



Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.journals.elsevier.com/spectrochimica-acta-part-a-molecular-and-biomolecular-spectroscopy



Spectrum classification of citrus tissues infected by fungi and multispectral image identification of early rotten oranges

Wei Luo ^{a,1}, Guozhu Fan ^{a,b,1}, Peng Tian ^a, Wentao Dong ^{a,1}, Hailiang Zhang ^{a,*}, Baishao Zhan ^{a,*}

^a College of Electrical and Automation Engineering, East China Jiaotong University, Nanchang 330013, China

^b State Grid Jiangxi Extra High Voltage Company, Nanchang 330013, China



ARTICLE INFO

Keywords:

Fruit quality and safety
Hyperspectral image processing
Wavelength image selection
Spectrum modeling
Rotten orange detection

ABSTRACT

Citrus fruit is susceptible to postharvest rot by fungal infection. The detection of early rot is difficult due to similar skin characteristics to sound area, which limits the ability of the grading system to evaluate the comprehensive quality of citrus. In this study, the visible and near infrared hyperspectral imaging system with the wavelength range of 325–1000 nm was used to collect hyperspectral images of oranges. Hyperspectral data of three types of tissues including sound tissue from 80 samples, rotten tissue infected by *Penicillium digitatum* from 100 samples and rotten tissue infected by *Penicillium italicum* from 100 samples were extracted. The bootstrapping soft shrinkage (BOSS) and BOSS-SPA (BOSS-Successive Projections Algorithm) combination algorithm were separately used to optimize spectrum variables. The partial least squares discriminant analysis (PLS-DA) model for classifying three types of tissues and PLS-DA model for classifying two types of tissues (sound tissue and rotten tissue) were constructed based on full-spectrum and the selected informative variables. Model comparison showed that the BOSS-PLS-DA model can effectively identify three types of tissues with the classification accuracy of 97.1%, while the BOSS-SPA-PLS-DA model was more effective for the binary classification of sound and rotten citrus tissues with the accuracy of 100%. Furthermore, the wavelength images corresponding to the nine informative variables extracted by BOSS-SPA were performed the principal component analysis (PCA), and four feature wavelength images (508, 568, 578 and 614 nm) were obtained by analyzing the weighting coefficients of each single-wavelength images constituting the optimal principal component (PC) image. Finally, a fast multispectral image processing algorithm combined with the global threshold theory was proposed for the rotten orange detection based on the extracted four wavelength images. A total of 280 samples including 80 sound and 200 rotten samples were used to evaluate the classification ability, which showed the proposed multispectral image detection algorithm can successfully differentiate between sound and rotten oranges with an overall classification accuracy of 98.6%.

1. Introduction

Citrus is one of the four major fruits in the world. The planting area and output rank first in the world, and its annual trade volume ranks third in the world agricultural trade. Citrus is fragrant, delicious and nutrient-rich, and has high planting income. These advantages make citrus become the fruit with the largest planting area, the highest yield and the largest consumption in China. Automatic grading after harvest is an important means to improve the quality and market competitiveness of citrus fruit. There are many quality grading indexes including external ones (such as size, color and skin defects) and internal ones (such as

sugar and acidity). According to these physical and chemical indexes, different grades of citrus can be sold to consumers with different needs and transported to different markets. However, no consumers or markets can accept citrus with signs of decay. Fungal infection, especially *Penicillium digitatum* and *Penicillium italicum*, is the main cause of decay of fresh citrus, [29,13]. The *Penicillium digitatum* (green mold) and *Penicillium italicum* (blue mold) are the most economically important pathogens in citrus, resulting in significant postharvest losses (up to 30% and 80%, respectively) [30,11].

Chemical and non-chemical treatments are the main means to control *Penicillium digitatum* and *Penicillium italicum* of citrus in the storage

* Corresponding authors.

E-mail addresses: hailiang.zhang@163.com (H. Zhang), weil_ecjtu@163.com (B. Zhan).

¹ Wei Luo and Guozhu Fan contributed equally to this work.

and transportation stage [30,10]. In terms of chemical control, the long-term use of chemical bacteriostatic agents will make the fungal pathogens produce antibodies themselves [15], which may require larger chemical doses. Thus, it will inevitably lead to food safety and environmental pollution. Although non-chemical control has a certain inhibitory effect on fungi and there is no potential danger, there are still some problems such as unsustainable efficacy or poor general application. In the early stage of decay, the correct identification and classification of pathogens is helpful to develop fungi-specific fungicides, so as to save product and probably reduce the use of chemicals and the impact of drug resistance [13]. The *Penicillium digitatum* and *Penicillium italicum* are common fungi. Their conidia are widely distributed in the environment and can easily invade citrus fruit through wounds during harvest and cause fruit decay [26,30]. This indicates that the pathogen already exists in some citrus and inevitably causes decay before any further control treatment. Therefore, the accurate sorting of early rotten fruit before storage and transportation can not only greatly control the quality of the whole batch, but also effectively prevent the disease from spreading to other sound citrus. This is of great significance to reduce economic losses.

Decay process in citrus fruit implies changes in enzymatic activity, resulting in an enhanced water-soluble pectin fraction, and consequently, weakening of the cell wall [24]. The subsequent water soaking and slight pigment change of the epidermal tissue are the early visible symptom of infection in citrus. Many studies used different techniques to identify this type of defect. These technologies include near infrared spectroscopy [22,12], RGB color imaging [5], fluorescence imaging [28], Laser-light backscattering imaging [23], structured light reflection imaging [16], hyperspectral imaging [19,17]. In terms of near infrared spectroscopy, it is effective in distinguishing sound tissue from rotten tissue. A maximum overall classification accuracy of 97.8% was achieved from study of Lorente et al. [22]. However, this technique does not have the ability to detect the whole surface information of citrus. Decay can cause chemical changes of fruit components. Some studies showed that full transmission spectroscopy can obtain most of the internal component information of fruit [37], but for slight decay, the full transmission spectrum is very limited in characterizing compositional variation. Therefore, the imaging technology was more commonly used for fruit defect detection due to the acquisition ability of spatial information.

RGB (Red, Green and Blue) color imaging was a widely used machine vision technology for fruit defect detection [6]. This technique was very effective in detecting some surface defects visible to the naked eye, but it was not suitable for detecting early decay of citrus due to the insignificant difference between the decayed and the sound area [18]. Fluorescence imaging was considered to be a potential technology for identification of rotten citrus fruit. Because the rotten area usually produces fluorescence under the induction of ultraviolet light, and this fluorescence phenomenon was easy to be captured by color camera [4,28]. Unfortunately, not all types of citrus fruits can excite fluorescence from rotten areas under the induction of ultraviolet light [33,27,28]. Even those citrus fruit that are easy to be excited to fluorescence, such as navel orange, also have different fluorescence intensity due to the difference of decay degree. Low intensity fluorescence was difficult to be effectively identified in the rapid detection of fruit quality. Laser-light backscattering imaging can detect the rotten area to a certain extent, because there were tissue structure differences between the rotten and sound areas. However, this technology has great limitations, and it usually requires laser aiming at the rotten area to obtain better identification results [22]. In fact, the fruit movement was random during the online inspection, and the rotten area of citrus cannot be directly facing the laser light. Structured light reflection imaging was a relatively new technology for the detection of rotten citrus. Based on this technology, it was usually necessary to obtain three independent structured light fringe images that have defined phase angles, and then the classical three-step phase shift method was used to demodulate the

structured light image to obtain the target image used for rot detection [21]. However, in the automatic detection of citrus quality, it was impossible to obtain the structured light image with specific phase angle.

In recent years, many studies used hyperspectral imaging to identify rotten citrus fruit [24,13,19,17,38]. These studies showed that the hyperspectral imaging technology can be successfully used to detect rotten citrus with an accuracy of 93%-98.6%. Different from near infrared spectroscopy and traditional imaging technology, hyperspectral imaging integrates the advantages of spectrum and image to achieve a complete fusion of both. The differences in chemical composition, physical structure and epidermal characteristics between rotten and sound tissues could be reflected in hyperspectral image to varying degrees. Thus, the early rotten area of citrus can be accurately detected by appropriate spectral image processing. The main problem of hyperspectral imaging lies in the long time of image acquisition and processing [31]. Therefore, for performing a fast detection task, some studies focused on the development of multispectral imaging system based on wavelength images obtained by analyzing hyperspectral data. For example, Huang et al. [14] developed a multispectral imaging system to detect apple bruising using three cameras coupling with wavelength images at 780, 850 and 960 nm. Study from Calvini et al. [7] indicated that the results from hyperspectral imaging can be transferred to a filter-based multispectral imaging system for on-line monitoring of green coffee. There were many wavelength selection algorithms for hyperspectral images, among which the weighting coefficient analysis based on principal component (PC) was the most commonly used method [20]. However, a hyperspectral image usually contains hundreds of single-wavelength images, thus, the principal component analysis (PCA) of a large number of wavelength images requires a long time of data processing. At the same time, it also puts forward higher requirements for the processing performance of computer, especially for the task of mass sample processing. Different from previous studies [19,17,38], this study tried to extract a small number of wavelength images by the combination of citrus tissue classification, spectral variable selection and wavelength image weighting coefficient analysis, so as to improve the processing speed of hyperspectral images and the selection efficiency of feature wavelength images.

The accurate segmentation of regions of interest (ROIs) based on target image was an important step to realize the effective recognition of rotten citrus fruit. Many image segmentation algorithms can be used, such as region growth algorithm [5], watershed algorithm [17], automatic threshold method, global threshold method, etc. In comparison, the simple threshold method was more suitable for rapid detection of fruit quality. Therefore, the automatic threshold method and global threshold method were preferred to be used in this study to segment rotten areas on citrus. In order to achieve accurate segmentation, it was necessary to obtain high quality target image. In this way, the selection of feature wavelength image was very crucial.

Based on the above analysis, the objectives of this study were: (1) to construct the classification models based on hyperspectral information of three types of tissues (sound tissue, rotten tissue infected with *Penicillium digitatum* and rotten tissue infected with *Penicillium italicum*) to verify the feasibility of hyperspectral data in classifying different tissues; (2) to extract the informative spectral variables based on hyperspectral data by BOSS and BOSS-SPA methods, and to compare the classification performance of full-spectrum and informative variable models for different tissues; (3) to determine a small number of feature wavelength images by combining informative spectral variable selection and wavelength image weighting coefficient analysis; (4) to develop the detection algorithm of early rotten citrus based on multispectral images and threshold theory, and the identification performance of the proposed algorithm was evaluated.

2. Materials and methods

2.1. Fruit samples

This study used 280 samples including 80 sound samples, 100 early rotten samples infected by *Penicillium digitatum* and 100 early rotten samples infected by *Penicillium italicum*. Orange (Newhall Navel Orange) was selected as the research samples, because this kind of citrus fruit is very popular and has high economic value in China. Two types of rotten samples were obtained by artificially inoculating sound oranges with *Penicillium digitatum* and *Penicillium italicum* fungi, respectively. For sample inoculation, the main steps were as follows: (1) collect some oranges naturally infected by fungi *P. digitatum* and *P. italicum*; (2) extract fungal spores from the infected fruit and dissolve in water to form the spore solution; (3) manually inject about 20 μL of fungal spore solution (5 mm below the epidermis) into each sample and (4) store the inoculated samples in box at 25 °C and 99% relative humidity. For more detailed information, please refer to Li et al. [17]. Three days after fungal inoculation, an early rotten area can be formed in the inoculated area. There was no obvious color difference between this area and sound area, but the rotten area was usually soft because the weakening of the cell walls due to changes in enzymatic activity [2,3]. With the increase of infection time, the infected area gradually grew a mold layer containing fungal spores. For the area infected by *Penicillium digitatum*, the mold layer was green. For the area infected by *Penicillium italicum*, the mold layer was cyan. At this time, it was easy to identify the infected area and fungi type by the naked eye. Fig. 1 shows RGB images of different samples. Fig. 1(A) shows the sound sample. Fig. 1(B) and (C) show typical decayed samples after three days of inoculation with *Penicillium digitatum* and *Penicillium italicum*, respectively. In Fig. 1(B) and (C), the upper right corner shows the mold layer growing in the infected area at the late stage of sample infection.

For all samples, 60 sound samples and 160 rotten samples (80 for each type) were randomly selected to form the training set, and the remaining samples were used as testing set. In this way, the training set including 220 samples was used for the construction of classification models and the development of detection algorithms. The testing set including 60 samples was used to assess the performance of the models and algorithms.

2.2. Hyperspectral imaging system, data acquisition and image correction

The hyperspectral images of all samples were acquired by using a hyperspectral imaging system as shown in Fig. 2. The spectral imaging range of system was 325–1100 nm, and the adjacent spectral interval was about 0.778 nm. The system mainly included an imaging spectrograph (ImSpector V10E-QE, Spectral Imaging Ltd., Oulu, Finland), an area-array CCD (Charge Coupled Device) camera (ICL-B1620, Imperx, Boca Raton, FL, USA), two halogen light source (Antefore International Co., Ltd., Taiwan, China), an electronically-controlled sample

movement stage (EZHR17EN, AllMotion, Inc., USA) and a computer (Lenovo, Inter(R)Core™i5, RAM 16.00G). The computer was equipped with sample image acquisition control software and sample movement stage control software (Isuzu Optics Corp., Taiwan). In order to avoid the distortion of the sample image and the saturation of the collected spectral data, after several tests, the moving speed of sample stage, the exposure time of camera and the object distance were adjusted to 15 mm/s, 5 ms and 45 cm, respectively. In the process of image acquisition, the orange was first placed on the sample stage (Note that the side containing the rotten area needs to face the camera when the image of rotten sample was acquired). Then, the hyperspectral image acquisition software and the movement control software of sample stage were turned on. When the sample stage moved the tested sample through the camera's field of view, the sample information was recorded by the imaging spectrograph and camera, and the image data of the tested sample was stored synchronically on the computer. The acquired hyperspectral image was the three-dimensional data cube with two spatial dimensions and one spectral dimension as shown in the lower right corner of Fig. 2. In other words, a spectrum can be extracted at each pixel in the hyperspectral image, and a wavelength image can be also extracted at each wavelength. It can be seen from the pixel spectrum curve shown in Fig. 2 that there were large noises at both ends of the spectrum, indicating that the signal-to-noise ratio of the wavelength images corresponding to these spectra was low. Therefore, this study only used hyperspectral images in the wavelength region of 450–1000 nm (a total of 707 wavelength images) for analysis.

After obtaining hyperspectral images (R_{raw}) of samples, white reference image (R_{white}) and dark reference image (R_{dark}) need to be collected under the same system parameters for image correction. The R_{dark} with 0% reflectance was collected by turning off light source. The R_{white} with approaching 100% reflectance was acquired from a white board (white calibration tile, Specim, Oulu, Finland). The corrected image R_c can be calculated by the following formula:

$$R_c = \frac{R_{\text{raw}} - R_{\text{dark}}}{R_{\text{white}} - R_{\text{dark}}} \quad (1)$$

The corrected hyperspectral images were used for subsequent analysis.

2.3. Spectral classification of citrus tissues

2.3.1. Different tissue classification model construction

One of the objectives of this study was to construct the classification model to verify the feasibility of classifying two types of citrus tissues (sound tissue and rotten tissue) and three types of citrus tissues (sound tissue and two types of rotten tissues) based on hyperspectral data. Therefore, the representative spectra of different tissues need to be first extracted (Note: since each pixel in the hyperspectral image has spectral information, in order to make the spectrum of each type of tissue more representative, a region containing about 1000 pixels was selected for each sample as the region of interest (ROI), and the average spectrum of

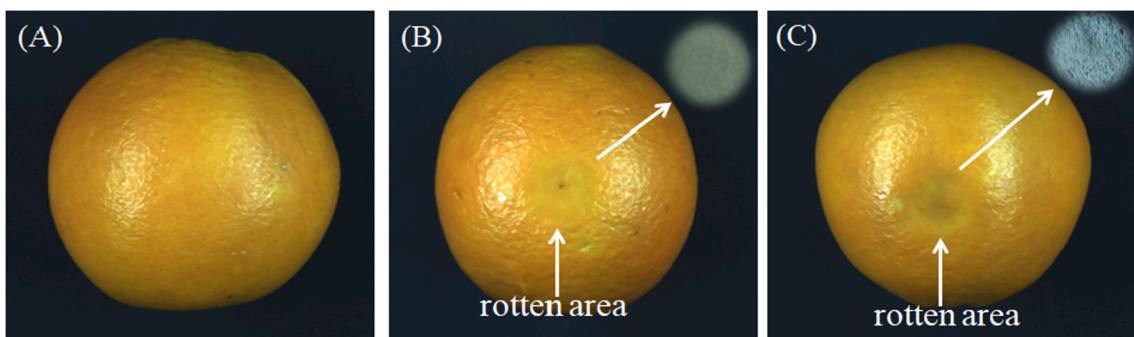


Fig. 1. RGB color images of the typical samples. (A) Sound sample, (B) sample infected by *Penicillium digitatum* and (C) sample infected by *Penicillium italicum*.

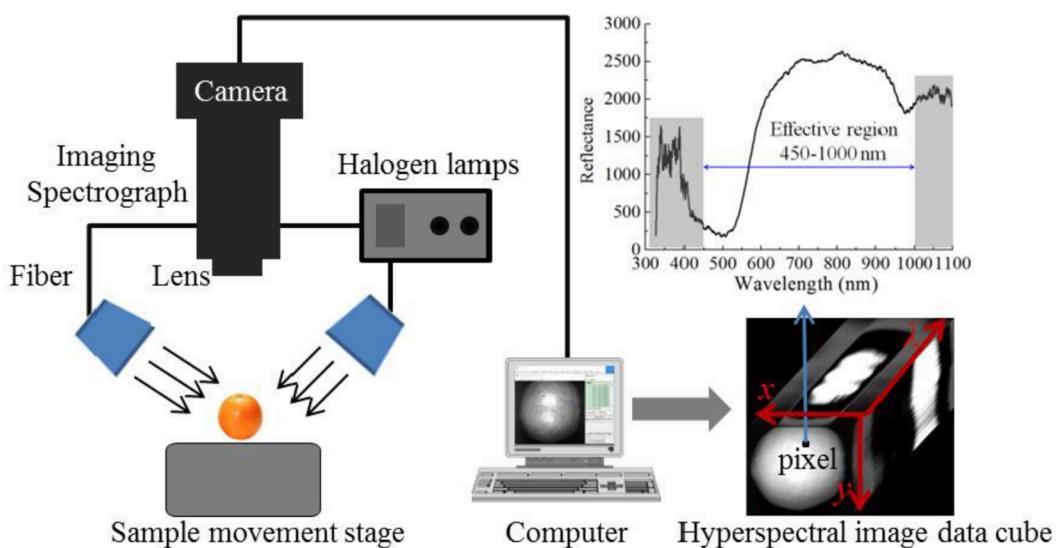


Fig. 2. Hyperspectral image collection system and data cube.

ROI was then calculated as the representative spectrum of the corresponding tissue). Then, the representative spectral data of tissues of samples in the training set were used as inputs to construct PLS-DA models.

The partial least squares for discrimination analysis (PLS-DA) [1], a linear classification method combining the properties of partial least squares regression with the discrimination power of a classification technique, was used for classification. This method established models between multivariate data and a vector coding different classes (here, the two and three classes). Multivariate data was represented by a matrix \mathbf{X} of size (n, p) where n was the observation number and p the variable number. The n observations were identified by their corresponding class in the vector \mathbf{y} of size $(n, 1)$ where values ranged from 1 to q , where q was the class number. The first step was to transform \mathbf{y} into a dummy matrix \mathbf{Y} of size (n, q) .

$$\mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \\ 2 \\ 3 \\ 3 \\ 3 \end{bmatrix} \rightarrow \mathbf{Y} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

Taking the three classes as an example, a dummy matrix was given in Eq. (2), in which nine observations belong to three classes. The matrix \mathbf{Y} contains binary values (0, 1) where each column corresponded to a class, reporting the 1 value if the corresponding row (object) belongs to the class, and 0 otherwise. Then, a PLS2 regression model was applied between \mathbf{X} and \mathbf{Y} [32,7]. The PLS-DA predicted value is a real number, not a dummy integer. In order to determine which class a tissue belongs to, a cutoff value needs to be set [35]. Here, the cutoff value was set as 0.5. Additionally, in this study, the PLS-DA model with leave-one-out CV was applied to prevent over-fitting of the calibration model and classify two classes and three classes of orange tissues. The optimal number of latent variables (LVs) was determined by the lowest value of predicted residual error sum of squares.

2.3.2. Classification model optimization

The original hyperspectral data contained 707 wavelengths, which were not all related to tissue classification. Moreover, too many variables can easily lead to over-fitting during modeling, which reduced the generalization performance of model, increased the calculation time of model, and even affected the prediction efficiency and accuracy of model [36]. In addition, too many variables were not conducive to explain the contribution of variables. In order to eliminate the influence of noise and useless variables, variable selection was usually the key step in the near-infrared spectrum modeling. Variable selection played an important role in simplifying the near-infrared spectrum prediction model and improving the prediction accuracy of model.

This study used a new spectrum variable selection method called bootstrapping soft shrinkage (BOSS) for variable selection. BOSS method was proposed by Deng et al. [9]. The algorithm introduced Bootstrap Sampling (BSS) and Weighted Bootstrap Sampling (WBS) methods to generate random variable combination to construct sub-models. Model population Analysis (MPA) [8] was used to extract variable information from sub-models. The BOSS selected a subset of valid variables as follows:

Let different types of citrus tissue data set $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, where x_i represents spectral data, dimension P ; y_i is the corresponding class value.

Step 1: BSS is applied on variable space to generate K subsets. Variables selected by BSS in each dataset will be extracted without duplication. In this step, all variables have the same probability of being selected, that is, all variables have equal weights.

Step 2: PLSR (Partial Least Squares Regression) sub-models are established based on the obtained subsets, and then the predictive error (the root mean square error for cross-validation, RMSECV) of all sub-models was calculated to extract best models (e.g. 10%) with the lowest RMSECV.

Step 3: After calculating the regression coefficients of each model, all elements in the regression vector are changed into the absolute values, and each regression vector is standardized. The new weight of variables can be obtained by summing the standardized regression vector.

Step 4: WBS is utilized to generate new subsets based on new variable weights. Similar to BSS, this method also requires that there are no duplicate variables in the extracted and used variables. Let the number of variables in the previous iteration be M , then number of replacement in WBS depends on the average number of M , and thus,

the number of variables contained in the new subset is close to 0.632 M. This step ensures that variables with larger absolute values of regression coefficients in the optimal sub-model are more likely to be selected into the sub-model.

Repeat steps 2–4 until the number of variables in the new subset is 1. Finally, the subset with the smallest RMSECV is the optimal variable set.

In addition to BOSS for variable selection, this study also first attempted to combine BOSS and SPA variable selection methods for variable selection, because many studies showed that the combination of two complementary wavelength selection theories can achieve the superposition effect to obtain better wavelength selection effect [34]. SPA was a forward wavelength selection technique designed to minimize collinearity problems for multivariate calibration. SPA employed simple projection operations in a vector space to obtain subsets of wavelengths with minimal collinearity. The final wavelength number was determined by the smallest root mean square error of prediction (RMSEP) in validation set of MLR (Multiple Linear Regression) calibration. For BOSS-SPA combination variable selection, BOSS was first used to extract a set of wavelengths related to citrus tissue classification, and then SPA was used to extract the most informative variables from the selected wavelengths by BOSS.

2.4. Identification of early rotten oranges

2.4.1. Feature wavelength image extraction

The selection of feature wavelength image was the key to build a fast multispectral imaging system for detection of the rotten citrus fruit. Generally, when the classification accuracy meets the requirements, the fewer the number of feature wavelength images used, the better. This study proposed a feature wavelength image extraction strategy by combining hyperspectral variable selection of citrus tissues with weighting coefficient analysis of target PC image. Firstly, the wavelength images corresponding to the informative variables that were obtained by BOSS or BOSS-SPA were extracted. Then, these wavelength images were used for principal component analysis to obtain the PC images. The PC image with high contrast between the rotten area and the sound area was selected as the target PC image. The weighting coefficients of all wavelength images constituting the target PC image were further analyzed. The wavelength images corresponding to the local maximum and minimum of the weighting coefficient curve were considered as the feature wavelength images [25].

2.4.2. Segmentation of region of interest

In order to realize the rapid and accurate identification of rotten citrus fruit, two threshold segmentation algorithms including Ostu and global threshold methods were compared. These two threshold theories were frequently used in the rapid detection of fruit defects. The Ostu method can obtain a satisfied result when the image histogram was bimodal. The global threshold value method was usually used when the object of interest was well contrasted against the background. Before the threshold segmentation algorithm was implemented, the target image for region of interest segmentation needs to be first obtained. Here, the target image was obtained by principal component analysis of feature wavelength image. After obtaining the target image, the background of target image was removed by constructing a binary mask image, and then filtering was performed to remove the noise in the target image. Finally, two threshold segmentation theories were used to segment potential rotten regions in the target image, respectively. If there was no white area (rotten area) in the resultant binary image, the sample was considered to be sound orange, otherwise, the sample was rotten fruit.

In this work, Environment for Visualizing Images Software Program (ENVI 4.6, Research System Inc., Boulder, CO., USA) was utilized to extract the spectra of ROIs, the Unscrambler V9.7 software (CAMO PRECESS AS, Oslo, Norway) was applied to develop PLS-DA models, and Matlab R2008b software (The Math Work, Inc., Natick, MA, USA) was

used for wavelength selection and image processing.

3. Results and discussion

3.1. Spectral analysis of different tissues

The spatial distribution of typical spectrum curves of different sample tissues was shown in Fig. 3. The shaded regions with different colors in the figure represent the spectrum distribution range of different tissues. The blue shadow represents the sound sample tissue, the light red shadow represents the rotten tissue infected by *Penicillium digitatum* fungi, and the light green represents the rotten tissue infected by *Penicillium italicum* fungi. It can be seen that, compared with spectra of sound tissue, the spectra of two types of rotten tissues have serious scattering, which was mainly because spectra of sound tissue were extracted from the middle region of all samples, while spectra of the rotten tissue were extracted from the edge or middle region of the fruit in the hyperspectral image due to the uncertainty of the distribution of rotten regions. Therefore, the extracted spectra of rotten tissues can be affected by the uneven light distribution on spherical fruit, which will lead to great differences in spectral intensity. It can also be seen that the change tendency of spectrum curves of the three types of tissues was similar, and the spectral intensity overlaps seriously. In the overlapping area (blue area), there was no obvious difference in intensity, indicating that it was difficult to distinguish different types of tissues only based on the spectral intensity. Therefore, it was necessary to construct the appropriate model to mine the physical and chemical characteristics contained in the spectra, so as to realize the classification and discrimination of different tissues.

3.2. Variable selection by BOSS and BOSS-SPA

The BOSS algorithm and the BOSS-SPA combination algorithm were respectively used to select the most informative variables from the original 707 hyperspectral variables to classify citrus tissues. Taking the classification of two types of tissues as an example, this study introduced the variable selection results obtained by BOSS algorithm. Fig. 4 shows the variable selection results by BOSS-SPA for the classification of sound and rotten orange tissues. The evolution of RMSECV, variable number and weights in sub-models in each iteration of BOSS were shown in Fig. 4(A), (B) and (C), respectively. It can be observed from Fig. 4(A) that the RMSECV in the sub-models decreased during the iteration and reached the minimum value at iteration 8. Therefore, the optimal variable set was obtained at iteration 8. As shown in Fig. 4(B), the optimal

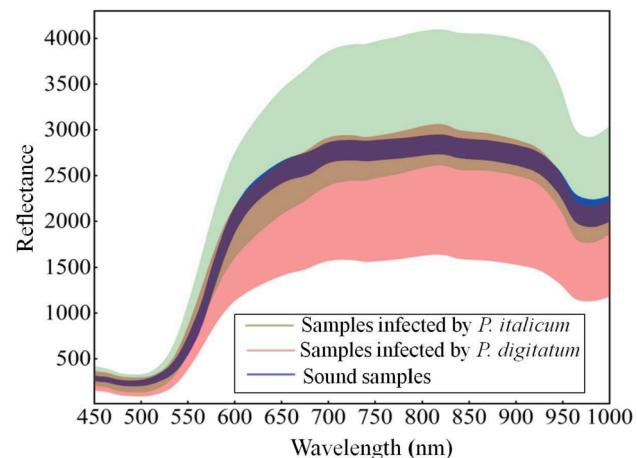


Fig. 3. Spectral curve change range of different tissues of all samples. Note: *P. digitatum* and *P. italicum* stand for *Penicillium digitatum* and *Penicillium italicum*, respectively.

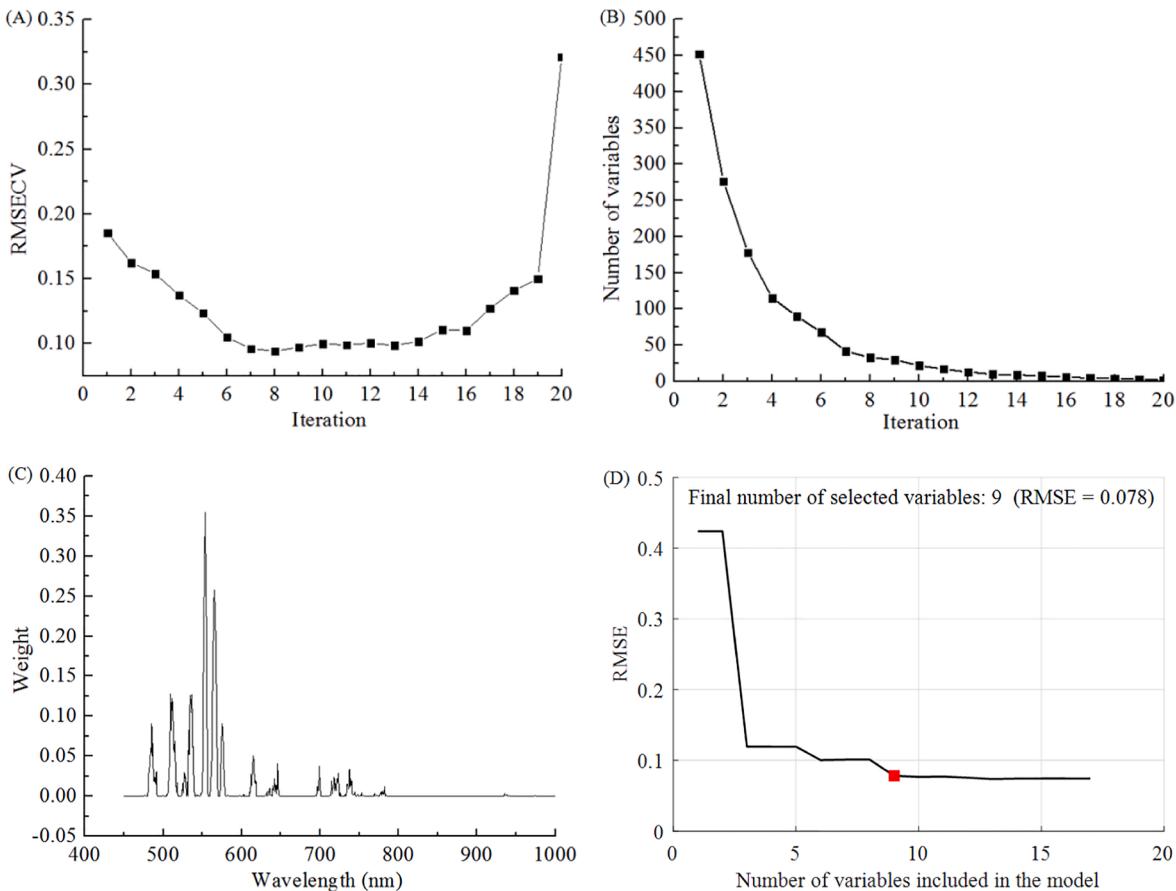


Fig. 4. Variable selection by BOSS-SPA for the classification of sound and rotten orange tissues. The evolution of RMSECV (A), variable number (B) and weights (C) in sub-models in each iteration of the BOSS algorithm. (D) Stand for variation of RMSE with the number of variables selected by SPA.

variable set included 33 variables (That is, the ordinate value corresponding to the eighth iteration). Moreover, it can be also found that, the number of variables decreased smoothly for each iteration and the number of variables decreased gradually from full-spectrum 707 variables to 1 in 20 iterations. This method considered the subset of variables at different level of variable number, which was reasonable because the optimal number of variables was unknown before and during variable selection. The optimal variable set included 33 variables. Fig. 4(C) shows the weights of variables at iteration 8. It can be seen that the most informative 33 variables were distributed at around 475–780 nm and 940 nm. Subsequently, SPA was used for variable selection based on 33 variables. Fig. 4(D) shows variation of RMSE with the number of variables during variable selection of SPA. The red solid block in the figure indicated the optimal number of the variables selected. It can be seen that only nine variables were extracted. Similarly, based on the BOSS algorithm, 21 variables for classifying three citrus tissues were selected from full spectra.

Fig. 5 shows distribution of the selected variables for classifying three classes of tissues and classifying two classes of tissues. It should be noted that the BOSS-SPA combination method was not used for variable selection in the classification of the three types of sample tissues, because the classification results of the model constructed based on the variables selected by BOSS-SPA were poor with an overall classification accuracy of only 65%. On the one hand, the reason may be that the 21 variables selected by BOSS were already the most informative variables with a small collinearity. On the other hand, the reason may be that the difference in physical and chemical properties of the three types of sample tissues, especially the two types of rotten tissues, was small, thus too few variables were not enough to effectively distinguish them. It can be seen from Fig. 5 that most of the selected informative variables were

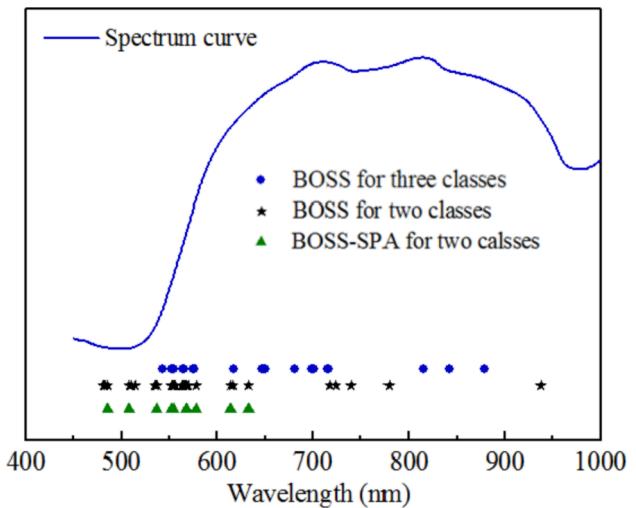


Fig. 5. Distribution of the selected variables for classifying three classes and two classes of tissues.

located in the visible region of 475–750 nm, regardless of the classification of three types of tissues or the classification of two types of tissues, indicating that the differences between different tissues mainly come from the peel features such as color and texture [2,3,24]. In addition to the main contribution of the visible light to tissue classification, it should also be noted that three wavelengths (815 nm, 842 nm and 879 nm) in the near-infrared region were obtained in the three types

of tissue classification, while they were not selected in the two tissue classification. It showed that these three wavelengths that were related to the absorption of C-H groups were more helpful to distinguish between rotten tissues affected by *Penicillium digitatum* and rotten tissues affected by *Penicillium italicum*. In the binary classification of sound tissue and rotten tissue, the BOSS algorithm not only selected the wavelengths in the visible spectrum region, but also selected a wavelength (938 nm) related to water absorption, indicating that the water content of the two classes of rotten tissues was also different. Further, only 9 wavelengths (485, 508, 537, 552, 554, 568, 578, 614 and 633 nm) in the 485–633 nm region were extracted from 33 wavelengths by BOSS-SPA algorithm, and most of wavelengths were located in the yellow-orange band, indicating that the binary classification of sound tissue and rotten tissue was largely based on the surface color difference, which can also be observed from the RGB image of sample.

3.3. Classification results of different tissues by PLS-DA models

The classification results of different tissues based on PLS-DA models with full-spectrum and the selected variables were shown in Table 1. As can be seen, the prediction performance of PLS-DA model based on the selected variables was better than that of the full-spectrum model regardless of the binary classification or the tripartite classification, indicating the effectiveness of the suggested BOSS and BOSS-SPA variable selection methods. Especially for the tripartite classification, variable selection can obviously improve the classification accuracy of the model for two types of rotten tissues. Based on the BOSS-PLS-DA model, the classification accuracies of three types of tissues (sound tissue, rotten tissue 1 and rotten tissue 2) were 100%, 97.5% and 96.3% for training set, 100%, 95% and 90% for testing set, respectively. Compared with classification of three types of tissues, the two classification results of sound tissue and rotten tissue were better. Both BOSS-PLS-DA model and BOSS-SPA-PLS-DA model obtain 100% classification accuracy for samples in the training set and testing set, especially for BOSS-SPA-PLS-DA model that used only about 1.27% (nine variables) of the original variable information. In general, the results in Table 1 showed that the selected spectral variables combined with appropriate classification models can be used to effectively distinguish three classes or two classes of tissues. Further consideration, in the commercial grading of citrus quality, all rotten citrus must be removed regardless of the category of rotten citrus. Therefore, the binary classification of sound and rotten fruit was only considered, combined with the selected 9 wavelengths by BOSS-SPA, in the development of multispectral algorithm.

3.4. Principal component analysis and multispectral image selection

3.4.1. Determination of the optimal principal component image

In the online detection of rotten citrus, the whole surface information

of fruit needs to be evaluated. Therefore, PLS-DA model for local tissue classification was insufficient to achieve this task due to the lack of spatial information. However, it can be seen from the above analysis that the nine wavelengths selected by BOSS-SPA combination algorithm can effectively distinguish rotten tissue from sound tissue. In this way, the wavelength images corresponding to nine variables were extracted to perform principal component analysis. The first seven PC images were shown in Fig. 6. RGB image for a reference, the black arrow points to the rotten area. In the PC image, the marked red elliptical areas were the bright spot areas. As can be seen from Fig. 6, the feature of rotten area was obvious in PC2, PC4, PC5 and PC6 images, but PC2 and PC4 were greatly affected by the uneven illumination of spherical fruit. In contrast, PC5 and PC6 images were more suitable for the extraction of rotten areas. Comparing PC5 and PC6 images, there were many small black spots in the PC6 image with the similar intensity to the rotten area, indicating that PC6 was more easily disturbed by the sample surface noise. Therefore, PC5 image was selected as the optimal principal component image. However, if PC5 was directly used as the target image for the development of rotten fruit detection algorithm, the development cost of multispectral imaging system will be very high because nine wavelength images were used. Moreover, too many wavelength images were also not conducive to the implementation of rapid detection task.

3.4.2. Multispectral image selection and principal component analysis

Each PC image was obtained by summing the single wavelength image multiplied by the corresponding weighting coefficient. By evaluating the weighting coefficients, those single wavelength images that contribute significantly to the optimal PC image (PC5 image shown in Fig. 6) can be selected as the feature wavelength images. Fig. 7(A) shows the weighting coefficient curve of PC5 image. It can be seen that the weighting coefficients corresponding to four wavelengths (508, 568, 578 and 614 nm) were located in the peaks and valleys of curve, so these images at four wavelengths were selected as the feature wavelength images. However, it should also be noted that two of four wavelengths (568 nm and 578 nm) had maximum and minimum weighting coefficients. Therefore, this study will compare the two groups of feature wavelength images, which were 568, 578 nm and 507, 568, 578, 614 nm, respectively. Fig. 7(B) and (C) show the result images obtained by PCA based on these two groups of feature wavelength images, respectively. In the PC images, the marked areas had high intensity, which was caused by uneven light distribution on the fruit surface. By contrast, PC2 image obtained using two feature wavelength images and PC4 image obtained using four feature wavelength images have the potential to be used as the target images for rotten fruit identification. In terms of PC4, the uneven illumination had little effect on the segmentation of rotten area, because the bright spot in the original image presented white area in PC4 image, while the rotten area presented black area.

Table 1

Classification results of different tissues based on PLS-DA models with full-spectrum and the selected variables, respectively.

Class	Input variables	Variable number	LVs	Classification accuracy (%) of training set (n = 220)			Classification accuracy (%) of testing set (n = 60)			Total (%)
				Sound (60)	Rotten tissue 1 (80)	Rotten tissue 2 (80)	Sound (20)	Rotten tissue 1 (20)	Rotten tissue 2 (20)	
Three classes	Full-spectrum BOSS	707	9	90.0	95.0	72.5	100.0	90.0	70.0	85.7
		21	5	100.0	97.5	96.3	100.0	95.0	90.0	97.1
Two classes	Full-spectrum BOSS BOSS-SPA	707	9	100.0	100.0	100.0	100.0	97.5	100.0	99.6
		33	6	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		9	5	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Note: Rotten tissue 1 and Rotten tissue 2 represent samples infected by *Penicillium digitatum* and *Penicillium italicum*, respectively.

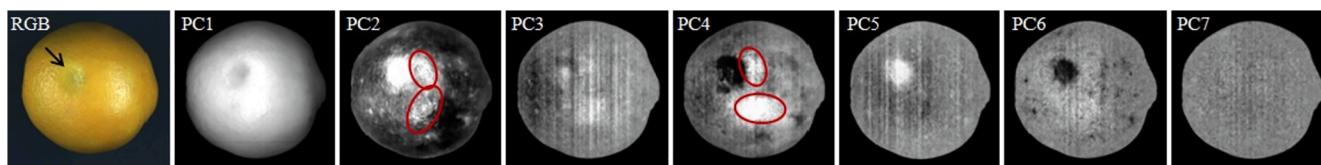


Fig. 6. The first seven PC images that are obtained by PCA based on wavelength images corresponding to the nine informative variables selected by BOSS-SPA.

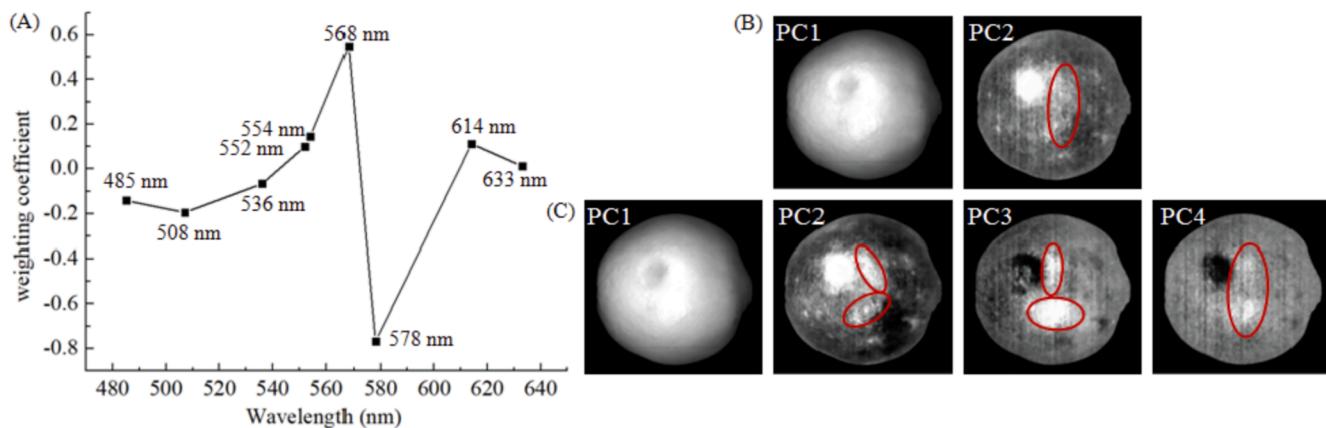


Fig. 7. Weighting coefficient curve of PC5 (A) and the PC images based on two feature wavelength images (568 and 578 nm) (B) and PC images based on four feature wavelength images (508, 568, 578 and 614 nm) (C).

3.5. Multispectral image identification of early rotten oranges

Fig. 8 shows the threshold segmentation algorithm developed based on four feature wavelength images, which was similar to the algorithm developed based on two feature wavelengths. In the multispectral image detection algorithm, four wavelength images at 508, 568, 578 and 614 nm

need to be first obtained, and then PCA was performed to obtain the target image PC4. Meanwhile, the binary mask image (Fig. 8(c)) constructed from single wavelength image at 614 nm was used for background segmentation to obtain the PC4 image as shown in Fig. 8(a). Next, the median filtering was applied to PC4 image for denoising, as shown in Fig. 8(b). Based on the denoised PC4 image, the rotten region

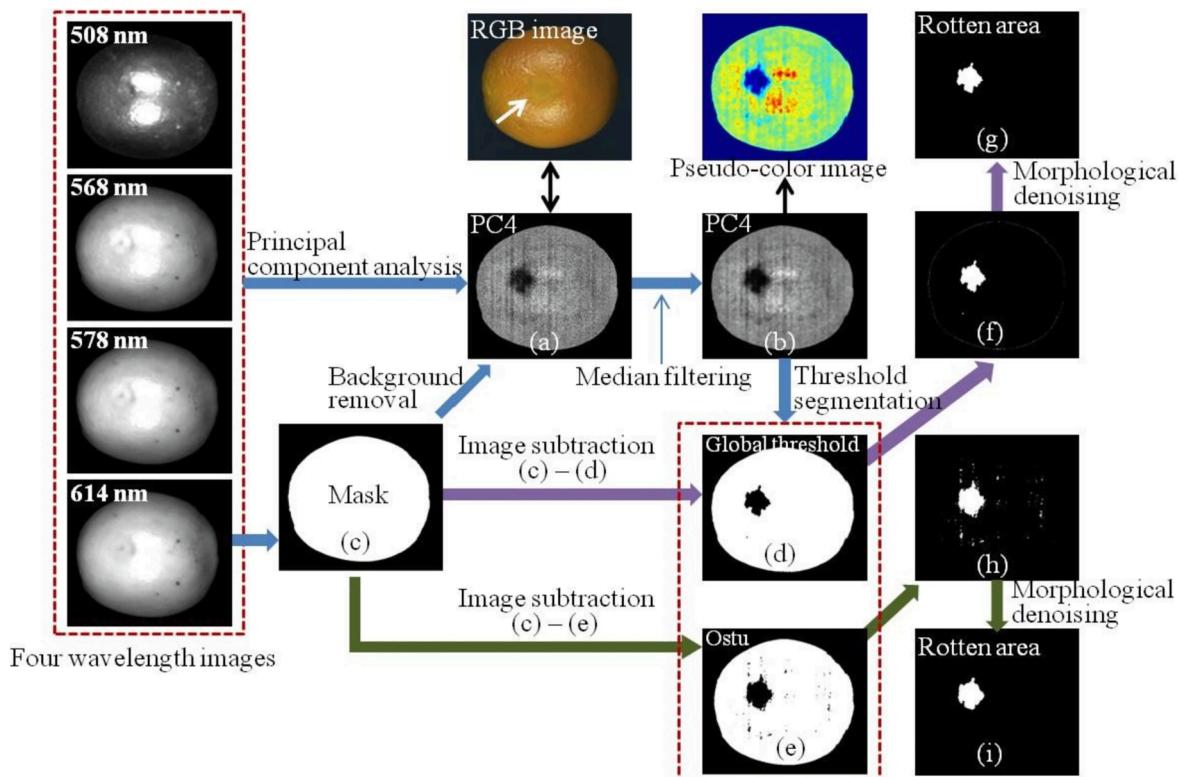


Fig. 8. Multispectral image identification algorithm of early rotten oranges.

was segmented by using the global threshold method or Ostu to obtain the binary image, as shown in Fig. 8(d) and (e), respectively. The binary image was subtracted from the mask image to obtain the final region of interest, as shown in Fig. 8(g) and (h), respectively. The final result image after image morphological denoising was displayed in Fig. 8(f) and (i), respectively. RGB image and pseudo-color image in Fig. 8 were used as reference images. It can be seen that the rotten area that was difficult to be identified based on RGB image was clearly presented in the PC4 image. It can be also clearly seen from the pseudo-color image that the rotten area had an obvious contrast with the sound area, and the bright spots on the fruit surface will not affect the segmentation of the rotten area.

As an example, the segmentation results of the typical orange samples based on PC2 and PC4 images combining with two threshold segmentation algorithms were shown in Fig. 9. RGB image was used as a reference, and the areas marked by green circle in RGB image were rotten areas. The first sample was sound orange and the rest were oranges with early decay. Considering the uncertainty of sample location in the actual online detection, the rotten areas on the samples shown in Fig. 9 may be located in the middle or edge of oranges and some sample images may contain stem-end (such as the sixth sample image). It can be seen from the figure that all rotten areas were very obvious in the PC2 image, but this image was greatly affected by uneven illumination. The most important thing was that bright spots, stem-end and rotten areas presented the same intensity in PC2 image. Although the global

threshold and Ostu methods were able to segment rotten areas, they could not effectively identify sound fruit. Compared with PC4 image, the bright spot and stem-end seem to have no effect on the segmentation of rotten areas, because the bright spot and stem-end showed a white area in the PC4 image, while the rotten area was black. Comparing two segmentation algorithms of global threshold and Ostu, the former obtained 100% recognition rate for all demonstration samples, while the latter had serious over-segmentation and only the second sample was correctly identified. Over-segmentation was mainly due to the uneven illumination on the sample surface, resulting in no effective bimodal histogram in PC4 image to segment the rotten area (Note: Effective bimodal histogram means that one peak represents the decayed tissue and the other peak represents sound tissue). Although the global threshold algorithm obtained good segmentation results, it should still be noted that the uneven illumination had a great impact on the segmentation accuracy of rotten regions.

Table 2 shows the identification results of all samples based on the proposed multispectral image detection algorithm. Similar to the above analysis of Fig. 9, the global threshold and Ostu segmentation method cannot detect sound samples based on PC2 image. Based on PC4 image, the global threshold algorithm achieved the highest recognition accuracy, which was 100% for all sound samples (80 oranges) and over 97.5% for all rotten samples (200 oranges), and the overall classification accuracy reached 98.6%. The results showed that the proposed multispectral image combined with global threshold segmentation algorithm

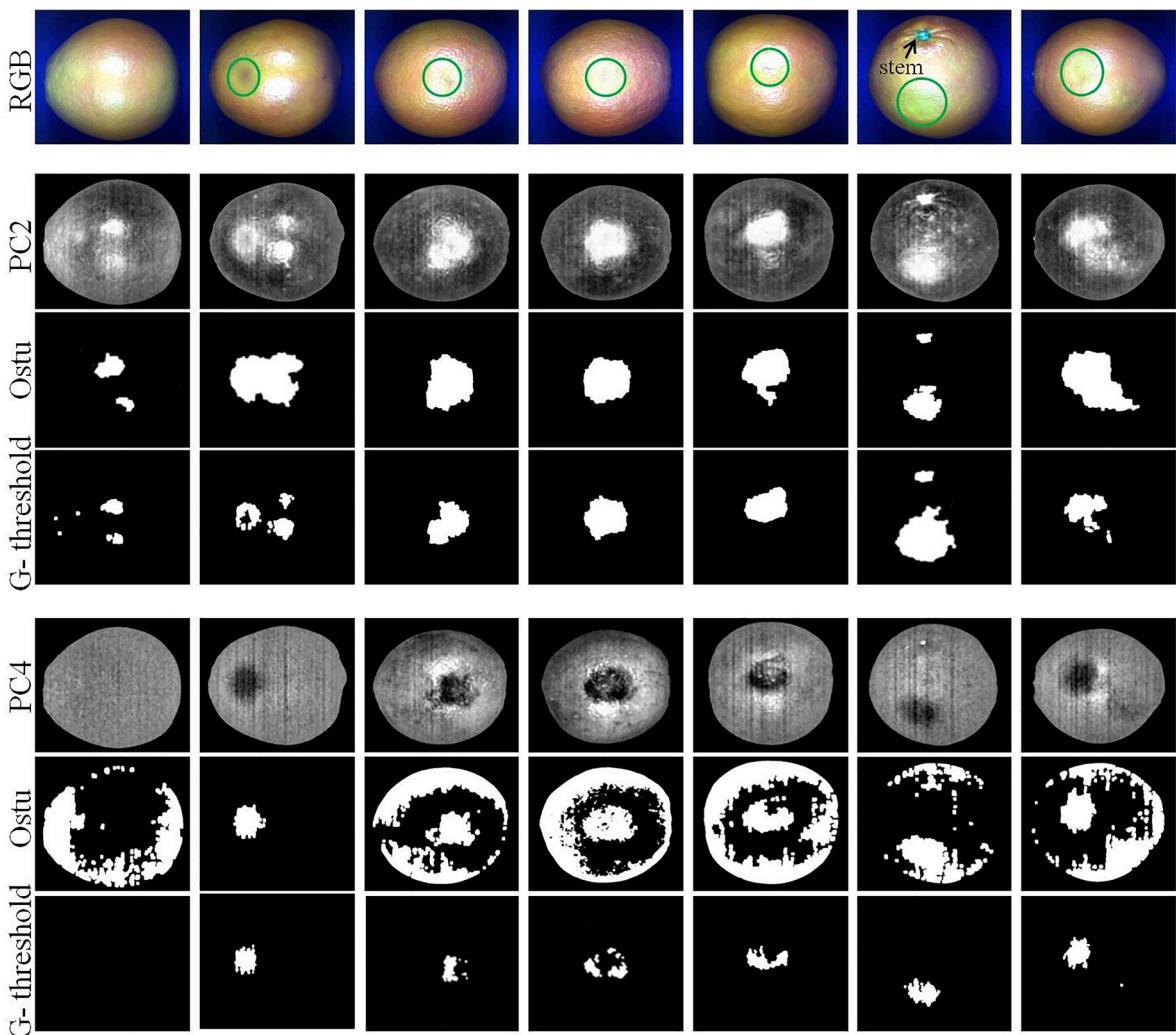


Fig. 9. Segmentation results of the typical orange samples based on different PC images and segmentation algorithms. G-threshold represents the global threshold method.

Table 2

The identification results of all samples based on the proposed multispectral image detection algorithm.

Multispectral PC image	Methods	Sample type	Training set			Testing set		
			Sample number	Misclassified	Accuracy (%)	Sample number	Misclassified	Accuracy (%)
PC2 image	Global threshold	Sound	60	52	13.3	20	18	10.0
		Rotten	160	2	98.8	40	0	100.0
		Total	220	54	75.5	60	38	30.0
	Ostu	Sound	60	50	16.7	20	18	10.0
		Rotten	160	3	98.1	40	1	97.5
		Total	220	53	75.9	60	19	68.3
PC4 image	Global threshold	Sound	60	0	100.0	20	0	100.0
		Rotten	160	3	98.1	40	1	97.5
		Total	220	3	98.6	60	1	98.3
	Ostu	Sound	60	60	0.0	20	20	0.0
		Rotten	160	136	15.0	40	37	7.5
		Total	220	196	10.9	60	57	5.0

Note: PC2 image is obtained by PCA of wavelength images at 568 nm and 578 nm; PC4 image is obtained by PCA of wavelength images at 508, 568, 578 and 614 nm.

can accurately detect the early rotten oranges, and the detection results were independent of the type of fungal infection.

In some similar studies, Gomez-Sanchis et al. [13] used hyperspectral LCTF-based system to classify decay in mandarins caused by *Penicillium digitatum* and *Penicillium italicum*. Ten wavelength images were extracted to build the classification models, and finally, about 93% classification accuracy was obtained. Li et al. [19] determined four wavelength images using hyperspectral imaging combining with PCA to detect decayed oranges and the overall detection accuracy of 98.6% was achieved. However, only *Penicillium digitatum* fungus was considered in their study. Subsequently, Li et al. [17] proposed an algorithm for identification of decayed oranges based on multispectral principal component analysis of seven wavelength images combining with bi-dimensional empirical mode decomposition (BEMD) and watershed segmentation method. Classification accuracy of 97.3% and 100% were obtained for decayed (infected by *Penicillium digitatum* and *Penicillium italicum*) and sound fruit, respectively. However, the used BEMD method and segmentation algorithm were very time-consuming. Zhang et al. [38] used line scanning hyperspectral imaging system to detect the decayed citrus. Two single-wavelength images at 680 nm and 715 nm were proved to be potential for the development of multispectral algorithm. Although only two wavelengths were used, the correct recognition rate of rotten citrus was relatively low (only 90.57%). In comparison, the multispectral image detection algorithm developed in this study was based on only four wavelength images and simple global threshold segmentation theory, which was very important for online citrus quality assessment. Furthermore, the developed algorithm considered two kinds of rotten citrus (caused by *Penicillium digitatum* and *Penicillium italicum*) and achieved up to 98.6% detection accuracy.

4. Conclusions

In this study, it was found that the performance of the PLS-DA model based on the informative variables was better than that of the full-spectrum model for classification of different tissues, especially for the binary classification of sound and rotten tissues. BOSS-PLS-DA and BOSS-SPA-PLS-DA models both achieved 100% classification accuracy by using only 4.67% (33 wavelengths) and 1.27% (9 wavelengths) of the full-spectrum information, respectively. Moreover, the results also showed that the BOSS-PLS-DA model had satisfactory results in the classification of three types of tissues, with an average classification accuracy of 97.1%, which was of great significance for the identification of fungal types in early citrus decay. It was helpful to develop fungi-specific fungicides, so as to save product and potentially reduce the use of chemicals.

In particular, this study proposed a feature wavelength image selection method by coupling hyperspectral data analysis with wavelength

image weighting coefficient analysis. Four images (508, 568, 578 and 614 nm) were finally determined as the feature wavelength images. Moreover, the PC4 image obtained based on PCA of the feature wavelength images can be selected as the target image for defect segmentation of the rotten fruit. In order to achieve fast detection of rotten citrus fruit, the global threshold and Ostu algorithm were respectively used to segment defects. The results showed that the former was not affected by fruit surface bright spots and stem-end, and segmentation result was better. Further research can try to improve the illumination quality of the original image from the two aspects of illumination system optimization and non-uniform illumination correction algorithm development. In this way, fewer wavelength images, such as two images at 568 nm and 578 nm in this study, can be used in the multispectral imaging system, which can further reduce the development cost of system and improve the detection efficiency of rotten fruit. Additionally, we will also try to apply the PLS-DA models directly to the hyperspectral images and analyze the corresponding classification images. This way will be further compared to the PCA-based approach proposed in this study for classification.

Ethical approval

This article has no any study with human participants or animals by any of the authors.

CRediT authorship contribution statement

Wei Luo: Model establishment, Image processing, Original manuscript writing. **Guozhu Fan:** Spectral analysis, Original manuscript writing. **Peng Tian:** Hyperspectral data acquisition. **Wentao Dong:** Sample collection. **Hailiang Zhang:** Supervision, Revision, Editing. **Baishao Zhan:** Methodology, Funding, Supervision, Revision, Editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Key Research and Development Project of Jiangxi Province (20203BBF63031).

References

- [1] M. Barker, W. Rayens, Partial least squares for discrimination, *J. Chemom.* 17 (3) (2003) 166–173.

- [2] C.R. Barmore, G.E. Brown, Role of pectolytic enzymes and galacturonic acid in citrus fruit decay caused by *Penicillium digitatum*, *Phytopathology* 69 (1979) 675–678.
- [3] C.R. Barmore, G.E. Brown, Polygalacturonase from citrus fruit infected with *Penicillium italicum*, *Phytopathology* 71 (1981) 328–331.
- [4] J. Blasco, N. Aleixos, J. Gómez, E. Moltó, Citrus sorting by identification of the most common defects using multispectral computer vision, *J. Food Eng.* 83 (3) (2007) 384–393.
- [5] J. Blasco, N. Aleixos, E. Moltó, Computer vision detection of peel defects in citrus by means of a region oriented segmentation algorithm, *J. Food Eng.* 81 (3) (2007) 535–543.
- [6] A. Bhargava, A. Bansal, Fruits and vegetables quality evaluation using computer vision: a review, *J. King Saud Univ.-Comput. Inform. Sci.* 33 (3) (2021) 243–257.
- [7] R. Calvini, J.M. Amigo, A. Ulrici, Transferring results from NIR-hyperspectral to NIR-multispectral imaging systems: a filter-based simulation applied to the classification of Arabica and Robusta green coffee, *Anal. Chim. Acta* 967 (2017) 33–41.
- [8] B.C. Deng, Y.H. Yun, Y.Z. Liang, Model population analysis in chemometrics, *Chemometr. Intell. Lab. Syst.* 149 (2015) 166–176.
- [9] B.C. Deng, Y.H. Yun, D.S. Cao, Y.L. Yin, W.T. Wang, H.M. Lu, Q.Y. Luo, Y.Z. Liang, A bootstrapping soft shrinkage approach for variable selection in chemical modeling, *Anal. Chim. Acta* 908 (2016) 63–74.
- [10] E.A. Elsherbiny, D.H. Dawood, N.A. Safwat, Antifungal action and induction of resistance by β -aminobutyric acid against *Penicillium digitatum* to control green mold in orange fruit, *Pestic. Biochem. Physiol.* 171 (2021) 104721.
- [11] M. El-Otmani, A. Ait-Oubâou, L. Zacarias, Citrus spp.: Orange, Mandarin, tangerine, clementine, grapefruit, pomelo, lemon and lime, in: E.M. Yahia (Ed.), *Postharvest Biology and Technology of Tropical and Subtropical Fruits*, Woodhead Publishing, 2011, pp. 437–516.
- [12] N.G. Ghooshkhaneh, M.R. Golzarian, M. Mamarabadi, *Spectral Pattern Study of Citrus Black Rot Caused by Alternaria Alternata and Selecting Optimal Wavelengths for Decay Detection*, Food Science & Nutrition, 2022.
- [13] J. Gomez-Sanchis, J. Blasco, E. Soria-Olivas, D. Lorente, P. Escandell-Montero, J. M. Martinez-Martinez, M. Martinez-Sober, N. Aleixos, Hyperspectral LCTF-based system for classification of decay in mandarins caused by *Penicillium digitatum* and *Penicillium italicum* using the most relevant bands and non-linear classifiers, *Postharvest Biol. Technol.* 82 (2013) 76–86.
- [14] W.Q. Huang, J.B. Li, Q.Y. Wang, L.P. Chen, Development of a multispectral imaging system for online detection of bruises on apples, *J. Food Eng.* 146 (2015) 62–71.
- [15] P. Kinay, M.F. Mansour, M.F. Gabler, D.A. Margosan, J.L. Smilanick, Characterization of fungicide-resistant isolates of *Penicillium digitatum* collected in California, *Crop Prot.* 26 (4) (2007) 647–656.
- [16] J.B. Li, W.Q. Huang, A recognition system and method of early rotten citrus fruit based on ring stripe light imaging, China National Invention patent, 2020, Patent No. 201910044898.5.
- [17] J. Li, R. Zhang, J. Li, Z. Wang, H. Zhang, B. Zhan, Y. Jiang, Detection of early decayed oranges based on multispectral principal component image combining both bi-dimensional empirical mode decomposition and watershed segmentation method, *Postharvest Biol. Technol.* 158 (2019) 110986.
- [18] J.B. Li, X.Q. Rao, F.J. Wang, W. Wu, Y.B. Ying, Automatic detection of common surface defects on oranges using combined lighting transform and image ratio methods, *Postharvest Biol. Technol.* 82 (2013) 59–69.
- [19] J.B. Li, W.Q. Huang, X. Tian, C.P. Wang, S.X. Fan, C.J. Zhao, Fast detection and visualization of early decay in citrus using Vis-NIR hyperspectral imaging, *Comput. Electron. Agric.* 127 (2016) 582–592.
- [20] D. Liu, D.W. Sun, X.A. Zeng, Recent advances in wavelength selection techniques for hyperspectral image processing in the food industry, *Food Bioprocess Technol.* 7 (2) (2014) 307–323.
- [21] R. Li, Y. Lu, R. Lu, Structured illumination reflectance imaging for enhanced detection of subsurface tissue bruising in apples, *Trans. ASABE* 61 (3) (2018) 809–819.
- [22] D. Lorente, P. Escandell-Montero, S. Cubero, J. Gomez-Sanchis, J. Blasco, Visible-NIR reflectance spectroscopy and manifold learning methods applied to the detection of fungal infections on citrus fruit, *J. Food Eng.* 163 (2015) 17–24.
- [23] D. Lorente, M. Zude, C. Idler, J. Gomez-Sanchis, J. Blasco, Laser-light backscattering imaging for early decay detection in citrus fruit using both a statistical and a physical model, *J. Food Eng.* 154 (2015) 76–85.
- [24] D. Lorente, J. Blasco, A.J. Serrano, E. Soria-Olivas, N. Aleixos, J. Gómez-Sanchis, Comparison of ROC feature selection method for the detection of decay in citrus fruit using hyperspectral images, *Food Bioprocess Technol.* 6 (12) (2013) 3613–3619.
- [25] W. Luo, H. Zhang, X. Liu, Hyperspectral/multispectral reflectance imaging combining with watershed segmentation algorithm for detection of early bruises on apples with different peel colors, *Food Anal. Methods* 12 (5) (2019) 1218–1228.
- [26] J. Obagwu, L. Korsten, Integrated control of citrus green and blue molds using *bacillus subtilis* in combination with sodium bicarbonate or hot water, *Postharvest Biol. Technol.* 28 (1) (2003) 187–194.
- [27] D. Obenland, D. Margosan, S. Collin, J. Sievert, K. Fjeld, M.L. Arpaia, J. Thompson, D. Slaughter, Peel fluorescence as a means to identify freeze damaged navel oranges, *HortTechnology* 19 (2) (2009) 379–384.
- [28] D. Obenland, D. Margosan, J.L. Smilanick, B. Mackey, Ultraviolet fluorescence to identify navel oranges with poor peel quality and decay, *HortTechnology* 20 (6) (2010) 991–995.
- [29] L. Palou, J.L. Smilanick, C. Montesinos-Herrero, S. Valencia-Chamorro, Novel approaches for postharvest preservation of fresh citrus fruits, in: D.A. Slaker (Ed.), *Citrus Fruits: Properties, Consumption and Nutrition*, Nova Science Publishers, Inc., NY, USA, 2011, pp. 1–45. ISBN: 978-1-61761-189-6.
- [30] K. Papoutsis, M.M. Mathioudakis, J.H. Hasperué, V. Ziogas, Non-chemical treatments for preventing the postharvest fungal rotting of citrus caused by *Penicillium digitatum* (green mold) and *Penicillium italicum* (blue mold), *Trends Food Sci. Technol.* 86 (2019) 479–491.
- [31] R. Qureshi, M. Uzair, K. Khurshid, H. Yan, Hyperspectral document image processing: Applications, challenges and future prospects, *Pattern Recogn.* 90 (2019) 12–22.
- [32] M. Ryckewaert, M. Metz, D. Heran, P. George, B. Grezes-Besset, R. Akbarinia, J. M. Roger, R. Bendoula, Massive spectral data analysis for plant breeding using parSketch-PLSDA method: discrimination of sunflower genotypes, *Biosyst. Eng.* 210 (2021) 69–77.
- [33] D.C. Slaughter, D.M. Obenland, J.F. Thompson, M.L. Arpaia, D.A. Margosan, Non-destructive freeze damage detection in oranges using machine vision and ultraviolet fluorescence, *Postharvest Biol. Technol.* 48 (3) (2008) 341–346.
- [34] W.-H. Su, S. Bakalis, D.-W. Sun, Chemometric determination of time series moisture in both potato and sweet potato tubers during hot air and microwave drying using near/mid-infrared (NIR/MIR) hyperspectral techniques, *Dry. Technol.* 38 (5–6) (2020) 806–823.
- [35] K.Q. Yu, Y.R. Zhao, Z.Y. Liu, X.L. Li, F. Liu, Y. He, Application of visible and near-infrared hyperspectral imaging for detection of defective features in loquat, *Food Bioprocess Technol.* 7 (11) (2014) 3077–3087.
- [36] Y.H. Yun, H.D. Li, B.C. Deng, D.S. Cao, An overview of variable selection methods in multivariate analysis of near-infrared spectra, *TrAC, Trends Anal. Chem.* 113 (2019) 102–115.
- [37] Y.F. Zhang, X.H. Yang, Z.L. Cai, S.X. Fan, H.Y. Zhang, Q. Zhang, J.B. Li, Online detection of watercore apples by Vis/NIR Full-transmittance spectroscopy coupled with ANOVA method, *Foods* 10 (12) (2021) 2983.
- [38] H. Zhang, S. Zhang, W. Dong, W. Luo, Y. Huang, B. Zhan, X. Liu, Detection of common defects on mandarins by using visible and near infrared hyperspectral imaging, *Infrared Phys. Technol.* 108 (2020) 103341.