CHARACTERISTIC WAVELENGTH SELECTION OF HYPERSPECTRAL IMAGES FOR PLANT DISEASE ANALYSES

Yu-Hui Flora Yeh¹, Wei-Chang Chung¹, Rong-Siou Lee¹, Jui-Yu Liao², Chia-Lin Chung³, Yan-Fu Kuo¹, Ta-Te Lin¹*

¹ Department of Bio-Industrial Mechatronics Engineering, National Taiwan University, Taiwan, R.O.C.
² Department of Plant Pathology and Microbiology, National Taiwan University, Taiwan, R.O.C.
*Corresponding Author-- Voice: +886-2-3366-5331, Email: m456@ntu.edu.tw

Abstract: Non-destructive methods have gained popularity in agricultural applications as they do not affect plant growth while examining plant health conditions. Hyperspectral imaging is one of the non-destructive techniques that have been widely applied in practices such as plant disease detection and assessment. It provides comprehensive spectral and spatial image information of observed objects. However for quick assessments, this huge amount of data needs to be filtered before processing. Two plant diseases were investigated in this research: strawberry Anthracnose and bok choy black spot disease. An early detection of these diseases can be beneficial to ensure the production and quality. This research aims to select the characteristic wavelengths across the spectral coverage of 400 to 1000 nm for each selected plant disease. With only a set of characteristic wavelengths, plant disease can be distinguished and assessed to avoid full spectral information analysis. The characteristic wavelengths were selected based on the correlation criterion. Plant disease assessment models were built based on these characteristic wavelengths with a simple decision tree structure. The performance of assessment models was evaluated and compared with a commonly used hyperspectral imaging method, stepwise discriminant analysis. Results have shown the feasibility and benefits of this approach for plant disease assessment applications. This characteristic wavelength selection method can further be applied to build multispectral system for more cost-effective and speedy examination. Furthermore, not only plant diseases can be detected but also production can be improved with proper quality controls.

Key Words: Hyperspectral Imaging, Plant Disease, Characteristic Wavelength

INTRODUCTION

The agricultural production and economic loses across the world can be affected critically by both physiological and infectious plant diseases. Hence, plant disease predetermination and
prevention has raised great interests in research, especially in real-time and non-destructive techniques. Hyperspectral image analysis has been applied to plant disease detections and assessments (Coops et al., 2003; Mahlein et al., 2012; Qin et al., 2012). Hyperspectral images provide both spatial image and spectral information of the observed object. Spectral information is resourceful but may be redundant for adjacent wavelengths (Bajscy and Groves, 2004; Martinez-Uso et al., 2007); hence, characteristic wavelengths with only essential information should be identified. Then a multispectral system can be built based on these characteristic wavelengths selected to cost down and speed up plant disease assessment.

This research aims to propose a method to select characteristic wavelengths from the spectral range of a hyperspectral image, especially for plant disease applications. The results verify the feasibility of this method and are comparable to stepwise discriminant analysis.

MATERIALS AND METHODS

PLANT DISEASE SAMPLES

For the strawberry foliage Anthracnose assessment, the most widely cultivated strawberry in Taiwan, Taoyuan No. 3, was selected to be inoculated with Colletotrichum gloeosporioides (Gc001). For the Bok Choy black spot disease samples, the Tainan No. 1 was the cultivar selected and spray inoculated with Alternaria brassicicola. Detached leaves, both strawberry leaves and Bok Choy, were placed in the sterile 15-cm diameter petri dishes. Every detached leaf was immobilized using a specialized black mesh. About 0.5 ml of spore suspension (10^6 conidia mL^-1) was evenly sprayed on each detached leaf (fixed on the plate) with an airbrush at 10-15 psi. The plates were then sealed with parafilm to maintain the RH at > 90%. After the one full day incubation period, strawberry foliage samples were scanned every 12 hours, while Bok Choy samples were scanned every 6 hours across the whole observation period. Extra details about the samples are provided in Table 1.

Table 1. Plant disease sample preparation

<table>
<thead>
<tr>
<th>Sample details</th>
<th>Strawberry Anthracnose</th>
<th>Bok Choy Black Spot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inoculation date</td>
<td>2012/12/11</td>
<td>2012/12/04</td>
</tr>
<tr>
<td>Observation period</td>
<td>2012/12/12–2012/12/18</td>
<td>2012/12/05–2012/12/07</td>
</tr>
<tr>
<td>Number of control samples</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Number of inoculated samples</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

HYPERSPECTRAL IMAGING

The hyperspectral images of inoculated samples were taken with Headwall HyperspecTM VNIR A-series hyperspectral imaging system set in a dark cabinet, two linear halogen light source and a moving platform. The spectral range of the hyperspectral camera is from 400 to 1000 nm with the spectral resolution set to be 4.6 nm. The movable platform is powered by a stepping motor which move the sample in steps for it to be line-scanned by the camera.

In this research, three different plant health statuses were assessed: healthy, incubated, and
symptomatic. For each plant disease, fifty 3x3 pixel blocks for each health status were selected from hyperspectral images scanned. All healthy areas were sampled randomly from the control samples across different days. As the infection severities across samples were not even, the symptomatic areas were chosen by the defined darkness level of the symptomatic spot. In other words, the number of symptomatic areas selected in each sample was not the same. This was the same case for the incubated areas. The incubated area was defined as the area which has no visible plant disease symptom but develops into visible symptoms 24 hours later.

CORRELATION MEASURE

For any particular plant disease, developing a method to select characteristic wavelengths from the whole spectral range is the main focus of this research. As the characteristic wavelengths are defined as those at which or the combination of those the health statuses can be distinguished, a measure expressing the difference between health statuses at each wavelength is what we are looking for. Hence, we designed to use the correlation to evaluate how different each health status pair is at each wavelength. Through greedy searching over every wavelength’s correlation coefficients across the whole spectral range, the characteristic wavelengths at which the health status pair is uncorrelated can be identified. The correlation coefficient value for characteristic wavelength selection is set to be as close to 0 as possible. Furthermore, this value can set as low as there is at least one characteristic wavelength representing every class.

RESULTS & DISCUSSION

In order to learn the efficiency of correlation measure method, there are two criteria to evaluate its performance: by the number of characteristic wavelength selected and by the accuracy of assessment model built on characteristic wavelengths. The assessment models are chosen to be built as decision trees as each characteristic wavelength is coupled with a health status pair. Furthermore, the correlation measure is compared with the stepwise discriminant analysis (SDA) method based on these criteria as SDA is a commonly applied hyperspectral imaging method for characteristic wavelength selection.

COLOR BOARD

Firstly, a standard color board problem was studied to validate this new method. Six colors were selected from the standard color board as shown in Fig. 1 with forty-five 3x3 pixel areas per color. The six colors were separated into two sets of three-class problem: red/ green/ light blue and peach/ apple green/ light grey. These colors’ average spectral curves are also plotted in Fig. 1 for reference.

The characteristic wavelength sets selected by stepwise discriminant analysis and correlation measure are both listed in Table 2. For the red/ green/ light blue color problem, the correlation measures have four characteristic wavelengths comparing with nine from SDA. Also, all characteristic wavelengths match (within 10 nm difference) with those selected by SDA. For the other peach/ apple green/ light grey color problem, similar pattern can be
observed. The correlation measure selected only three characteristic wavelengths and two out of them match with those from SDA.

Applying these characteristic wavelengths from correlation measure to build a decision tree assessment model for each problem, the leave-one-out cross validation (LOOCV) accuracy for both color problems is 100%. The red/ green/ light blue problem only used 502 and 630 nm to build the decision tree model; while the peach/ apple green/ light grey assessment model required only one wavelength (595 nm). The SDA achieved same accuracy (100%) with more wavelengths in both color problems. An exciting point found is the characteristic wavelengths used to build the decision tree model are those match with wavelengths found by SDA.

Table 2. The color board characteristic wavelengths.

<table>
<thead>
<tr>
<th>Characteristic wavelengths (nm)</th>
<th>Red/ Green/ L Blue</th>
<th>Peach / Apple Green/ L Grey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stepwise Discriminant Analysis (SDA)</td>
<td>476, 493, <strong>498</strong>, 507, 604, <strong>626</strong>, 648, 666, <strong>777</strong></td>
<td><strong>555</strong>, 604, 613, 622, 626, 644, 648, 661, 750, 852, 870, 874, 878, 927, <strong>945</strong></td>
</tr>
<tr>
<td>Correlation Measure</td>
<td><strong>502</strong> (RG), <strong>630</strong> (GB), <strong>635</strong> (RG), <strong>781</strong> (GB)</td>
<td><strong>595</strong> (PA), 825 (PL), <strong>954</strong> (AL)</td>
</tr>
</tbody>
</table>

**PLANT DISEASES**

As color board problems in the previous section had shown the correlation measure to be a valid measure for selecting characteristic wavelengths, same methodology was applied to plant diseases mentioned: Strawberry Anthracnose and Bok Choy black spot disease. In the following, the health statuses are abbreviated to its respective first letter.

**Strawberry Anthracnose**
For plant disease problems, as the reflectance was acquired from the same material (leaf), the spectral curve pattern is very similar but the reflectance value varies for different health statuses. The characteristic wavelength selection results for strawberry Anthracnose are presented in Table 3. It is clear that characteristic wavelengths selected by correlation measure are less than SDA. With so many characteristic wavelengths, SDA can reach 80% for its LOOCV accuracy; however, with only two characteristic wavelengths, the LOOCV accuracy drops to 69.3%. This result is as accurate as the decision tree model built on characteristic wavelengths selected by correlation measure as shown in Table 3 and Fig. 2. Although the assessment models are different (linear discriminant model and decision tree), the characteristic wavelengths match exactly with each other. This finding coincides with the results in color board problems.

Table 3. The strawberry Anthracnose characteristic wavelengths.

<table>
<thead>
<tr>
<th>Characteristic wavelengths (nm)</th>
<th>Assessment model wavelengths (nm)</th>
<th>LOOCV accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stepwise Discriminant Analysis (SDA)</td>
<td>515, 613, 626, 635, 653, 719, 741, 750, 914, 989</td>
<td>635 and 750</td>
</tr>
<tr>
<td>Correlation Measure</td>
<td>591 (HS), 635 (HI), 684 (HI), 746 (HS), 861 (HS)</td>
<td>635 and 746</td>
</tr>
</tbody>
</table>

Figure 2. The decision tree assessment model for strawberry Anthracnose.

**Bok Choy Black Spot**

Unlike as in the strawberry Anthracnose problem, the characteristic wavelengths selected by the correlation measure do not match well with those from SDA in the Bok Choy black spot disease problem. Only one out of three wavelengths selected matches those from SDA selection. The decision tree model built at the end was based on this wavelength matched and showed an accuracy of only 55.3%. However, the SDA method could not build an assessment model with less than four wavelengths for this black spot disease problem with an accuracy as high as 95.3%. This result seems to be disappointing with such big difference in accuracy. Yet taken into consideration; two assessment models were built on different number of characteristic wavelengths and different model type. Not matching characteristic wavelengths imply that the problem is much more complicated and requires more than a few characteristic wavelengths.
Table 4. The Bok Choy black spot disease characteristic wavelengths.

<table>
<thead>
<tr>
<th>Characteristic wavelengths (nm)</th>
<th>Assessment model wavelengths (nm)</th>
<th>LOOCV accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stepwise Discriminant Analysis (SDA)</td>
<td>409, 555, 635, 653, 715, 737, 746, 852, 905, 918, 936</td>
<td>639, 710, 759, 852</td>
</tr>
<tr>
<td>Correlation Measure</td>
<td>414 (HI / IS), 427 (HI), 449 (HI)</td>
<td>414</td>
</tr>
</tbody>
</table>

CONCLUSIONS

A new and simple characteristic wavelength selection method, correlation measure, has been introduced in this paper. This method has been applied and assessed in four different problems. A standard method for characteristic wavelength selection in hyperspectral imaging, stepwise discriminant analysis, was chosen to be compared with this correlation measure for performance evaluation. With similar accuracy performance and matching characteristic wavelengths, the empirical results from the color board and strawberry Anthracnose problems have validated the efficiency of this correlation measure method. In particular, the correlation measure eliminates redundant wavelengths faster right from the beginning. With a simple decision tree model, the characteristic wavelengths can further trimmed to only one or two wavelengths. Interestingly, these wavelengths match with those selected by the stepwise discriminant analysis method. However, as shown in the Bok Choy black spot disease problem, the characteristic wavelengths selected by these two methods do not match, then the assessment model may not work as well as expected. These results conclude to an interesting direction for future works: both correlation measure and stepwise discriminant analysis should be applied to select characteristic wavelengths. The correlation measure can further define and trim down the characteristic wavelengths if both assessment models have matched selected wavelengths.

REFERENCES