DEVELOPMENT OF GROUND-BASED SENSOR SYSTEM FOR AUTOMATED AGRICULTURAL VEHICLE TO DETECT DISEASES IN CITRUS PLANTATIONS

S. Sankaran and R. Ehsani

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

C. Dima

National Robotics Engineering Center
Carnegie Mellon University
Pittsburgh, Pennsylvania

ABSTRACT

In recent years, citrus-devastating diseases such as citrus canker and huanglongbing (HLB) have spread to various parts of Florida and are threatening the Florida citrus industry. The detection of HLB infected trees in the early stages is imperative to eliminate trees as sources of inoculum and to prevent further spread of the disease. Currently, grove scouting by a trained scouting crew is the most widely used technique for disease detection. There is a need for the development of new field-based, real-time, accurate sensing technologies for rapid plant disease detection. The present work evaluates the spectroradiometer as a prospective sensor for HLB detection in citrus. The reflectance from the HLB-infected and healthy leaves were collected in the range 350-2,500 nm. After preprocessing, the spectral data were classified using two classification algorithms: quadratic discriminant analysis (QDA) and k-nearest neighbor (kNN). Results indicated that the classification accuracies for classifying healthy and diseased classes were > 80%. The classification accuracies for predicting healthy class were found to be higher than diseased class prediction in most cases. In general, the classification accuracies of QDA were found to be higher than kNN with accuracies greater than or equal to 90%. Thus, the optical sensors provide a potential for disease detection in citrus orchards.

Keywords: Disease detection, Huanglongbing, citrus greening, visible-near infrared spectra.

INTRODUCTION

Huanglongbing (HLB) or citrus greening is a major citrus disease that has greatly affected the citrus production in Florida. Since HLB was discovered in

southern Florida in 2005 (Chung and Brlansky, 2009), it continues to spread to other citrus production areas in Florida. The symptoms of HLB are distinctively visible in the leaves. The leaves develop blotchy mottle with asymmetric yellowing or chlorosis. These symptoms are often confused with nutrient deficiencies. The citrus growers, related industries, and researchers are working together to prevent the spread of HLB.

The HLB can be detected only through appearance of symptoms in the citrus leaves. After the infection of a citrus tree with causative bacterium, *Candidatus* Liberibacter asiaticus, it takes about six months to a year or two (based on the age of the tree) for the symptoms to appear. By the time, the symptoms appear and disease is identified in the trees; the infected tree has already acted as a source of inoculum for further spread of the disease. Therefore, there is a need for the development of new field-based sensing technologies for rapid plant disease detection. The optical sensing technique offers a real-time, non-destructive, accurate, and rapid field-based detection of citrus diseases.

The visible and infrared is a fast-developing technology being used for the detection of stress, injury, and diseases in plants (Polischuk et al., 1997, Spinelli et al., 2006, Naidu et al., 2009). The visible and infrared regions of the electromagnetic spectra are known to provide the maximum information on the physiological stress levels in the plants (Muhammed, 2002; Muhammed, 2005; Xu et al., 2007) and thus, some of these wavebands specific to a disease can be used to detect plant diseases (West et al., 2003), even before the symptoms are visible. In general, visible spectroscopy in combination with infrared spectroscopy is used for disease detection in plants (Malthus and Madeira, 1993; Bravo et al., 2003; Huang et al., 2004; Larsolle and Muhammed, 2007).

An integrated USDA-funded project involving Carnegie Mellon University, University of Florida, Cornell University and John Deere is ongoing, to develop an autonomous tractors for sustainable specialty crop farming. The research teams have come together to develop an automated system for detecting plant stress, estimating yields, and reducing chemical usage through precision spraying for specialty crops. The goals of the automation process are to reduce the tractor-related labor costs, reduce the scouting labor costs, to improve equipment utilization, and reduce chemical usage. As a part of these efforts, a sensor platform was developed to monitor the diseases in citrus trees through multiple-sensing techniques. These systems can be used for a broad range of specialty crops. The present work involves the evaluation of an optical sensor system for field- and laboratory-based detection of citrus greening.

METHODOLOGY

Sensor system

A high resolution field-portable spectroradiometer, SVC HR-1024 (Spectra Vista Cooperation, NY) was used for spectral data collection. The spectral range of the spectroradiometer was 350 nm to 2,500 nm. A 4º field of view optic was used in order to identify the spectral wavelengths specific for HLB detection in citrus. During the field-based data collection, the sensor was placed in a stand in a moving sensor platform. Additional light source, 500-watt portable halogen lamp

was used during data collection. Our preliminary studies indicated that additional light source was required (even during day time) to prevent noise in the spectral data, especially in the near infrared region. Two lamps were angled at 45º to provide uniform light source. The white and dark background correction was performed as required.

Data collection

The spectral reflectance data from symptomatic citrus leaves and healthy leaves from Southern Garden citrus grove were collected using the spectroradiometer interfaced with computer. The HLB-infected samples were pre-marked with the help of Southern Garden's head scout. After the field spectral data collection, all the samples were analyzed in laboratory, followed by sending the leaf samples for polymerase chain reaction (PCR) analysis. The number of healthy and HLB-infected orange trees analyzed was 100 and 101, respectively. Five spectra were collected from each sample and averaged for analysis. During the data collection, the distance between the spectroradiometer and the leaf sample was maintained at about 0.5-0.6 m.

Data analysis

Data preprocessing

The spectral reflectance data were preprocessed for specific wavelength identification and classification purposes. Each sample spectral reflectance values were normalized using equation 1.

$$R_{norm(i)} = \frac{R_i}{\left[\sum R_i^2 + R_{i+1}^2 .... R_n^2\right]^{1/2}} \quad (1)$$

where, $R_{norm(i)}$ = normalized reflectance for a particular wavelength in a sample, R_i is measured reflectance of a particular wavelength in a sample, and i varies from 1 to n (989), referring to each wavelength within a sample. After normalization, the data were averaged every 25 nm. The 989 spectral data features were reduced to 86 data features for each sample.

Classification of visible-infrared reflectance spectra

The data were analyzed in two ways. In first, the entire averaged spectral data were analyzed; while in second, specific wavelengths that possess the discriminatory power were used for classification. During the first method of analysis, the Savitzky-Golay filter was applied to determine the first and the second derivatives. A window size of 7 and a polynomial order of two were used for filtering. Following this, preprocessed data were merged with first and second derivatives, followed by principal component analysis (PCA). The principal components (PCs) were selected such that the variance of the data > 99.9%. The principal components were used as the input features for classification. Two classifiers, quadratic discriminant analysis (QDA) and k-nearest neighbor (kNN) were used for classification. The k for kNN algorithm was varied from 1 to 15,

and optimized based on the maximum classification accuracy. The healthy samples were referred as class 1, while diseased (HLB) samples were referred as class 2.

In the second method of analysis, the critical wavelengths contributing to the variations between the classes (healthy and HLB) were identified. These wavelengths were selected using stepwise discriminant analysis. The field and the laboratory data were classified based on these wavelengths as classifier input features. In addition, common wavelengths were derived from laboratory and field data and analyzed in a similar manner. Two classifiers, QDA and kNN were used for classification. All the data analysis was performed using MATLAB®7.6 with an exception of stepwise discriminant analysis, which was performed using SAS®9.2.

RESULTS AND DISCUSSION

Classification of visible-infrared reflectance spectra

The spectral reflectance data (Fig. 1.) were acquired from the SVC spectroradiometer in visible-near infrared region (350-2,500 nm) of the electromagnetic spectrum. The resolution of the reflectance data depended on the wavelength of the spectra. The resolution of the spectral reflectance was between 0.9-1.5 nm for wavelengths from 350 to 980 nm; while, resolution was between 2.1-3.9 nm for wavelengths from 980 to 2,500 nm. After averaging the spectral reflectance every 25 nm, the spectral reflectance features were reduced from 989 to 86 spectral features.

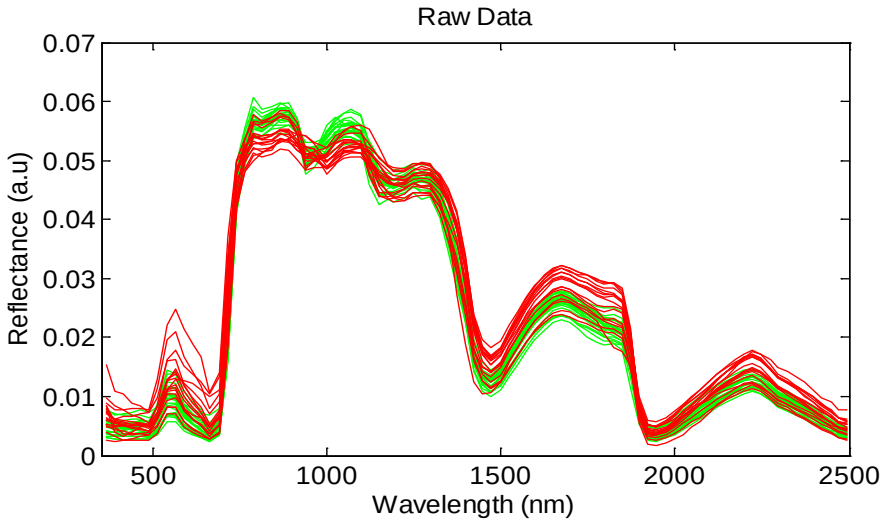

Figure 1. Spectral reflectance of the representative field data from 350-2,500 nm.

The Savitzky-Golay filter was applied to determine first and second derivatives of the spectra. As a window size of 7 was used during filtering, a total 80 spectral features from first and second derivations (each) were acquired. A total of 246 spectral features were obtained for each sample. The number of

features was reduced through PCA. Twenty four and sixteen PCs (99.9%) were used for the further analysis of field and laboratory data, respectively. It could be observed that fewer PCs were required in analyzing laboratory data in comparison with field data, due to low variability. In the field condition, there were many challenges during data collection that lead to some variations. Some of them are presence of wind, non-uniform canopy from tree to tree, changing light conditions, and presence of symptomatic leaves in-between the canopy of the tree. Fig.2. shows the PC plot for the field and laboratory data.

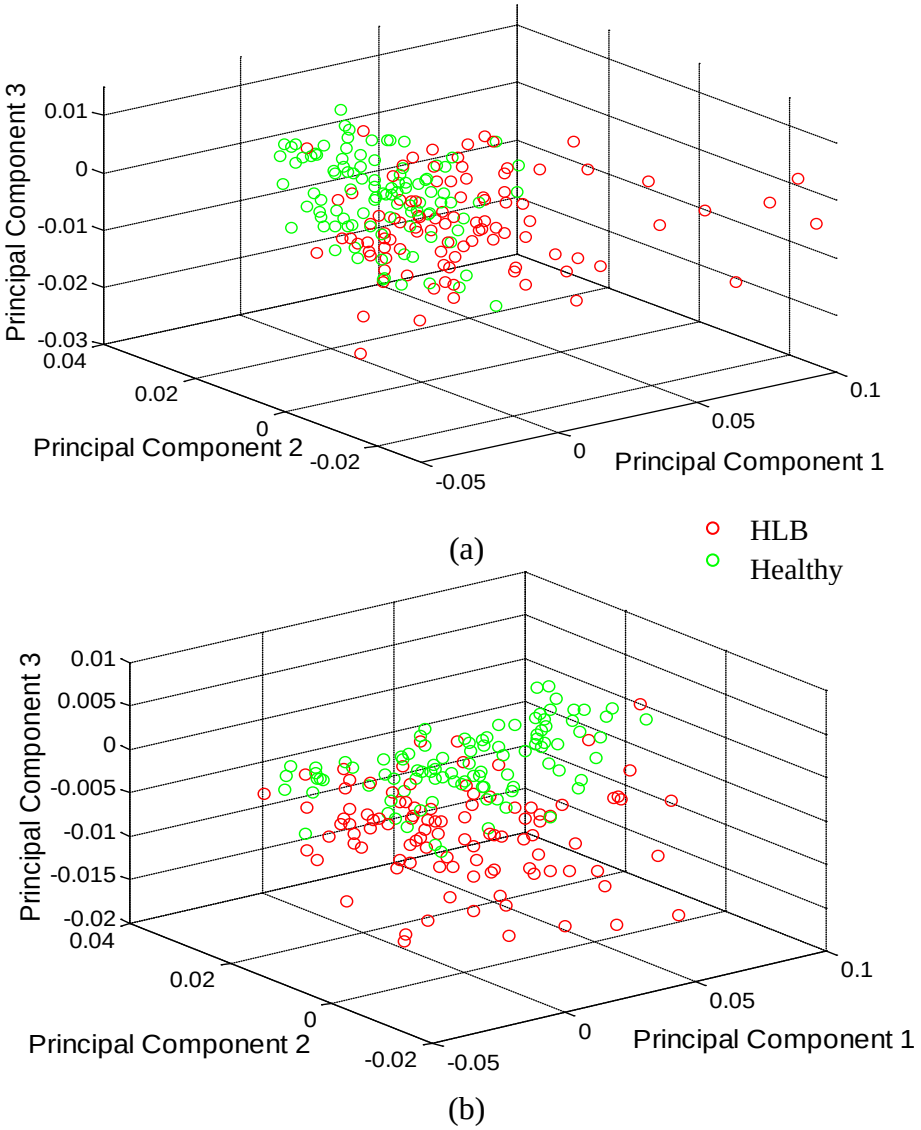

Figure 2. Principal component analysis of the spectral reflectance data features.

The PCs were used as input features for the classification algorithm. The data were randomized and separated into train and test class in the ratio 4:1. For the same set of data, the classification accuracy of QDA and kNN were determined. The procedure was performed three times such that the data was randomized three

times before dividing the data into train and test data, and classification. The results of QDA- and kNN-based classification are summarized in Table 1 for the field and laboratory dataset.

Table 1. Classification accuracies (%) of healthy and diseased class using entire spectral data.

Accuracy (%)	QDA					kNN			
Run	I	II	III	Average	k	I	II	III	Average
Field Data									
Total	98	93	90	93		83	85	90	86
Class1	94	100	85	93	7	94	100	90	95
Class2	100	88	95	94		73	75	90	79
Laboratory Data									
Total	90	92	85	89		90	85	92	89
Class1	96	100	89	95	8	91	94	89	91
Class2	81	86	81	83		88	77	95	87

The classification accuracies of both the classifiers for classifying healthy and diseased sample data were found to be > 80%. In field data analysis, the QDA overall average classification accuracies (93%) were distinctly higher than the kNN overall average classification accuracies (86%). In laboratory-based data analysis, though the classification accuracies of QDA were slightly higher than that of kNN, the accuracies were comparable. Comparing the false positive, the false negatives were found be higher with a lower classification accuracies in most cases, except QDA-based classification of field data. Thus, QDA is a better classifier than kNN for spectral data classification.

Wavelength selection

The spectroradiometer captures reflectance from the tree canopy for a broad wavelength range. For these reasons, the spectroradiometer instruments are expensive. The cost of sensors can further be decreased by reducing the number of wavelength bands used for data collection. This will also help in the fabrication of a more rugged and potable sensor. In addition, selection of wavebands specific to a particular disease will increase the accuracy of the sensor systems. For these reasons, the wavelengths possessing the ability to classify the healthy and diseased samples were identified. Stepwise discriminant analysis selects the quantitative variables that contribute to the discrimination among the classes. The stepwise discriminant analysis was performed in field and laboratory data. In addition, a set of wavelength variables were identified that were found to be occurring in wavelengths selected by stepwise discriminant analysis from both the datasets (field and laboratory). Table 2 summarizes the wavelengths showing discriminatory power selected from field and laboratory data analysis.

Table 2. Wavelengths selected by stepwise discriminant analysis.

Data	No. of wavelt.	Significant Wavelengths (nm)
Laboratory data	12	612, 638, 662, 713, 813, 1445, 1622, 1923, 1997, 2047, 2098, 2471
Field data	24	537, 612, 688, 713, 763, 813, 963, 998, 1023, 1120, 1148, 1272, 1296, 1498, 1524, 1597, 1647, 1822, 1873, 1898, 2073, 2121, 2172, 2273
General	12	537, 612, 688, 713, 813, 913, 1120, 1346, 1445, 1597, 1622, 1898

Based on the selection of wavelengths that contributes to classification of the classes from the field and laboratory data, a set of common wavelengths were derived. Except some wavelengths in green and red region of the visible spectra, most of the wavelengths were found to be in mid-infrared region.

Classification of spectral reflectance data based on selected wavelengths

The specific wavelengths derived from stepwise discriminant analysis of the indoor reflectance data, and commonly derived wavelengths were utilized for further analysis. The reflectance values from these selective wavelengths were used as input features in the classification model. Table 3 and 4 summarizes the classification results when the specific wavelength reflectances from the laboratory data were used and commonly derived wavelength reflectance values were used as model input features, respectively. Though, there were variations in the optimum 'k' values, the classification accuracies were high. The kNN-based algorithm yielded low classification accuracies (especially field data) in comparison to QDA-based model. Moreover, the kNN-based classification accuracy to predict diseased class was lower then healthy class, resulting in higher false negative values. The QDA classification accuracies were 10% higher than kNN-based classification. The average overall classification accuracy of QDA was > 90%. The overall classification accuracies of laboratory data were higher than field data.

Table 3. Classification accuracies (%) after stepwise discriminant analysis of laboratory data.

Accuracy (%)	QDA				kNN				
Run	I	II	III	Average	k	I	II	III	Average
Field Data									
Total	90	93	88	90	8	88	83	85	85
Class1	89	95	86	90		89	95	90	92
Class2	90	89	89	90		86	68	79	78
Laboratory Data									
Total	95	87	100	94	8	95	77	92	88
Class1	90	88	100	93		100	76	95	91
Class2	100	86	100	95		89	77	88	85

Table 4. Classification accuracies (%) based on selective wavelengths derived from field and laboratory data.

Accuracy (%)	QDA					kNN			
Run	I	II	III	Average	k	I	II	III	Average
Field Data									
Total	93	90	88	90		80	83	78	80
Class1	94	90	89	91	9	94	95	79	89
Class2	91	90	86	89		68	70	76	71
Laboratory Data									
Total	92	92	87	91		87	79	82	83
Class1	94	91	88	91	9	88	78	88	85
Class2	91	94	87	91		86	81	78	82

CONCLUSION

In the present work, a spectroradiometer was used as an optical sensor to detect citrus greening in leaves. The spectral reflectance in the range 350-2,500 nm was collected from the HLB-infected and healthy leaves under field and laboratory conditions. The reflectance data was preprocessed by averaging every 25 nm in order to reduce the number of features during analysis. Two classifiers, quadratic discriminant analysis and k-nearest neighbor were used for classification. The classification accuracies for classifying healthy and diseased class were found to about 80% or higher during kNN-based analysis. The QDA-based algorithm yielded high classification accuracies of 90% or higher, with a very low false negatives, as desired for real-world applications. In general, the classification accuracies for predicting healthy class were found to be higher than disease class prediction. Efforts were also made to identify the wavelengths that provide the discriminatory power within the spectra among the classes. It was found that most of the wavelengths contributing to the discrimination were present in the near-infrared region. This study demonstrates the potential of optical sensors for disease detection.

REFERENCES

Bravo, C., Moshou, D., West, J., McCartney, A., Ramon, H., 2003. Early disease detection in wheat fields using spectral reflectance. Biosystems Engineering 84 (2), 137-145.

Chung, K-.R., Brlansky, R.H. 2009. Citrus diseases exotic to Florida: Huanglongbing (citrus greening). Florida cooperative extension service/IFAS, University of Florida, Report No. PP-210.

Huang, M. Y., Huang, W. H., Liu, L. Y., Huang, Y. D., Wang, J. H., Zhao, C. H., Wan, A.M., 2004. Spectral reflectance feature of winter wheat single leaf infested with stripe rust and severity level inversion. Transactions of the CSAE 20 (1), 176-180.

Larsolle, A., Hamid Muhammed, H., 2007. Measuring crop status using multivariate analysis of hyperspectral field reflectance with application to disease severity and plant density. Precision Agriculture 8 (1-2), 37-47.

Malthus, T. J., Madeira, A. C., 1993. High resolution spectroradiometry: spectral reflectance of field bean leaves infected by *Botrytis fabae*. Remote Sensing of Environment 45, 107-116.

Muhammed, H. H., 2002. Using hyperspectral reflectance data for discrimination between healthy and diseased plants, and determination of damage-level in diseased plants. IEEE: Proceedings of the 31st Applied Imagery Pattern Recognition Workshop 49-54.

Muhammed, H. H., 2005. Hyperspectral crop reflectance data for characterizing and estimating fungal disease severity in wheat. Biosystems Engineering 91 (1), 9-20.

Naidu, R. A., Perry, E. M., Pierce, F. J., Mekuria, T., 2009. The potential of spectral reflectance technique for the detection of Grapevine leafroll-associated virus-3 in two red-berried wine grape cultivars. Computers and Electronics in Agriculture 66, 38-45.

Polischuk, V. P., Shadchina, T. M., Kompanetz, T. I., Budzanivskaya, I. G., Sozinov, A., A., 1997. Changes in reflectance spectrum characteristic of *Nicotiana debneyi* plant under the influence of viral infection. Archives of Phytopathology and Plant Protection 31(1), 115-119.

Spinelli, F., Noferini, M., Costa, G., 2006. Near infrared spectroscopy (NIRs): Perspective of fire blight detection in asymptomatic plant material. Proceeding of 10th International Workshop on Fire Blight, Acta Hort. 704, 87-90.

West, J. S., Bravo, C., Oberti, R., Lemaire, D., Moshou, D., McCartney, H. A., 2003. The potential of optical canopy measurement for targeted control of field crop disease. Annual Review of Phytopathology 41, 593-614.

Xu, H. R., Ying, Y. B., Fu, X. P., Zhu, S. P., 2007. Near-infrared spectroscopy in detecting leaf miner damage on tomato leaf. Biosystems Engineering 96 (4), 447-454.