Wavelength Selection for Detection of Slight Bruises on Pears Based on Hyperspectral Imaging

Hao Jiang 1, Chu Zhang 1, Yong He 1,*, Xinxin Chen 1, Fei Liu 1 and Yande Liu 2,*

1 College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China; 11413017@zju.edu.cn (H.J.); chuzh@zju.edu.cn (C.Z.); qhchenxinxin@163.com (X.C.); fliu@zju.edu.cn (F.L.)
2 Mechanical and Electrical Engineering Institution, East China Jiaotong University, Nanchang 330013, China
* Correspondence: yhe@zju.edu.cn (Y.H.); liuyande@ecjtu.jx.cn (Y.L.);
Tel.: +86-0571-88982143 (Y.H.); +86-0791-87046175 (Y.L.)

Academic Editor: Kuanglin Kevin Chao
Received: 5 October 2016; Accepted: 15 December 2016; Published: 19 December 2016

Abstract: Hyperspectral imaging technology was employed to detect slight bruises on Korla pears. The spectral data of 60 bruised samples and 60 normal samples were collected by a hyperspectral imaging system. To select the characteristic wavelengths for detection, several chemometrics methods were used on the raw spectra. Firstly, principal component analysis (PCA) was conducted on the spectra ranging from 420 to 1000 nm of all samples. Considering that the reliability of the first two PCs was more than 90%, five characteristic wavelengths (472, 544, 655, 688 and 967 nm) were selected by the loading plot of PC1 and PC2. Then, each of the wavelength variables was considered as an independent classifier for bruised/normal classification, and all classifiers were evaluated by the receiver operating characteristic (ROC) analysis. Two wavelengths (472 and 967 nm) with the highest values under the curve (0.992 and 0.980) were finally selected for modeling. The classifying model was built by partial least squares discriminant analysis (PLS-DA) and the bruised/normal classification accuracy of the modeling set (45 damaged samples and 45 normal samples) and prediction set (15 damaged samples and 15 normal samples) was 98.9% and 100%, respectively, which is similar to that of the PLS-DA model based on the whole spectral range. The result shows that it is feasible to select characteristic wavelengths for the detection of slight bruises on pears by the methods combining the PCA and ROC analysis. This study can lay a foundation for the development of an online detection system for slight bruise detection on pears.

Keywords: pear; damage detection; hyperspectral imaging; characteristic wavelength

1. Introduction

The Korla pear is characterized by its faint scent, thin skin, delicious flesh and sweet juice. In the international market, it is known as “the treasure of pears”. Bruising, usually caused by external forces during the process of picking and transporting [1], is one of the most common factors that affect the quality of the pears. However, some light bruises on the pears are difficult to detect by the naked eye or traditional computer vision technology [2] because there is no significant difference in color between lightly bruised and normal areas. So, it is very important to develop a method to detect light bruises on the surface of pears quickly and effectively.

Spectral technology is one of the most efficient tools for damage detection. Cao et al. [3] detected the damage degree of pears by visible and near-infrared (vis/NIR) spectroscopy and got a great result. Cho et al. [4] used a hyperspectral infra-red (1000–1700 nm) imaging technique to detect bruise damage underneath the pear skin and the results demonstrated good potential. Compared with the visible and near-infrared spectral analysis, hyperspectral image technology, which combines traditional 2D image technology with spectral technology, has a better ability to detect and show the internal or external...
attributes of objects [5]. With the advantage of multi-bands and a high resolution, hyperspectral image technology can meet the needs of slight bruise detection of Korla pears.

Hyperspectral imaging technology has been widely used in the damage detection of agricultural products, such as the external damage of tomatoes, freeze damage of mushroom, pest damage of soybeans, and bruising of pears [6–10]. However, the analysis of spectral data has always been difficult because of the complexity of the data. More and more scholars began to pay attention to the selection of feature bands to improve the efficiency of detection. Li et al. [11] built a hyperspectral imaging system to acquire reflectance images of orange samples in the spectral region between 400 and 1000 nm. Principal component analysis (PCA) was then used to select six characteristic wavelengths (630, 691, 769, 786, 810 and 875 nm). By developing an algorithm based on images at these six wavelengths, the accuracy rate of damage identification reached 91.5%. Wang et al. [12] detected the external damage of jujubes by acquiring hyperspectral reflectance images at the spectral region between 400 and 720 nm. Then, stepwise discriminant analysis (SDA) was used for modeling and the result showed that the accuracy rate of the damage identification was 97.0%. Rivera et al. [13] identified the mechanical damage of mangos by using a 650–1080 nm hyperspectral imaging system. They chose 700–780 nm, 890–900 nm, 1070–1080 nm as characteristic bands using a few methods and the results showed that the recognition rate of models based on those wavelengths was 91.4%. These studies reduced the wavelength range of the model to a certain extent and had ideal results in damage detection. So the extraction of characteristic wavelengths had no significant influence on the detection accuracy but accelerated the detection speed of the system.

Receiver operating characteristic (ROC) analysis is one of the methods used for choosing characteristic wavelengths. Luo et al. [14] analyzed the spectra of different kinds of apples. They used ROC analysis to extract wavelength combinations \((R(\lambda_1) - R(\lambda_2))\) which can identify damaged areas from normal areas for four kinds of apples separately. Comparing the results with that of the partial least squares discriminant analysis (PLS-DA) model based on the full-wave band, they found that the recognition rates were similar. Lorente et al. [15] combined ROC analysis with an artificial neural network to identify rot areas of citrus and found the recognition accuracy reached 89%.

The specific objectives of this paper were as follows: (1) to select the characteristic wavelengths for bruise detection of Korla pears by the method of combining ROC analysis with PCA; and (2) to compare the performance of the PLS-DA model based on selected wavelengths with that of the PLS-DA model based on all of the wavelength variables.

2. Materials and Methods

2.1. Pear Samples

Korla pears were purchased from a local market in Hangzhou in October 2015. One hundred twenty pears free from damage on surface and similar in size, were picked out as trial samples. Sixty samples were dropped from the height of 25 cm to the horizontal ground by artificial simulation and slight bruises were caused near the equator line of pears. These areas were difficult to be recognized by naked eyes but softer than normal areas. As shown in Figure 1a, the bruised area (in black frame) was similar to normal area around it. However, Figure 1b shows that when the peel of bruised area was cut off, there were obvious differences between normal and bruised area in color.

After bruising, all pears were left in the laboratory at room temperature (18–20 °C and 60%–70% Relative Humidity (RH)) for 24 h to allow the bruises to develop. Sixty hyperspectral images of bruised peers were collected as bruised samples and 60 hyperspectral images of normal peers were collected as normal samples. Forty-five bruised samples and 45 normal samples were chosen randomly as the model set and the others as the prediction set.
Figure 1. The bruised area of a Korla pear after the pear was left in the laboratory at room temperature for 24 h (to allow the bruise to develop). (a) The bruised area (in black frame) was similar to normal area around it; (b) when the peel of bruised area was cut off, there were differences between normal and bruised area in color.

2.2. Hyperspectral Image Acquisition System

As shown in Figure 2, the hyperspectral image acquisition system used in this experiment consists of an imaging spectrometer (Imspector V10E, Specim, Oulu, Finland), an imaging lens (OLE22, Imspector V10E), a light source (Fiber Lite Illuminator) equipped with an optical fiber linear lamp, a precision mobile platform, a step-motor and a computer. The imaging spectral region of the hyperspectral imaging system is 380–1030 nm, and the spectral resolution is 2.8 nm. The whole system is placed in a dark box in order to reduce the influence of ambient light.

Figure 2. Schematic diagram of the hyperspectral imaging system. The system consists of a black box, spectrometer, lens, light source, step-motor and computer.

2.3. Hyperspectral Image Acquisition and Correction

In order to obtain the high spectral image, the speed of the transport platform and the height of the lens to samples need to be adjusted. In this study, the exposure time was set to 0.05 s, the running speed of the transport platform was 2.3 mm/s and the average height of the lens to the samples’ surface was 36 cm. In order to eliminate the noise generated by the dark current and the intensity distribution of the light source, Equation (1) was used for correction [16]:

\[
I_c = \frac{R_e - R_d}{R_{\text{aw}} - R_d}
\]  

(1)
where $I_c$ is corrected image, $R_w$ is calibration image obtained by scanning the standard white correction board (reflectance close to 99.9%), $R_d$ is the calibration image when the lens is covered with its cap (reflectance close to 0%). $R_r$ is the original hyperspectral image.

2.4. Data Processing and Analysis

After the hyperspectral images were acquired, ENVI 4.6 (ITT visual information solutions, Boulder, CO, USA, 2009), The Unscrambler X 10.3 (CAMO PROCESS AS, Oslo, Norway, 2015) and Matlab R2012a (version: 7.14, The Math Works, Natick, MA, USA, 2012) were used for data processing and analysis.

2.4.1. Principal Component Analysis

PCA is an effective method for dimension reduction and has been widely used in the field of spectral analysis [17]. On the basis of preserving the original information as far as possible, the samples in high dimensional space can be mapped to the lower dimension principal component space. The basic idea of PCA is to use an optimization method to concentrate measured data (Y), the data matrix is simplified to reduce the number of principal component for a few linear combinations of the original variables, in order to reveal the characteristics of Y data structure and extract basic features. The main components after the dimension reduction of PCA can be expressed in formula (2).

$$PC_s = \sum_{i=1}^{m} \beta_{si} \lambda_i$$

where $PC_s$ is the $s$th principal component, $\lambda_i$ is the $i$th wavelength (A total of $m$ wavelengths), $\beta_{si}$ is the load factor of the $s$th principal component on the $i$th wavelength. The bigger the load factor is, the more effective the principle component is.

Choosing some representative characteristic wavelengths is very important because the data processing and modeling will be influenced by the abundant spectral data and strong inter band redundancy [18]. This study firstly chose five characteristic wavelengths according to the load factor.

2.4.2. ROC Curve Analysis

The ROC curve is a comprehensive index of the specificity of reaction and the sensitivity of continuous variables [19]. This method reveals the reciprocal relationships between sensitivity and specificity by composition method (true positive rate (TPR) as vertical coordinate, false positive rate (FPR) as horizontal coordinate). For the two classification problem in this study, the classification results may appear in four cases as follows. Normal samples are correctly identified (true positive, $TP$), normal samples were misjudged as damage samples (false negative, $FN$), damaged samples were correctly identified (true negative, $TN$), and damaged samples were misjudged as normal samples (false positive, $FP$). For a given threshold, $TPR$ and $FPR$ can be calculated as follows:

$$TPR = \frac{TP}{(TP + FN)}$$

$$FPR = \frac{FP}{(FP + TN)}$$

To a separate classifier, the change of the threshold is finally drawn into a ROC curve. The larger the area under the curve (AUC) is, the better the classification effect is. In this study, each wavelength of PCA is extracted as a classifier to draw the ROC curve, and two characteristic wavelengths were selected to establish the final PLS-DA model by comparing the AUC value of each classifier.
2.4.3. PLS-DA Modeling

PLS-DA is a kind of supervised linear classification method, which establishes the model by giving different kinds of variables the virtual integer value. In this study, all bruised samples were signed to “0” and all normal samples were signed to “1”. Finally, the predicted results of the model were real values, and the attribution of the samples can be judged by taking a certain threshold (0.5 in this study) [20].

3. Results and Discussion

3.1. Reflectance Spectra of Pears

Because the pear is kind of a ball-shaped object, the surface of the pear is bright in the middle but dark at the edge. In order to minimize the effect of illumination on the experimental results, this study chose the normal skin region of interest (ROI) near the equator line of the pears. The spectra were extracted from the ROI region of 60 normal samples and 60 bruised samples, and the average spectra of all bruised samples and normal samples were calculated separately. Considering that noises exist at the beginning and end of the spectrum, a 420 to 1000 nm spectral range was selected for the final analysis.

It can be seen from Figure 3 that the average reflectance of the damaged area was lower than that of the normal region in the 420 to 1000 nm spectral range. In addition, there were two obvious absorption peaks in the graph. The absorption peak at 680 nm is mainly caused by the absorption of chlorophyll on the surface of the fruit [21] which reflects the color information of the fruit surface, while the absorption peak near 960 nm is mainly caused by the water absorption [22].

![Figure 3. Averaged spectra between 420 and 1000 nm of bruised and normal pears of all samples. Two absorption peaks in the black frame represent the absorption of chlorophyll (680 nm) and water (960 nm), respectively.](image)

3.2. The Results of PCA

PCA was performed on all samples in the 420–1000 nm spectrum range. The results are shown in Figure 4. Since the first two principal components of cumulative credibility reached 92.97% (above 90%), the main information of the original spectra could be represented by the first two principal components.
Figure 4 is a clustering plot of the PC1 and PC2 scores for modeling samples, in which the horizontal coordinate represents the first principal component and the vertical coordinate represents the second principal component of each sample. It can be seen from the graph that the bruised samples clustered mainly in the lower-right corner while the normal samples clustered mainly in the upper-left corner. Though these two kinds of samples overlapped in some places, they could be well distinguished in general. Generally speaking, PC1 and PC2 had good effects on clustering.

Figure 5 shows the loading plots of PC1 and PC2 in the spectral region of 420–1000 nm. The peaks and valleys of the loading plots indicated the importance of the wavelengths. Wavelengths 472, 544, 655, 688 and 967 nm were selected as the characteristic wavelengths. The first three wavelengths (472, 544 and 655 nm) were related to the color information of blue (around 450 nm), green (around 550 nm) and red (around 650 nm); the wavelength of 688 nm represented the absorption of chlorophylls; the wavelength of 967 nm represented the absorption of water. As seen in Figure 1, the differences between water and color could be obviously visualized in the bruise region, indicating the efficiency of the characteristic wavelengths’ selection.
3.3. Results of ROC Curve Analysis

To reduce the number of characteristic wavelengths used for making the final model, this study used ROC curve analysis to extract the characteristic wavelengths of the PCA. Wavelengths of 472, 544, 655, 688 and 967 nm were used as a classifier separately, and the results are shown in Table 1 and Figure 6.

Table 1. The area under the curve (AUC) of different wavelengths by the receiver operating characteristic (ROC) curve analysis.

<table>
<thead>
<tr>
<th>Wavelength/nm</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>472</td>
<td>0.992</td>
</tr>
<tr>
<td>544</td>
<td>0.864</td>
</tr>
<tr>
<td>655</td>
<td>0.728</td>
</tr>
<tr>
<td>688</td>
<td>0.663</td>
</tr>
<tr>
<td>967</td>
<td>0.980</td>
</tr>
</tbody>
</table>

Figure 6. The receiver operating characteristic (ROC) curves at different wavelengths of modeling samples: (a) 472 nm; (b) 544 nm; (c) 655 nm; (d) 688 nm; (e) 967 nm.

As previously mentioned, the greater the AUC value of the ROC curve analysis of a classifier is, the better the classification performance of this classifier is. It can be found from the chart that the AUC values at wavelengths 472 and 967 nm were higher than those in other regions, so these two bands are more sensitive to the identification of bruised samples and can better distinguish between normal and bruised samples. Finally, the spectral data at 472 and 967 nm were selected to establish the final PLS-DA model.

3.4. PLS-DA Modeling Based on Two Characteristic Wavelengths

Finally, the two characteristic wavelengths were extracted by the PCA-ROC analysis: 472 and 967 nm. The spectral data of 90 samples (45 bruised samples and 45 normal samples) at the characteristic wavelengths were used to establish the PLS-DA model as the modeling set, and 30 samples (15 bruised samples and 15 normal samples) were used to predict the accuracy of the model as the prediction set. The results were compared with those of the full-wave-band PLS-DA modeling result and are shown in Table 2.
Table 2. Discriminant results of the partial least squares discriminant analysis (PLS-DA) model of modeling samples (45 normal pears and 45 bruised pears) and prediction samples (15 normal pears and 15 bruised pears).

<table>
<thead>
<tr>
<th>PLS-DA Type of Pear Sample</th>
<th>Modeling Set</th>
<th>Prediction Set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample Number</td>
<td>Misjudgment Number</td>
</tr>
<tr>
<td>Full spectral bands</td>
<td>normal 45</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>bruised 45</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Total 90</td>
<td>0</td>
</tr>
<tr>
<td>Characteristic spectral bands</td>
<td>normal 45</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>bruised 45</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Total 90</td>
<td>1</td>
</tr>
</tbody>
</table>

It can be concluded from Table 2 that the recognition results by the model based on characteristic spectral bands were ideal. All of the 45 normal samples of the modeling set and the 15 normal samples of the prediction set were correctly identified, while 44 of the 45 bruised samples of the modeling set and all 15 bruised samples of the prediction set were correctly identified. The overall accuracy of the modeling and prediction sets of the model established on the two characteristic wavelengths (472 and 967 nm) reached 98.9% and 100%. The accuracy was similar to that of the model established on the full band, which was 100% of the modeling set and 100% of the prediction set. However, by extracting the characteristic wavelength, the number of wavelengths needed in the detection process was greatly reduced, which was more suitable for on-line detection. Compared with previous studies which detected bruised pears using the hyperspectral imaging technique [4,9,10], this paper reduced the amount of wavelengths to two using PCA-ROC analysis and improved the detection accuracy of the model.

4. Conclusions

In this study, the characteristic wavelengths for the damage detection of Korla pears were extracted by using hyperspectral technology combined with stoichiometry, and the feasibility of the method was verified. Firstly, five wavelengths (472, 544, 655, 688 and 967 nm) sensitive to damage were extracted by PCA of the original spectra of the pear samples. Then, two characteristic wavelengths were extracted by ROC analysis when every single wavelength was used as a classifier. Finally, the PLS-DA model was established on the two characteristic wavelengths (472 and 967 nm) and the recognition accuracy of the modeling and prediction sets reached 98.9% and 100%, which are similar to that of the model established on the full band. In conclusion, it is feasible to use this method to detect bruised pears, in which the PCA and ROC curve analysis were used to methodically select the inspection to detect bruised pears. Also this research will provide a reference for future research on damage detection systems of pears and other fruits.


Author Contributions: Hao Jiang, Yong He and Yande Liu conceived and designed the experiments; Hao Jiang and Xinxin Chen performed the experiments; Hao Jiang, Chu Zhang and Fei Liu analyzed the data; Hao Jiang, Chu Zhang, Yong He, Xinxin Chen, Fei Liu and Yande Liu wrote the paper.

Conflicts of Interest: The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References


