and Detar, W. R. 2004. Spider mite detection and canopy component mapping in cotton using hyperspectral imagery and spectral mixture analysis. Precision Agriculture **5**, 275–289.

Gitelson, A. A. and Merzlyak, M. N. 1998. Remote sensing of chlorophyll concentration in higher plant leaves. Advanced Space Research **22**, 689–692.

Guyot, G. 1990. Optical properties of vegetation canopies. In: *Applications of Remote Sensing in Agriculture*, edited by M. D. Steven and J. Clark (Butterworths, London UK), p. 19–43.

Hatfield, J. L. and Pinter, P. J., Jr. 1993. Remote sensing for crop protection. Crop Protection **12**, 403–414.

Huete, A. C. 1988. A soil adjusted vegetation index (SAVI). Remote Sensing of the Environment **17**, 37–53.

Huete, A. C., Justice, C. and Liu, H. 1994. Development of vegetation and soil indices for MODIS-EOS. Remote Sensing of the Environment **49**, 224–234.

Keegan H. J., Schleter J. C., Hall W. A., Jr. and Haas G. M. 1956. Spectrophotometric and colorimetric study of diseased and rust resisting cereal crops Report (4591), National Bureau of Standards Washington, DC, USA.

Kimes, D. S., Markham, B. L., Tucker, C. J. and Mcmurtrey, J. E. 1981. Temporal relationships between spectral response and agronomic variables of a corn canopy. Remote Sensing of Environment **11**, 401–411.

Kurschner, E., Walter, H. and Koch, W. 1984. Measurements of spectral reflectance of leaves as a method for assessing the infestation with powdery mildew. Journal of Plant Disease Protection **91**, 71–80.

Lathrop, L. D. and Pennypacker, S. 1980. Spectral classification of tomato disease severity levels. Photogrammetry Engineering and Remote Sensing **46**, 1133–1138.

Lillesand, T. M. and Kiefer, R. W. 1994. Remote Sensing and Image Interpretation. 3rd edn (John Wiley & Sons, New York, USA).

Mcdonald, A. J., Gemmell, F. M. and Lewis, P. E. 1998. Investigation of the utility of spectral vegetation indices for determining information on coniferous forests. Remote Sensing of Environment **66**, 250–272.

Shibayama, M., Takahashi, W., Morinaga, S. and Akiyama, T. 1993. Canopy water deficit detection in paddy rice using a high resolution field spectrometer. Remote Sensing of Environment **45**, 117–126.

Steven M. D. and Clark J. A. 1990. Applications of Remote Sensing in Agriculture (Butterworths, London, UK), p. 427.

Thomas, J. R. and Oerther, G. F. 1972. Estimating nitrogen content of sweet pepper leaves by reflectance measurements. Agronomy Journal **64**, 11–13.

Toler, R. W., Smith, B. D. and Harlan, J. C. 1981. Use of aerial color infrared photography to evaluate crop disease. Plant Disease **65**, 24–31.

Tucker, C. J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment **8**, 127–150.

United States Department of Agriculture (USDA). (2002). The US processed tomato industry situation. National Agricultural Statistics Service. http://www.fas.usda.gov/htp/horticulture/Proc.Veg/The%20U.S.%20Processed%20Tomato%20Industry.pdf. November (2003) accessed.

Wiegand, C. L., Richardson, A. J., Escobar, D. E. and Gerbermann, A. H. 1991. Vegetation indices in crop assessments. Remote Sensing of Environment **35**, 105–119.

Yang K., Lin T. Y., Sun J. B. and Liu J. 1988. Digital Processing of Remotely Sensed Imagery (Surveying & Mapping Press Beijing, China).

Zhang, M., Ustin, S. L., Rejmankova, E. and Sanderson, E. W. 1997. Monitoring pacific coast marshes using remote sensing. Ecological Applications **7**, 1039–1053.

Zhang, M., Liu, X. and Oneill, M. 2002. Spectral Discrimination of *Phytophthora infestans* infection on tomatoes based on principal component and cluster analyses. International Journal of Remote Sensing **23**(6), 1095–1107.

Zhang, M., Qin, Z., Liu, X. and Ustin, S. L. 2003. Hyperspectral remote sensing applications in detecting late blight infection on tomatoes. International Journal of Applied Earth Observation and Geoinformation **4**, 295–310.