

Analytical Methods

Prediction of banana color and firmness using a novel wavelengths selection method of hyperspectral imaging

Chuanqi Xie^a, Bingquan Chu^{b,*}, Yong He^{a,*}^a College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China^b School of Biological and Chemical Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China

ARTICLE INFO

Keywords:

Banana
Hyperspectral imaging
Color
Firmness
Two-wavelength combination
Prediction

ABSTRACT

This study investigated the feasibility of using hyperspectral imaging for determining banana color (L^* , a^* and b^*) and firmness as well as classifying ripe and unripe samples. The hyperspectral images at wavelengths 380–1023 nm were acquired. Partial least squares (PLS) models were built to predict color and firmness. Two-wavelength combination method ($\frac{\lambda_i - \lambda_j}{\lambda_i + \lambda_j}$, $\frac{\lambda_i^2 - \lambda_j^2}{\lambda_i^2 + \lambda_j^2}$, $\frac{\lambda_i}{\lambda_j}$ and $\lambda_i - \lambda_j$) was used to identify the effective wavelengths.

Based on the selected wavelengths, PLS models obtained good results with the coefficient of determination in prediction (R_p^2) of 0.795 for L^* , 0.972 for a^* , 0.773 for b^* and 0.760 for firmness. The corresponding residual predictive deviation (RPD) values were 2.234, 6.098, 2.119 and 2.062, respectively. The classification results of ripe and unripe samples were excellent in two different principal components spaces (PC1 + PC2 and PC1 + PC3). It indicated hyperspectral imaging can be used to non-destructively determine banana color and firmness as well as classify ripe and unripe samples.

1. Introduction

Fruits quality is important for determining the harvest time and storage condition (Lleó, Roger, Herrero-Langreo, Diezma-Iglesias, & Barreiro, 2011). Quality properties in fruits change rapidly along with the ripeness (Glew et al., 2005). If the fruits are harvested at a very early stage, the consumption quality will be poor (Lleó et al., 2011), while a late harvest may reduce the shelf life and make the fruits easily susceptible to diseases (Jha, Chopra, & Kingsly, 2007). In order to keep good quality, the fruits should be harvested at a proper time. Thus, it is significant to know the ripeness stages. Firmness has been considered as an important indicator of ripening and one predictor of the potential shelf life (Crisosto, 1999). The detection of firmness includes destructive and non-destructive measurements. Texture analyzer is a destructive method, which cannot be implemented in large-scale detection. The optical technique is a good substitute, which is non-destructive and fast, and can be used in large-scale and online detection. Color is another index for evaluating the degree of ripeness (Jha et al., 2007). It has been assumed that component a^* might be correlated with the chlorophyll content and related with the fruit maturity stage (Zude, 2002). The internal (i.e., firmness) and external characteristics (color components) will change when the fruits become ripe. For example, the color of bananas will change to yellow when they become mature, and

the firmness will decrease as well. All of these features will result in the change of external optical properties. This is why optical technique can be used to detect fruits quality (i.e., color and firmness). However, traditional non-contact optical detection method (spectral technique) can only collect the spectral reflectance information one by one. Multispectral imaging can only provide the images at limited wavebands and no spectral reflectance information.

An advanced optical detection technique (hyperspectral imaging) was used to detect color and firmness as well as classify unripe and ripe samples in this study. This technique can provide the spectral reflectance of any pixel, and more samples can be scanned one time, which is faster and more effective. Compared with multispectral imaging technique, hyperspectral imaging can provide more images at hundreds of wavebands. The hyperspectral image can be created when the sample is scanned by the hyperspectral imaging camera. It is composed of a series of monochromatic images at different wavelengths. Each pixel of the hyperspectral image contains both spectral and spatial information (Xie, Li, Shao, & He, 2014). This technique has been proved to be an excellent method in detecting color of pork (Barbin, ElMary, Sun, & Allen, 2012), beef (Wu, Peng, Li, Chen, & Dhakal, 2012), lamb (Kamruzzaman, ElMasry, Sun, & Allen, 2012), salmon fillets (Wu, Sun, & He, 2012), Turkey ham (Iqbal, Sun, & Allen, 2013) and soybeans (Huang, Wang, Zhang, & Zhu, 2014).

* Corresponding authors.

E-mail addresses: chubingquan10@163.com (B. Chu), yhe@zju.edu.cn (Y. He).

It was also used in firmness determination of bananas (Rajkumar, Wang, ElMasry, Raghvan, & Gariepy, 2012) and blueberries (Leiva-Valenzuela, Lu, & Aguilera, 2013). Because of the excellent performance, hyperspectral imaging was already applied in fruit ripeness detection (Lleó et al., 2011; Wei, Liu, Qiu, Shao, & He, 2014; Yang, Lee, & Gader, 2014). All of these studies showed the feasibility of using hyperspectral imaging for color and firmness determination as well as ripeness degree detection.

The specific objectives of this study were to: (1) determine color and firmness of bananas using spectral reflectance information; (2) select useful wavelengths using a novel wavelengths selection method $\left(\frac{\lambda_i - \lambda_j}{\lambda_i + \lambda_j}, \frac{\lambda_i^2 - \lambda_j^2}{\lambda_i^2 + \lambda_j^2}, \frac{\lambda_i}{\lambda_j} \text{ and } \lambda_i - \lambda_j\right)$; and (3) classify ripe and unripe samples based on spectrum signature.

2. Materials and methods

2.1. Samples

Green bananas (*Musa* spp., AAA group cv. 'Brazil') were harvested in Guangxi province and then transported to the laboratory at Zhejiang University within 3 days (transportation temperature: 13–15°C; humidity: 90–95%). In order to promote the ripening process, some unripe bananas were dipped in a solution for 5 min, which contained 1000 µL/L ethephon (Huayi Group, Shanghai, China). After air-dried for 2 h, the samples were stored at 22 °C in a climate chamber for 6 days for ripening (Wang, Luo, & Du, 2015). Some other samples were dipped in distilled water for 5 min, which only contained purified water (for control group). After air-dried for 2 h, these samples were used as the unripe ones. All samples have the similar size (length: 18–24 cm, diameter: about 3.5 cm) in order to make them representative. Finally, 87 unripe and 81 ripe bananas were used in this study.

2.2. Imaging system and software

The visible/near infrared hyperspectral imaging system, covering the wavelengths of 380–1023 nm, was used to capture hyperspectral images in this study. It consisted of an imaging spectrograph (V10E, Specim, Oulu, Finland), a charge coupled device (CCD) camera (C8484-05, Hamamatsu City, Japan), a lens (OLE-23), two light sources (Oriel Instruments, Irvine, USA) with two 150W quartz tungsten halogen lamps, a moving conveyor and a computer. The area CCD array detector of the camera has 672 × 512 (spatial × spectral) pixels, and the spectral resolution is 2.8 nm. The samples were scanned by the camera line by line with the moving speed of 2.2 mm/s. The camera exposure time was 0.13 s, and the frame rate was 7.0. ENVI 4.7 software (ITT Visual Information Solutions Inc., Boulder, CO, USA) was used for extracting spectral reflectance information from hyperspectral images. IBM SPSS Statistics V22 (IBM, Armonk, NY, USA), Unscrambler V10.1 software (CAMO Software AS, Oslo, Norway) and MATLAB R2014a (The Math Works Inc., Natick, MA, USA) were used to analyze the data, establish prediction models and select effective wavelengths.

2.3. Images calibration

The white and dark images should be collected by the camera firstly. The white image was acquired using a white Teflon board with the spectral reflectance of about 99%. The dark image was obtained by turning off the light and covering camera lens with the cap. Then, each sample was tiled on the conveyor to be scanned by the camera in sequence. Finally, hyperspectral images covering the wavelengths of 380–1023 nm were obtained. For each hyperspectral image, it consisted of 512 gray images at different wavebands. All raw hyperspectral images were calibrated by the dark and white images according to Eq. (1).

$$I_{cal} = \frac{I_{raw} - I_{dark}}{I_{white} - I_{dark}} \quad (1)$$

where I_{cal} is the calibrated hyperspectral image, I_{raw} is the raw hyperspectral image, I_{dark} is the dark image, and I_{white} is the white image.

2.4. Color and firmness measurement

In this study, three color components in $L^*a^*b^*$ space were determined by the colorimeter (Konica Minolta, CR-400, Japan). Before color acquisition, the colorimeter should be calibrated by a standard white plate. For each measurement, the three color components (L^* , a^* and b^*) can be obtained at the same time. L^* represents the brightness component of samples. The higher the L^* value, the brighter the sample. Positive value of a^* demonstrates the sample is redder than standard, and negative means greener than standard. The positive b^* shows the sample is yellower than standard, and negative indicates bluer than standard (Liu, Fu, & Nian, 2014). Firmness was taken using the texture analyzer (TA-XT2i, Stable Microsystems Texture Technologies Inc., UK) fitted with a 5 mm diameter flat probe. The test speed was 1 mm/s, and the penetration depth was 10 mm (Wang et al., 2013). The firmness unit was recorded as N/cm², and samples with higher firmness values mean they are harder. The sampling spot was in the middle of each sample, which was corresponding to the area for spectrum information extraction.

2.5. Models

Partial least squares (PLS) was used to predict the color (L^* , a^* and b^*) and firmness values. This method has been used for prediction in many previous studies (Cen, Bao, He, & Sun, 2007; Zhu, Cheng, Wu, & He, 2011; Zou et al., 2011). It can obtain good performance when the spectral variables and Y variables are linear, which is particularly efficient when the number of X variables is more than the samples (Xie, Li, Nie, & He, 2013). Based on the PLS model, a set of orthogonal factors (latent variables, LV) can be obtained, which contain most of the useful information for prediction (Kamruzzaman et al., 2012). It can also be used for discrimination analysis in the form of partial least squares-discrimination analysis (PLS-DA) (Bylesjö et al., 2006). In this study, PLS-DA was used to classify ripe and unripe bananas. Principal component analysis (PCA) can be used for data compression and effective information selection in hyperspectral imaging analysis (Qin, Burks, Kim, Chao, & Ritenour, 2008). It summarizes data by forming a few new variables (PCs), which can represent the raw data (Kamruzzaman, ElMasry, Sun, & Allen, 2011). Based on these PCs, the useful information in the spectral reflectance data can be acquired.

2.6. Two-wavelength combination

In order to remove redundant information among contiguous wavelengths and simplify the prediction model, two-wavelength combination method was used to select the effective wavelengths. This method aims at identifying the most important wavelengths by seeking the highest correlation value between the two-wavelength combination and Y variables. The combination can be a ratio or difference $\left(\frac{\lambda_i - \lambda_j}{\lambda_i + \lambda_j}, \frac{\lambda_i^2 - \lambda_j^2}{\lambda_i^2 + \lambda_j^2}, \frac{\lambda_i}{\lambda_j} \text{ and } \lambda_i - \lambda_j\right)$ of all possible two wavelengths, where λ_i and λ_j are wavebands at 400–1023 nm. For different types of samples, they have different spectral reflectance at the same waveband. Big difference means that the waveband may play an important role for prediction or classification. Therefore, the ratio or difference of any two wavelengths can be used for effective wavelengths selection. This method has been used in previous studies and obtained good results. Bulanon, Burks, Kim, and Titenour (2013) selected the optimal wavelengths using $\frac{\lambda_i}{\lambda_j}$ and $\lambda_i - \lambda_j$ for citrus black spot disease detection, producing four useful wavelengths (493, 629, 713 and 781 nm). Lee, Kang,

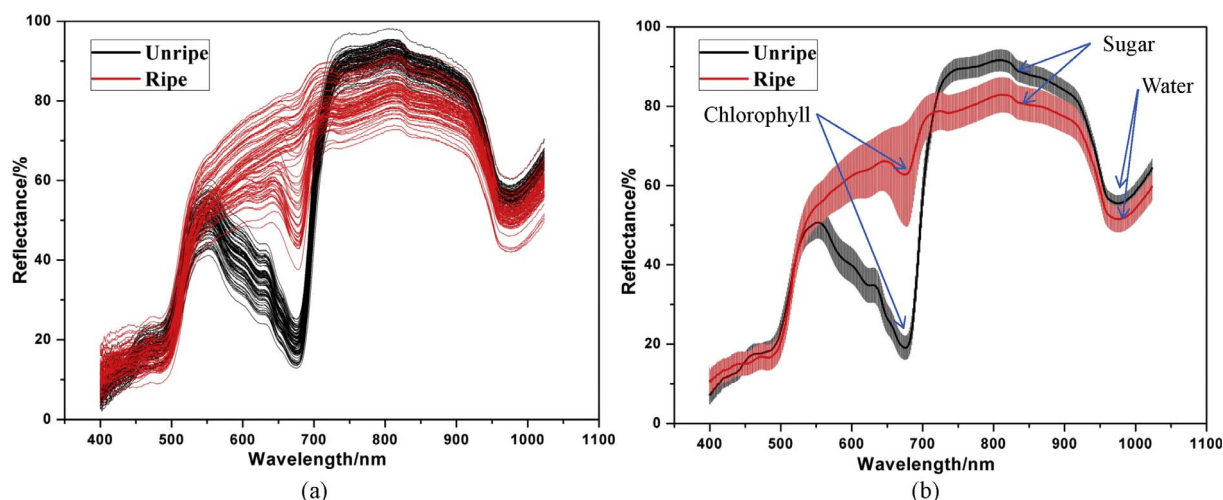


Fig. 1. Spectral reflectance curves: (a) full and (b) average with SD as well as the specific absorbance wavelengths.

Delwiche, Kim, and Noh (2008) used $\frac{\lambda_i}{\lambda_j}$ and $\lambda_i - \lambda_j$ to identify effective wavelengths for detecting defects on apples. For each formulation, only one highest value can be acquired, which is corresponding to two different wavelengths. Consequently, four wavelengths (670, 684, 734 and 828 nm) were finally generated by the two formulations. Only $\frac{\lambda_i}{\lambda_j}$ was investigated to select useful wavelengths for citrus canker lesion detection by Niphadkar, Burks, Qin, and Ritenour (2013), and two wavelengths (729 and 834 nm) were obtained. In our study, four different two-wavelength combinations (Eqs. (2)–(5)) were carried out to identify the most significant wavelengths. Taking $\frac{\lambda_i}{\lambda_j}$ for example, the correlation value between $\frac{\lambda_i}{\lambda_j}$ and Y variables (L^* , a^* , b^* and firmness) can be calculated, and each obtained value is located on either hand of the diagonal of the square plot, in which the horizontal-axis is corresponding to the wavelength i, and the vertical-axis is corresponding to the wavelength j. For this method, the values on the diagonal should be 0, while only one highest value can be obtained for each equation. Finally, full spectral reflectance data changed into a new matrix with the dimension of $m \times n$ (m is the number of the samples, and n is the number of the selected wavelengths). Each equation was calculated by MATLAB R2014a.

$$\text{Equation 1} = \frac{\lambda_i - \lambda_j}{\lambda_i + \lambda_j} \quad (2)$$

$$\text{Equation 2} = \frac{\lambda_i^2 - \lambda_j^2}{\lambda_i^2 + \lambda_j^2} \quad (3)$$

$$\text{Equation 3} = \frac{\lambda_i}{\lambda_j} \quad (4)$$

$$\text{Equation 4} = \lambda_i - \lambda_j \quad (5)$$

where λ_i stands for wavelength i from 400 to 1023 nm, and λ_j represents wavelength j from 400 to 1023 nm.

2.7. Performance evaluation

Each prediction model was evaluated by the values of coefficient of determination in calibration (R_c^2), coefficient of determination in cross-validation (R_{cv}^2), coefficient of determination in prediction (R_p^2), residual predictive deviation (RPD), root mean square error in calibration (RMSEC), root mean square error in cross-validation (RMSECV) and root mean square error in prediction (RMSEP). RPD is defined as standard deviation (SD) of the data divided by RMSECV. It stands for the relative prediction performance more directly than only using either R^2 or RMSE (Kamruzzaman et al., 2012). If the RPD value is less than

1.0, it means the prediction result is very poor. Between 1.0 and 1.4 indicates a poor prediction result. Between 1.4 and 1.8 means a fair prediction result, which may be used for prediction. A RPD between 1.8 and 2.0 signifies a good result and can be used for quantitative prediction. A RPD between 2.0 and 2.5 proves a very good prediction result, and RPD greater than 2.5 demonstrates an excellent result (Rossel, McGlynn, & McBratney, 2006). A good model should have high values of R_c^2 , R_{cv}^2 , R_p^2 and RPD as well as low values of RMSEC, RMSECV and RMSEP (Wu, Chen, Zhu, Guan, & Wu, 2011).

2.8. Experiment procedure

After hyperspectral images collection, color values were obtained using the colorimeter, and firmness values were measured by the texture analyzer. All raw hyperspectral images were calibrated by the white and dark reference images. For each calibrated image, a region of interest (ROI) with 40×40 pixels which was corresponding to the same area of color and firmness measurement was identified. Spectral reflectance values of all pixels at each ROI were extracted and averaged into one value at each waveband from 400 to 1023 nm. Thus, a matrix with $m \times b$ were obtained, in which m is the number of samples and b is the number of the bands. Based on the spectral reflectance data, PLS models were built to predict color and firmness values. Then, significant wavelengths were selected by two-wavelength combination method. PLS models were built again based on the selected wavelengths for the prediction, and PCA as well as PLS-DA were used for differentiating ripe and unripe samples.

3. Results and discussion

3.1. Spectral reflectance analysis

The full and average spectral reflectance curves as well as SD values for both unripe and ripe samples were shown in Fig. 1(a and b), respectively. The large difference of spectral reflectance at 550–950 nm means the spectrum information could be used for classifying the two different types of samples. The wavelength at around 680 nm was the fruit chlorophyll absorption band (Subedi & Walsh, 2011). With unripe bananas becoming ripe, the color will change from green to yellow, and the chlorophyll will decrease as well. Thus, the average spectral reflectance of unripe samples at this wavelength (680 nm) was lower (Lleó et al., 2011). The spectral wavelengths between 550 and 700 nm displayed a strong correlation with the banana peel chlorophyll level, and thus unripe samples which had higher chlorophyll content showed lower spectral reflectance values in this spectral region (Li,

Table 1
Color (L^* , a^* and b^*) and firmness values for unripe, ripe, calibration and prediction sets.

Type	Statistics	L^*	a^*	b^*	Firmness/N
Unripe	Minimum	56.63	−20.11	33.93	4.17
	Maximum	68.82	−15.39	38.42	19.60
	Mean	62.78	−18.20	36.10	11.32
	SD	2.88	1.16	0.92	4.82
Ripe	Minimum	66.15	−5.55	39.20	1.05
	Maximum	74.06	−0.02	54.67	2.80
	Mean	70.62	−2.52	46.73	1.59
	SD	1.90	1.22	4.00	0.45
Calibration	Minimum	56.63	−20.11	33.93	1.05
	Maximum	74.06	−0.02	54.67	19.60
	Mean	66.61	−10.67	41.17	6.58
	SD	4.63	7.96	6.03	5.97
Prediction	Minimum	58.02	−20.01	34.26	1.10
	Maximum	73.64	−0.29	54.07	18.68
	Mean	66.47	−10.58	41.33	6.72
	SD	4.66	7.99	6.11	6.06

SD Standard Deviation.

Slaughter, & Thompson, 1997). The wavelength at 840 nm was the sugar absorption band (ElMasry, Wang, ElSayed, & Ngadi, 2007), and ripe samples showed lower reflectance at 840 nm meaning they contained more sugar. Ripe bananas have more water content (Egbebi & Bademosi, 2011), thus the spectral reflectance at 970 nm (water absorption band) were lower for ripe samples (Xie et al., 2013).

The details of the specific absorbance wavebands can also be seen in Fig. 1(b). The difference for the two curves might be caused by many factors, such as the color, firmness, water content and sugar content, etc. This is why spectral reflectance information could be used for color and firmness prediction. Due to the noise at the beginning of the wavelengths, only 400–1023 nm was studied.

3.2. Color and firmness values

The minimum, maximum, mean and SD values of L^* , a^* , b^* and firmness were shown in Table 1. It can be found the mean value of L^* for ripe bananas is higher than unripe samples. Unripe bananas are greener, thus a^* values are lower. Ripe bananas are yellower, resulting in higher values of b^* . The firmness values of unripe samples were between 4.17 and 19.60 with the average of 11.32, while the values were between 1.05 and 2.80 with the average of 1.59 for ripe samples. Unripe samples have higher firmness values, which is because they are harder than ripe samples. To avoid bias in subset partition (Li, Xie, He, Qiu, & Zhang, 2012), all samples were arranged in an ascending order in accordance with the Y variables (L^* , a^* , b^* and firmness, respectively), and then they were divided into the calibration and prediction sets at a ratio of 2:1. One sample was picked out from every three ones consecutively, resulting in 112 samples in the calibration set and 56 ones in the prediction set. The minimum, maximum, mean and SD values of each parameter for both calibration and prediction sets were also shown in Table 1. It must be noted that no one sample was used in both sets, which meant each sample was either in calibration or

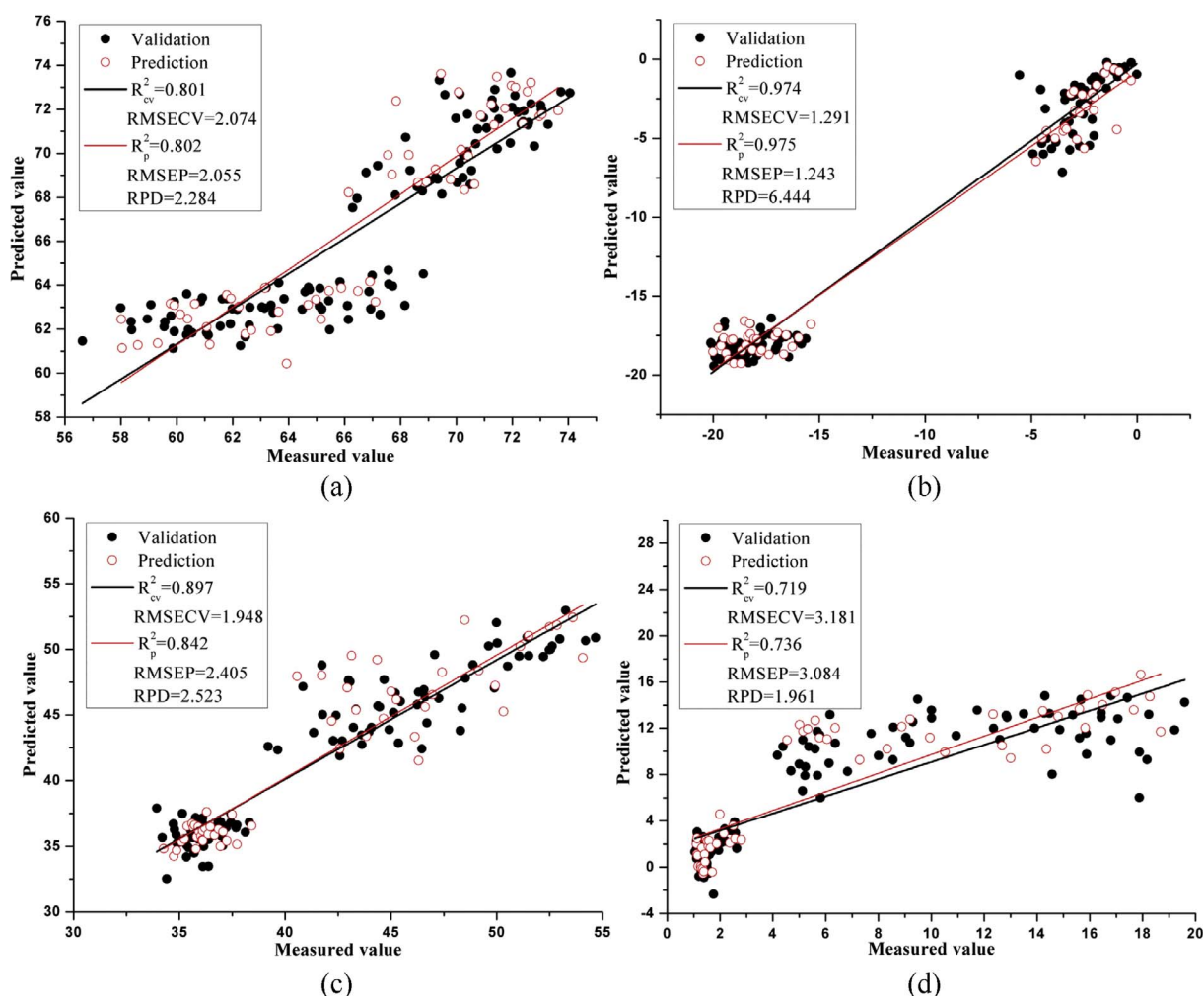


Fig. 2. Cross-validation and prediction results for (a) L^* , (b) a^* , (c) b^* and (d) firmness of PLS models using full wavelengths.

prediction set. A wide range of the four features can be seen in both calibration and prediction sets, indicating that the samples in both sets can represent the range of all possible values, thus was helpful to build an accurate and robust model (Wu, Chen, et al., 2012). In this study, those samples used for calibration set were also used as the full cross-validation set. Full cross-validation was applied in testing the robustness of the calibration set. If there are large differences between calibration and cross-validation results, the models cannot be trusted.

3.3. Prediction results using full wavelengths

Four different partial least squares (PLS) prediction models were established using the full wavelengths (400–1023 nm) for L^* , a^* , b^* and firmness, respectively. The values of R_{cv}^2 , R_p^2 , RPD , $RMSECV$ and $RMSEP$ for each model were shown in Fig. 2(a–d). The R_p^2 of L^* , a^* , b^* and firmness were 0.802, 0.975, 0.842 and 0.736, and the corresponding $RMSEP$ were 2.055, 1.243, 2.405 and 3.084. From the results, it can be found each model performed good with high values of R_{cv}^2 , R_p^2 and RPD as well as low values of $RMSECV$ and $RMSEP$. The samples in both cross-validation and prediction sets distributed near the ideal lines, demonstrating that the obtained values were trustable. The cross-validation results were used to test the robustness of the models. The results obtained from calibration (R_c^2) and cross-validation (R_{cv}^2) sets

were very similar (L^* : 0.805/0.801, a^* : 0.978/0.974, b^* : 0.919/0.897 and firmness: 0.759/0.719), indicating the models were robust and trusted. The best prediction model was a^* -based PLS model, which produced the highest values of R_{cv}^2 (0.974), R_p^2 (0.975) and RPD (6.444). Most of the RPD values for L^* , a^* , b^* and firmness (2.284, 6.444, 2.523 and 1.961) were higher than 2.0, demonstrating the prediction results were good or excellent. The overall results proved that hyperspectral imaging was trusted and objective for color and firmness detection.

3.4. Effective wavelengths

Fig. 3(a–d) showed the obtained results of component L^* . Each small figure (a–d) were corresponding to the results of $\frac{\lambda_i - \lambda_j}{\lambda_i + \lambda_j}$, $\frac{\lambda_i^2 - \lambda_j^2}{\lambda_i^2 + \lambda_j^2}$, $\frac{\lambda_i}{\lambda_j}$ and $\lambda_i - \lambda_j$, respectively. This method aimed at looking for the highest correlation value between the two-wavelength combination and Y variables (L^*). In each contour plot, both horizontal-axis and vertical-axis were corresponding to the wavebands from 400 to 1023 nm. Each point represented one value, and different colors stood for the corresponding correlation values. The brighter point meant a higher correlation between the two-wavelength combination and the Y variable. Similarly, the darker point indicated the correlation was lower. For each two-

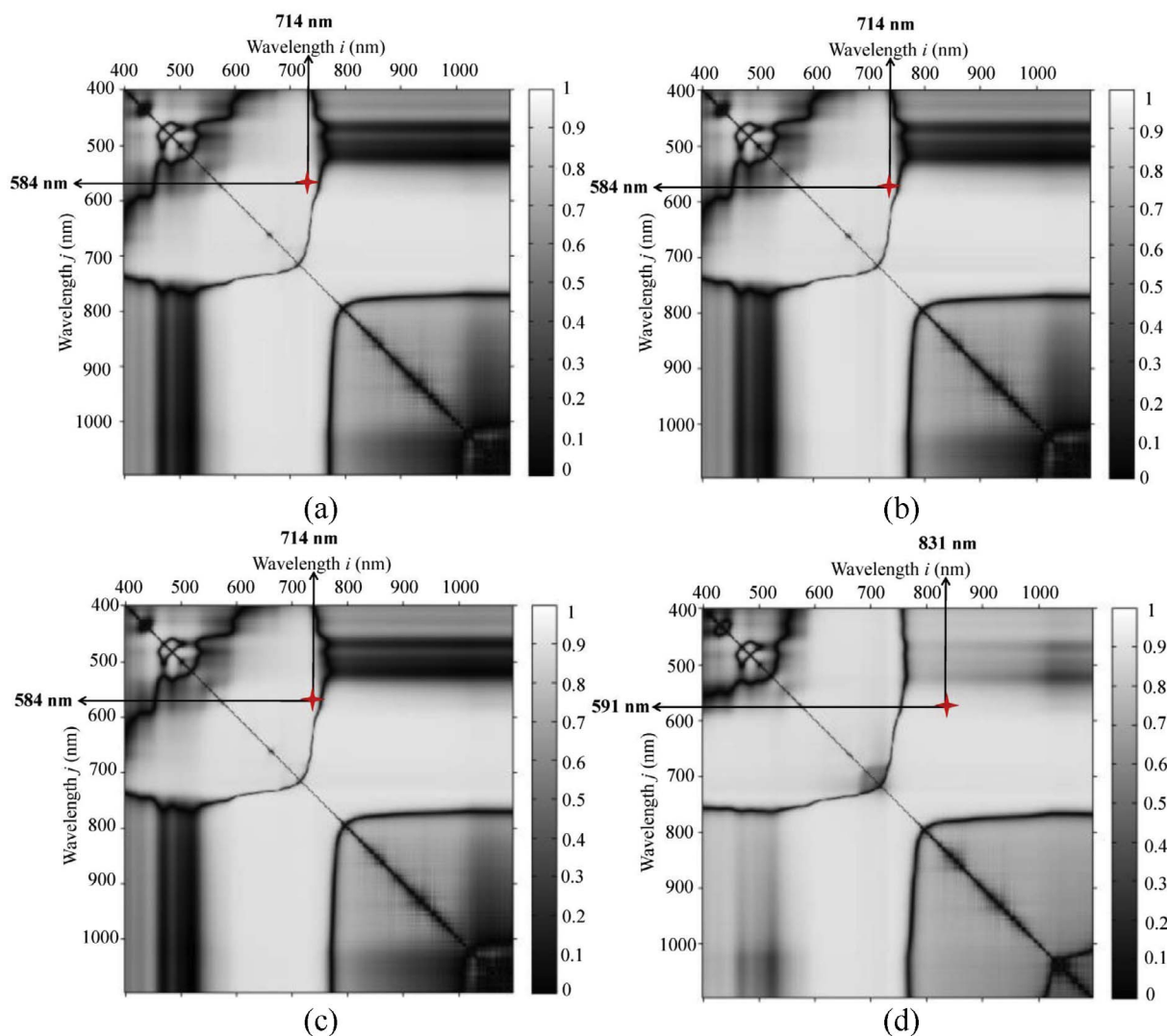


Fig. 3. Contour plots for effective wavelengths selection of L^* : (a) $\frac{\lambda_i - \lambda_j}{\lambda_i + \lambda_j}$, (b) $\frac{\lambda_i^2 - \lambda_j^2}{\lambda_i^2 + \lambda_j^2}$, (c) $\frac{\lambda_i}{\lambda_j}$ and (d) $\lambda_i - \lambda_j$.

Table 2
Correlation analysis and ANOVA for selected wavebands.

Components	Wavelength/nm	Analysis of correlation		ANOVA	
		Correlation	Significance	F value	Significance
L*	584	0.869	0.000	14.529	0.001
	591	0.879	0.000	18.257	0.001
	714	0.307	0.000	2.522	0.120
	831	-0.601	0.000	6.274	0.013
a*	534	0.209	0.007	1.613	0.205
	540	0.318	0.000	1.601	0.209
	568	0.738	0.000	2.749	0.040
	595	0.906	0.000	6.076	0.002
	614	0.932	0.000	7.299	0.001
	647	0.961	0.000	11.260	0.000
	649	0.961	0.000	11.426	0.000
b*	507	-0.466	0.000	1.295	0.257
	519	-0.195	0.011	1.185	0.343
	525	-0.066	0.399	1.191	0.338
	532	0.047	0.541	1.233	0.303
	563	0.521	0.000	1.945	0.043
	566	0.602	0.000	2.304	0.016
	575	0.720	0.000	3.717	0.001
	618	0.830	0.000	10.175	0.000
firmness	638	-0.757	0.000	/	/
	675	-0.738	0.000	/	/
	681	-0.731	0.000	/	/
	696	-0.700	0.000	/	/

p = .01.

wavelength combination, two wavelengths can be acquired. Taking the first equation $\left(\frac{\lambda_i - \lambda_j}{\lambda_i + \lambda_j}\right)$ of L* for example, the highest correlation coefficient (the brightest point) occurred at 584 and 714 nm (Fig. 3a), and the correlation between $\frac{\lambda_{714 \text{ nm}} - \lambda_{584 \text{ nm}}}{\lambda_{714 \text{ nm}} + \lambda_{584 \text{ nm}}}$ and L* was 0.9082 (the highest). The asterisk point indicated the highest value, and the two arrow lines pointed to the two different wavelengths. Thus, eight wavelengths can be acquired for each Y variable. Considering there were some repeated wavelengths, four wavelengths for L* (584, 591, 714 and 831 nm) were obtained finally. For other three components, seven wavelengths for a* (534, 540, 568, 595, 614, 647 and 649 nm), eight for b* (507, 519, 525, 532, 563, 566, 575 and 618 nm) and four for firmness (638, 675, 681 and 696 nm) were identified by this method. The numbers of the selected wavelengths only accounted for 0.81%, 1.41%, 1.62% and 0.81% of the full spectral wavebands (400–1023 nm), respectively. The new wavelengths were then used to replace the full wavebands for the prediction of L*, a*, b* and firmness. Most of the selected wavelengths for a* and b* were located from 550 to 700 nm. This is because the wavelengths at this range have a strong correlation with the chlorophyll level, which is related to the two color components (a* and b*). All of the four selected wavelengths for firmness were located from 550 to 700 nm. This is because firmness is closely related to the maturity stage, and fruit maturing process is usually accompanied with the change of chlorophyll content. Among the four wavelengths, the wavelengths at 675 and 681 nm are the chlorophyll absorption bands. The selected wavelengths had the potential to be used for designing a multispectral detection sensor, which can be used in on-line detection.

The Pearson correlation and significance as well as F value and significance were used to evaluate each selected wavelength, and the corresponding results can be seen in Table 2. Most of the selected wavelengths had good correlation with the corresponding Y variables. For L*, the waveband 591 nm had the highest correlation (0.879), and the corresponding F value (18.257) was also the highest, while waveband 714 nm had the lowest correlation (0.307), and the corresponding F value (2.522) was also the lowest. The Analysis of variance (ANOVA) significance values for the two wavelengths (591 and 714 nm) were 0.001 and 0.120, which proved they had good performance. For a*, the

most significant wavelength was 649 nm, which had the highest correlation (0.961) and F value (11.426), and 534 and 540 nm performed the worst among the seven wavelengths due to the low correlation and F values as well as high values of significance. The wavelength 618 nm performed the best, and another three wavelengths (519, 525 and 532 nm) produced worse results for b*. For firmness, each wavelength obtained a good correlation result. Since SPSS did not produce ANOVA results for the data of firmness, we could only talk about the four wavelengths based on the analysis of correlation.

3.5. Performance of selected wavelengths

3.5.1. Prediction results

PLS models were re-established using the selected wavelengths, and the corresponding prediction results can be seen in Fig. 4(a–d). The samples in both sets (cross-validation and prediction) also distributed ideally near the lines. Each selected wavelengths-based PLS model also obtained a good result with high values of R_{cv}^2 , R_p^2 , RPD as well as low values of $RMSECV$ and $RMSEP$. The R_p^2 of L*, a*, b* and firmness were 0.795, 0.972, 0.773 and 0.760, and the corresponding $RMSEP$ values were 2.090, 1.318, 2.885 and 2.941. For firmness prediction, the R_{cv}^2 , R_p^2 and RPD values were even higher than those in full wavelengths-based PLS model. This might be because full wavelengths contained too much redundant information, and most of the useful information was kept after wavelengths selection. For L*, a* and b*, there were a little decrement of R_p^2 values. However, the input variables decreased largely, which simplified the model and had the potential to be used in online detection. The RPD values for L* (2.234), a* (6.098), b* (2.119) and firmness (2.062) were higher than 2.0, indicating the prediction results were still good or excellent. It can be concluded that the selected wavelengths can be equal to or even more efficient than the whole wavelengths, which is corresponding to a previous study (Kamruzzaman et al., 2011). The overall results proved again the spectral reflectance information extracted from hyperspectral images can be applied in determining color and firmness values, and two-wavelength combination method was very efficient for selecting significant wavelengths.

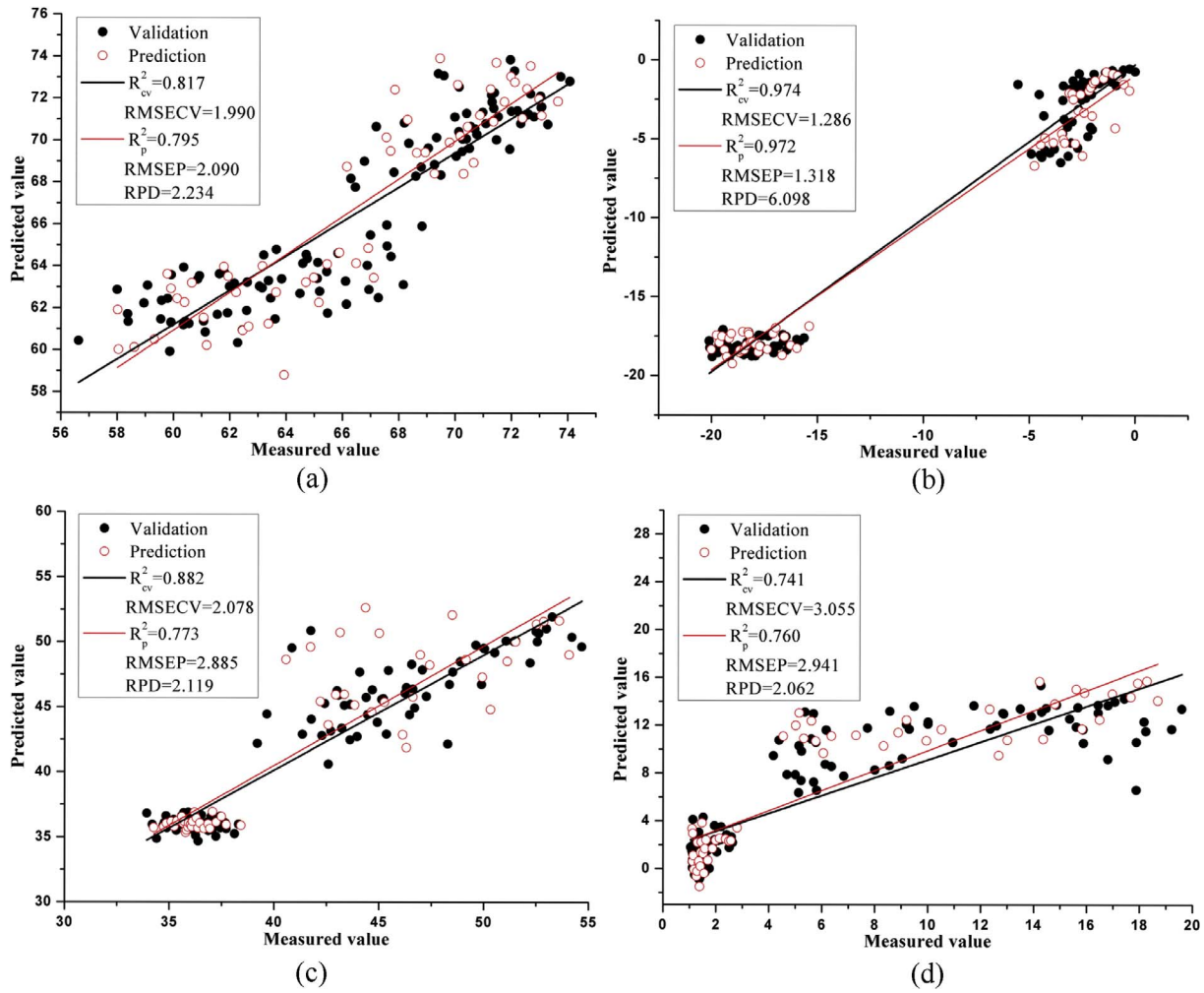


Fig. 4. Cross-validation and prediction results for (a) L^* , (b) a^* , (c) b^* and (d) firmness of PLS models using selected wavelengths.

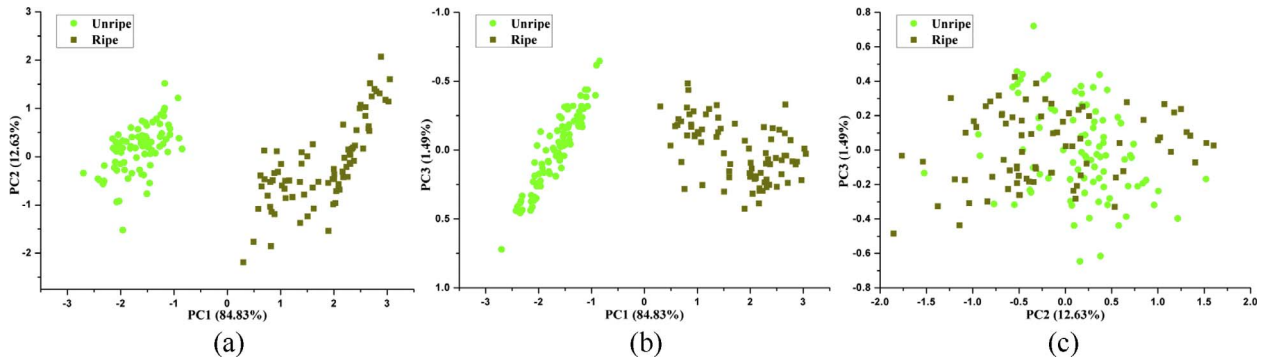


Fig. 5. Scores plots of unripe and ripe samples in different PC spaces: (a) PC1 + PC2, (b) PC1 + PC3 and (c) PC2 + PC3.

3.5.2. Linear regression equations

The linear equations between spectral reflectance and Y values were shown as follows. Based on the equations, different features (L^* , a^* , b^* and firmness) can be calculated only by the spectral reflectance values at the selected wavelengths. Eqs. (6)–(9) are corresponding to L^* , a^* , b^* and firmness, respectively. Also, the multispectral detection sensors have the potential to be designed based on the following equations.

$$L^* = 16.8011\lambda_{584 \text{ nm}} + 19.8003\lambda_{591 \text{ nm}} - 4.2278\lambda_{714 \text{ nm}} - 14.9409\lambda_{831 \text{ nm}} + 64.1245 \quad (6)$$

$$a^* = -18.5618\lambda_{534 \text{ nm}} - 17.6008\lambda_{540 \text{ nm}} - 10.2875\lambda_{568 \text{ nm}} + 1.1919\lambda_{595 \text{ nm}} + 6.3315\lambda_{614 \text{ nm}} + 19.0733\lambda_{647 \text{ nm}} + 19.9707\lambda_{649 \text{ nm}} - 8.9725 \quad (7)$$

$$b^* = -53.9909\lambda_{507 \text{ nm}} - 37.3977\lambda_{519 \text{ nm}} - 14.3529\lambda_{525 \text{ nm}} + 2.91\lambda_{532 \text{ nm}} + 27.2881\lambda_{563 \text{ nm}} + 28.5041\lambda_{566 \text{ nm}} + 23.161\lambda_{575 \text{ nm}} + 4.6412\lambda_{618 \text{ nm}} + 32.0407 \quad (8)$$

$$\text{firmness} = -47.613\lambda_{638 \text{ nm}} - 598.4697\lambda_{675 \text{ nm}} + 586.5086\lambda_{681 \text{ nm}} + 49.0891\lambda_{696 \text{ nm}} + 0.0442 \quad (9)$$

where $\lambda_{i \text{ nm}}$ is the spectral reflectance value at wavelength i .

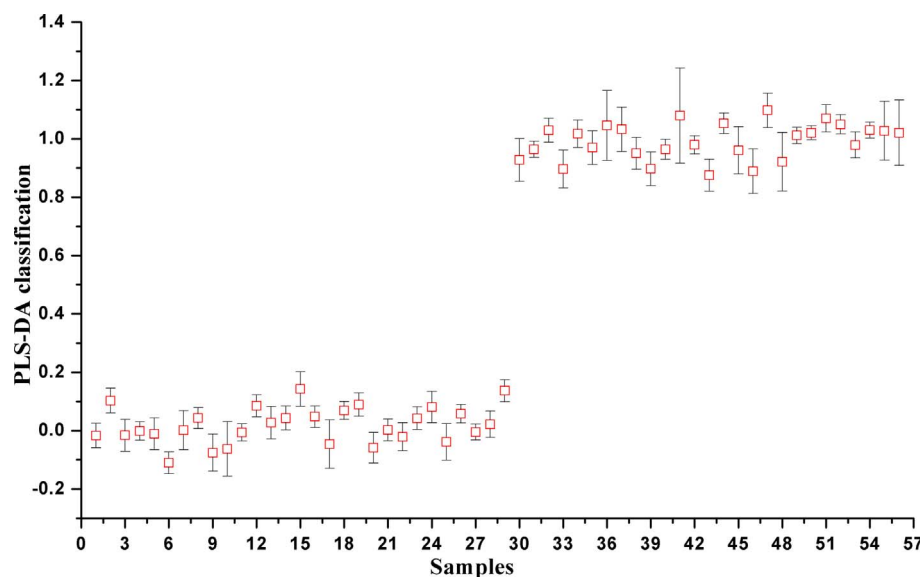


Fig. 6. PLS-DA classification results for the prediction set.

3.6. Classification results

This study was also carried out to classify unripe and ripe bananas using the spectral reflectance information. Fig. 5(a–c) showed the scores and loading plots of different combinations for the first three PCs (PC1 + PC2, PC1 + PC3 and PC2 + PC3). The first three PCs explained 98.95% of the variance (PC1: 84.83%, PC2: 12.63% and PC3: 1.49%), indicating these PCs could represent most of the useful information in the raw data. In Fig. 5(a–c), x-axis and y-axis were corresponding to different PCs (PC1, PC2 and PC3). In Fig. 5(a and b), it can be found unripe bananas were in the negative side of PC1, while ripe were in the positive side of PC1. There was no overlap for the two types of samples, indicating that PC1 + PC2 and PC1 + PC3 can produce excellent classification results. In Fig. 5(c), most of the samples were mixed, which might be because the total variance of PC2 and PC3 were only 14.12% and could not represent the raw wavebands. PLS-DA was also built to classify the two types of samples, and the corresponding results were shown in the plot of predicted values with estimated uncertainties (Fig. 6). The samples 1–29 had predicted values close to 0 and were classified to class “unripe”. The samples 30–56 with predicted values around 1 were assigned to class “ripe”. It also obtained a good classification result.

4. Conclusions

The obtained results demonstrated that spectral reflectance information extracted from hyperspectral images can be applied in the determination of color (L^* , a^* and b^*) and firmness as well as classification of ripe and unripe bananas. Two-wavelength combination was very efficient for selecting useful wavelengths. Both full and selected wavelengths-based PLS models obtained good prediction results. Based on the four different two-wavelength combinations, no more than eight wavelengths can be acquired for each Y variable. The numbers of the selected wavelengths for L^* , a^* , b^* and firmness only accounted for 0.81%, 1.41%, 1.62% and 0.81% of the whole wavebands. Based on the selected wavelengths, the R_p^2 of L^* , a^* , b^* and firmness were 0.795, 0.972, 0.773 and 0.760, and the corresponding RPD were 2.234, 6.098, 2.119 and 2.062. For firmness, the selected wavelengths even obtained a better result than full wavelengths. This might be because the redundant information was deleted by wavelengths selection. Therefore, a polychromatic on-line detection sensor can be designed using the selected wavelengths. Other non-contact optical techniques, such as spectral technique, can only collect the spectral reflectance data for one

sample one time, while hyperspectral imaging can scan more samples. Compared with multispectral imaging technique, hyperspectral imaging can provide more images at hundreds of wavelengths. To date, this is the first time to determine banana color and firmness using two-wavelength combination method of hyperspectral imaging.

Acknowledgments

This work was supported by National Key Foundation for Exploring Scientific Instrument of China (2014YQ47037703) and the National Science and Technology Support Program of China (2015BAD19B03).

Author contributions

The first two authors contributed equally to this work and shared the first authorship.

Additional information

The authors declare no conflict of interest.

References

- Barbin, D. F., ElMary, G., Sun, D. W., & Allen, P. (2012). Predicting quality and sensory attributes of pork using near-infrared hyperspectral imaging. *Analytica Chimica Acta*, 719, 30–42.
- Bulanon, D. M., Burks, T. F., Kim, D. G., & Titenour, M. A. (2013). Citrus black spot detection using hyperspectral image analysis. *Agricultural Engineering International: CIGR Journal*, 15(3), 171–180.
- Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J. K., Holmes, E., & Trygg, J. (2006). OPLS discriminant analysis: Combining the strengths of PLS-DA and SIMCA classification. *Journal of Chemometrics*, 20, 341–351.
- Cen, H. Y., Bao, Y. D., He, Y., & Sun, D. W. (2007). Visible and near infrared spectroscopy for rapid detection of citric and tartaric acids in orange juice. *Journal of Food Engineering*, 82(2), 253–260.
- Crisosto, C. H. (1999). Optimum procedures for ripening stone fruit. *Management of Fruit Ripening. University of California, Davis, Postharvest Horticulture Series*, 9, 28–30.
- Egbebi, A. Q., & Bademosi, T. A. (2011). Chemical compositions of ripe and unripe banana and plantain. *International Journal of Tropical Medicine and Public Health*, 1(1), 1–5.
- ElMasry, G., Wang, N., ElSayed, A., & Ngadi, M. (2007). Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *Journal of Food Engineering*, 81(1), 98–107.
- Glew, R. H., Ayaz, F. A., Millson, M., Huang, H. S., Chang, L. T., Sanz, C., et al. (2005). Changes in sugars, acids and fatty acids in naturally parthenocarpic date plum persimmon (*Diospyros lotus* L.) fruit during maturation and ripening. *European Food Research and Technology*, 221(1–2), 113–118.
- Huang, M., Wang, Q. G., Zhang, M., & Zhu, Q. B. (2014). Prediction of color and moisture content for vegetable soybean during drying using hyperspectral imaging technology. *Journal of Food Engineering*, 128, 24–30.

- Iqbal, A., Sun, D. W., & Allen, P. (2013). Predicting of moisture, color and pH in cooked, pre-sliced turkey hams by NIR hyperspectral imaging system. *Journal of Food Engineering*, 117(1), 42–51.
- Jha, S. N., Chopra, S., & Kingsly, A. R. P. (2007). Modeling of color values for non-destructive evaluation of maturity of mango. *Journal of Food Engineering*, 78(1), 22–26.
- Kamruzzaman, M., ElMasry, G., Sun, D. W., & Allen, P. (2011). Application of NIR hyperspectral imaging for discrimination of lamb muscles. *Journal of Food Engineering*, 104(3), 332–340.
- Kamruzzaman, M., ElMasry, G., Sun, D. W., & Allen, P. (2012). Prediction of some quality attributes of lamb meat using near-infrared hyperspectral imaging and multivariate analysis. *Analytica Chimica Acta*, 714, 57–67.
- Lee, K. J., Kang, S. W., Delwiche, S. R., Kim, M. S., & Noh, S. (2008). Correlation analysis of hyperspectral imagery for multispectral wavelength selection for detection of defects on apples. *Sensing and Instrumentation for Food Quality and Safety*, 2(2), 90–96.
- Leiva-Valenzuela, G. A., Lu, R. F., & Aguilera, J. M. (2013). Prediction of firmness and soluble solids content of blueberries using hyperspectral reflectance imaging. *Journal of Food Engineering*, 115(1), 91–98.
- Li, M., Slaughter, D. C., & Thompson, J. F. (1997). Optical chlorophyll sensing system for banana ripening. *Postharvest Biology and Technology*, 12(3), 273–283.
- Li, X. L., Xie, C. Q., He, Y., Qiu, Z. J., & Zhang, Y. C. (2012). Characterizing the moisture content of tea with diffuse reflectance spectroscopy using wavelet transform and multivariate analysis. *Sensors*, 12(7), 9847–9861.
- Liu, S. B., Fu, Y. Q., & Nian, S. (2014). Buffering colour fluctuation of purple sweet potato anthocyanins to acidity variation by surfactants. *Food Chemistry*, 162, 16–21.
- Lleó, L., Roger, J. M., Herrero-Langreo, A., Diezma-Iglesias, B., & Barreiro, P. (2011). Comparison of multispectral indexes extracted from hyperspectral images for the assessment of fruit ripening. *Journal of Food Engineering*, 104(4), 612–620.
- Niphadkar, N. P., Burks, T. F., Qin, J. W., & Ritenour, M. A. (2013). Estimation of citrus canker lesion size using hyperspectral reflectance imaging. *International Journal of Agricultural and Biological Engineering*, 22(3), 41–51.
- Qin, J. W., Burks, T. F., Kim, M. S., Chao, K. L., & Ritenour, M. A. (2008). Citrus canker detection using hyperspectral reflectance imaging and PCA-based image classification method. *Sensing and Instrumentation for Food Quality and Safety*, 2(3), 168–177.
- Rajkumar, P., Wang, N., ElMasry, G., Raghvan, G. S. V., & Gariepy, Y. (2012). Studies on banana quality and maturity stages using hyperspectral imaging. *Journal of Food Engineering*, 108(1), 194–200.
- Rossel, R. A. V., McGlynn, R. N., & McBratney, A. B. (2006). Determining the composition of mineral-organic mixes using UV-vis-NIR diffuse reflectance spectroscopy. *Geoderma*, 137(1), 70–82.
- Subedi, P. P., & Walsh, K. B. (2011). Assessment of sugar and starch in intact banana and mango fruit by SWNIR spectroscopy. *Postharvest Biology and Technology*, 62(3), 238–245.
- Wang, Y. S., Luo, Z. S., & Du, R. X. (2015). Nitric oxide delays chlorophyll degradation and enhances antioxidant activity in banana fruits after cold storage. *Acta Physiologiae Plantarum*, 37(4), 1–10.
- Wang, Y. S., Luo, Z. S., Du, R. X., Liu, Y., Ying, T. J., & Mao, L. C. (2013). Effect of nitric oxide on antioxidative response and proline metabolism in banana during cold storage. *Journal of Agricultural and Food Chemistry*, 61(37), 8880–8887.
- Wei, X., Liu, F., Qiu, Z. J., Shao, Y. N., & He, Y. (2014). Ripeness classification of as-tringent persimmon using hyperspectral imaging technique. *Food and Bioprocess Technology*, 7(5), 1371–1380.
- Wu, D., Chen, J. Y., Lu, B. Y., Xiong, L. N., He, Y., & Zhang, Y. (2012). Application of near infrared spectroscopy for the rapid determination of antioxidant activity of bamboo leaf extract. *Food Chemistry*, 135(4), 2147–2156.
- Wu, D., Chen, X. J., Zhu, X. G., Guan, X. C., & Wu, G. C. (2011). Uninformative variable elimination for improvement of successive projections algorithm on spectral multivariate selection with different calibration algorithms for the rapid and non-destructive determination of protein content in dried laver. *Analytical Methods*, 3(8), 1790–1796.
- Wu, J. H., Peng, Y. K., Li, Y. Y., Chen, J. J., & Dhakal, S. (2012). Prediction of beef quality attributes using VIS/NIR hyperspectral scattering imaging technique. *Journal of Food Engineering*, 109(2), 267–273.
- Wu, D., Sun, D. W., & He, Y. (2012). Application of long-wave near infrared hyperspectral imaging for measurement of color distribution in salmon fillet. *Innovative Food Science & Emerging Technologies*, 16, 361–372.
- Xie, C. Q., Li, X. L., Nie, P. C., & He, Y. (2013). Application of time series hyperspectral imaging (TS-HSI) for determining water content within tea leaves during drying. *Transactions of the ASABE*, 56(6), 1431–1440.
- Xie, C. Q., Li, X. L., Shao, Y. N., & He, Y. (2014). Color measurement of tea leaves at different drying periods using hyperspectral imaging technique. *PLoS One*, 9(12), e113422.
- Yang, C., Lee, W. S., & Gader, P. (2014). Hyperspectral band selection for detecting different blueberry fruit maturity stages. *Computers and Electronics in Agriculture*, 109, 23–31.
- Zhu, F. L., Cheng, S. X., Wu, D., & He, Y. (2011). Rapid discrimination of fish feeds brands based on visible and short-wave near-infrared spectroscopy. *Food and Bioprocess Technology*, 4(4), 597–602.
- Zou, X. B., Shi, J. Y., Hao, L. M., Zhao, J. W., Mao, H. P., Chen, Z. W., et al. (2011). In vivo noninvasive detection of chlorophyll distribution in cucumber (*Cucumis sativus*) leaves by indices based on hyperspectral imaging. *Analytica Chimica Acta*, 706(1), 105–112.
- Zude, M. (2002). Non-destructive prediction of banana fruit quality using VIS/NIR spectroscopy. *Fruits*, 58(3), 135–142.