Study on Test method of Kiwifruit Hardness Based on Hyperspectral Technique

Xinyuan Chen¹, Lina Zheng² and Zhiliang Kang*²

¹ College of economics, Sichuan Agriculture University, Chengdu, Sichuan Province, 625014, China
² College of mechanical and electrical engineering, Sichuan Agriculture University, Ya’an, Sichuan Province, 625014, China
*Corresponding author’s e-mail: kangzhiliang96@163.com

Abstract. Flesh hardness is an important index to measure the quality and storage character of fresh fruit. At present, destructive test method is usually used to test kiwifruit hardness by sampling, while a new method based on hyperspectral technique is proposed in this study. Firstly, hyperspectral images of kiwifruit samples are collected in the spectral range of [400nm, 1000nm], then the hyperspectral images are pre-processed by standard normal transformation, and three sub-ranges of the best combination are selected through synergy interval Partial Least Square (siPLS). Secondly, kernel principal component analysis is adopted for dimension reduction of spectral bands of the best combination sub-intervals, and the first two principal components are input into the trained partial least square (PLS) as characteristic spectral bands. Finally, the test results are compared with physicochemical measurement values. The experimental results show that the Cross-validation Root Mean Square Errors (RMSEC) and the correlation coefficients of the training set and the testing set obtained by the proposed method are 0.2698/0.9315, and 0.3573/0.8738 respectively, indicating that it can effectively test kiwifruit hardness.

1. Introduction

The hardness of kiwifruit directly affects its taste and price. At present, the hardness of kiwifruit is usually measured by a fruit hardness tester or the farmers' pinches, which is subjective or destructive test. Hyperspectral technique is an image technique based on multi-narrow bands¹⁻², which studies the spectral characteristics of various substances by presenting images with the electromagnetic spectrum³. The solubility of pectin and cellulose in kiwifruit affects the hardness of kiwifruit. These substances can be absorbed by hyperspectrum. Therefore, the hardness of kiwifruit can be detected according to the different spectral sensitive groups of each kiwifruit’s hardness, and then the farmers can be guided to determine the picking time and storage time.

In this study, a nondestructive testing method of kiwifruit hardness based on hyperspectral images is designed, which includes the following steps: (1) collecting kiwifruit’s hyperspectral images in the spectrum range of [400nm, 1000nm]; (2) extracting the range of interest, pre-processing the spectral information, and selecting three sub-intervals of the best combination through synergy interval partial least square (siPLS) and obtaining 32 spectral bands totally; (3) reducing dimension for 32 spectral bands through kernel principal component analysis (KPCA)⁴; (4) inputting the characteristic spectra of the first two principal components into a trained partial least square (PLS) and comparing the
results with physicochemical values (that is, actual values), which validates the effectiveness of the proposed method.

2. Test materials
A total of 130 green-core kiwifruits with similar size and non-destructive surface are selected in Yucheng District, Ya'an City, Sichuan Province as experimental samples, and numbered in order. At room temperature of about 20°C in April, 2019, the hyperspectral images of the samples are collected and the hardness is measured by physicochemical tests.

"Gaia Sorter" hyperspectral sorter is adopted. The sorter mainly comprises a uniform source of light, a spectral camera, an electrically-controlled mobile platform and a computer. Wherein, the uniform source of light has four LSTS-200 bromine tungsten lamps, which are arranged in a trapezoidal structure. The spectral resolution of the spectral camera is 2.8 nm. The electrically-controlled mobile platform is used to place the sample to be tested, and the computer is used to control the whole system. After collecting the hyperspectral images of the kiwifruits and removing the peel, the GY-4 fruit hardness tester with 8mm in diameter is used for physicochemical measurement. The kiwifruits are placed directly under the hardness tester head. The force of the hardness tester head changed with the depth of the test head inserted into a kiwifruit. In this process, the maximum force is recorded, representing the physicochemical values of kiwifruit hardness [5].

A total of 90 kiwifruits are selected as the training set and the remaining 40 kiwifruits are used as the testing set. Table 1 lists the physicochemical values of hardness of the training set and testing set.

| Sample Set   | Sample Number | Minimum | Maximum | Mean    | Standard Deviation |
|--------------|---------------|---------|---------|---------|--------------------|
| Training set | 90            | 6.17    | 15.7    | 9.4656  | 1.7553             |
| Testing set  | 40            | 7.67    | 11.63   | 9.6085  | 0.9387             |

3. Test method
Focus the hyperspectral imaging system and set the acquisition parameters. The exposure time of the spectral camera is 13s, and the scanning distance of the electrically-controlled mobile platform is 25cm. The forward and backward velocities of the motor controlling the motion of the electrically-controlled mobile platform are 0.4cm/s and 1cm/s, respectively. The spectral resolution is 2.8nm, and the spectral band range corresponding to the hyperspectral images is 400nm–1000nm where are 256 spectral bands totally.

3.1. Hyperspectral image acquisition and pre-processing
Hyperspectral image of a kiwifruit sample is shown in Figure 1. Certain region of kiwifruit is selected as the region of interest (ROI) after black-and-white correction of the images (see Figure 2), and the hyperspectral information curve of the ROI is drawn as the original spectral curve of the hardness test of the kiwifruit.

![Figure 1. Hyperspectral image of a kiwifruit.](image1)

![Figure 2. Image of RIO.](image2)

The original spectral curves of 130 kiwifruit samples are very similar, each of which has greater noise at the head and the back. The spectral values increase significantly from 700nm spectrum band to 883nm spectrum band. The peak appears near the 883nm, which is related to the spectral absorption...
bands of the cellulose. And there is a significant trough of wave near the 950nm. This indicates that the spectral sensitive groups contain O-H and C-H chemical bonds, which can provide the basis for measuring kiwifruit hardness by hyperspectral technique. The noise spectral bands at the head and the back are removed, and the hyperspectral information in the band range from 450nm to 1000nm is retained, thus a total of 217 spectral bands are obtained. The corresponding hyperspectral curves are shown in Figure 3 (a). Figure 3 (a) is pre-processed through standard normal transformation to avoid the effect of surface scattering. The spectral curves after preprocessing are shown in Figure 3 (b).

3.2. Extraction of characteristic spectral bands

3.2.1. Interval partial least square. The whole spectral bands of hyperspectral images of kiwifruit are divided into n subintervals of equal width by interval partial least square (iPLS), and the hyperspectral information of sub-intervals is correlated with the physicochemical measurement values of kiwifruit hardness, thus an iPLS model for measuring kiwifruit hardness is established. The measurement effect of each regression model is evaluated by Cross-validation Root Mean Square Error (RMSEC), and the sub-interval corresponding to the minimum value of RMSEC is selected as the characteristic spectral band extracted through the iPLS model [6-7].

A total of 217 spectral bands in Figure 3 (b) are divided into 20 sub-intervals of equal width, and the corresponding iPLS models are established for the 20 sub-intervals. The optimal principal components and the corresponding RMSEC extracted through the iPLS models of the intervals are shown in Table 2. In Table 2, Interval 0 represents the 217 spectral bands in Figure 3 (b), namely the full spectral bands. The RMSEC of the iPLS model established on the basis of the first 6 principal components extracted from the 19th sub-interval (949.36nm–973.32nm) is the minimum value of 0.794, which is better than spectral bands of other sub-intervals and the full spectral bands. From the experimental results in Table 2, it can be seen that the RMSECs of the iPLS models in sub-intervals are large, which indicates that the test accuracy of the iPLS models with a single sub-interval is low.
### Table 2. Experimental results of iPLS regression models in the sub-intervals.

| Interval number | The spectral bands /nm | Principal components | RMSEC  |
|-----------------|------------------------|----------------------|--------|
| 0               | 450.00-1000.00         | 3                    | 0.8773 |
| 1               | 450.00-471.40          | 4                    | 0.8690 |
| 2               | 476.26-498.19          | 3                    | 0.8628 |
| 3               | 503.07-525.12          | 4                    | 0.8582 |
| 4               | 530.03-552.20          | 4                    | 0.9072 |
| 5               | 557.14-579.42          | 3                    | 0.9044 |
| 6               | 584.39-606.79          | 3                    | 0.9243 |
| 7               | 611.79-634.31          | 3                    | 0.9286 |
| 8               | 639.33-661.98          | 1                    | 0.9701 |
| 9               | 667.03-689.79          | 1                    | 0.9752 |
| 10              | 694.96-717.75          | 1                    | 0.9588 |
| 11              | 722.85-745.86          | 1                    | 0.9613 |
| 12              | 750.98-774.11          | 1                    | 0.9736 |
| 13              | 779.26-802.51          | 2                    | 0.9317 |
| 14              | 807.69-831.06          | 4                    | 0.9076 |
| 15              | 836.26-859.75          | 2                    | 0.9411 |
| 16              | 864.98-888.59          | 3                    | 0.9370 |
| 17              | 893.85-917.58          | 4                    | 0.8565 |
| 18              | 922.86-946.71          | 6                    | 0.9043 |
| 19              | 949.36-973.32          | 6                    | 0.794  |
| 20              | 975.99-1000.00         | 3                    | 0.8647 |

3.2.2. Extraction through synergy interval partial least square. siPLS is an extension of iPLS, which combines multiple sub-intervals to detect the hardness of Kiwifruit and selects the combination intervals corresponding to the minimum RMSEC as the best combination intervals \(^8\), namely the characteristic spectral band extracted through the siPLS model.

The 20 sub-intervals in Figure 3 (b) are verified by 5-fold cross-validation, and 2, 3 and 4 sub-intervals are combined respectively. By analysis and calculation, the siPLS model is favorable for extracting characteristic spectral bands when combined with 3 sub-intervals. The combination intervals corresponding to siPLS models and their RMSEC are shown in Table 3.

### Table 3. Experimental results of siPLS models of the combination intervals

| Principal components | The combination intervals | RMSEC | Principal components | The combination intervals | RMSEC |
|----------------------|---------------------------|-------|----------------------|---------------------------|-------|
| 6                    | [16 17 19]                | 0.6134| 7                    | [13 17 19]                | 0.6833|
| 6                    | [2 17 19]                 | 0.6687| 7                    | [10 17 19]                | 0.6834|
| 7                    | [3 17 19]                 | 0.6692| 7                    | [12 17 19]                | 0.6969|
| 8                    | [6 17 19]                 | 0.6767| 6                    | [16 17 18]                | 0.7077|
| 8                    | [7 17 19]                 | 0.6818| 7                    | [16 18 20]                | 0.7104|

As shown in Table 3, the test effect of siPLS models of combination intervals is better than that of iPLS model of single sub-interval in measuring the hardness of kiwifruit. When the combination intervals are [16 17 19], the first six principal components are extracted by siPLS model, and the corresponding RMSEC is the minimum value of 0.6130. The spectral bands corresponding to the combination intervals [16 17 19] are 797.34nm–823.26nm, 825.85nm–851.91nm and 880.71nm–904.38nm, where are 32 spectral bands totally.

The 32 spectral bands extracted through siPLS models are reduced in dimension by KPCA, and the first two principal components are extracted. Wherein, the contribution rate of the first principal
component is 82.8131%, and the contribution rate of the second principal component is 13.0634%. Therefore, the first two principal components can represent 217 spectral bands in Figure 3 (C) and can be input into the Partial least square (PLS) test model for later experiment.

4. Partial least square test Model
PLS is a many-to-many linear regression method, which is widely used in spectral analysis. The PLS test model can be expressed as follows:

\[ Y = Xb + e \]

Where, \( Y \) is the hardness value of kiwifruit, \( X \) is the hyperspectral information of kiwifruit, \( b \) is the regression coefficient, and \( e \) is the test error.

RMSEC and correlation coefficient of calibration (\( R_C \)) of training set, as well as root mean squared error of prediction (RMSEP) and correlation coefficient of prediction (\( R_P \)) of testing set, are selected as the evaluation indexes of the PLS test model. Wherein, RMSEC and RMSEP respectively reflect the deviation between the test values and physicochemical values of the training set and the testing set. And the smaller the value, the smaller the deviation degree is. \( R_C \) and \( R_P \) respectively reflect the fitting degree between the test values and the physicochemical values of the training set and the testing set. And the closer to 1 the fitting degree, the higher the accuracy of the test is.

PLS also uses the first six principal components for modeling. Firstly, the hyperspectral information and the corresponding hardness physicochemical values of the training set are used to train the PLS test model. Then the PLS test model is used to measure the hardness of the testing set. The results are evaluated by the evaluation indexes, and the experimental results are shown in Table 4.

| Test method       | RMSEC  | \( R_C \) | RMSEP  | \( R_P \) |
|-------------------|--------|-----------|--------|-----------|
| iPLS+PLS          | 0.7940 | 0.5849    | 0.8475 | 0.4921    |
| siPLS+ PLS        | 0.6130 | 0.7736    | 0.6853 | 0.7068    |
| siPLS+KPCA+PLS    | 0.2698 | 0.9315    | 0.3573 | 0.8738    |

As seen from Table 4, when kiwifruit hardness is tested by siPLS + KPCA + PLS, then \( R_C \), RMSEC, \( R_P \), and RMSEP are 0.9315, 0.2689, 0.8738 and 0.3573, respectively. The test results are better than those of the other methods. The test results of siPLS + PLS are better than those of iPLS + PLS, which indicates that the combination intervals own more information about kiwifruit hardness than that of a single interval. The test accuracy of siPLS+ KPCA + PLS is higher than that of siPLS+ PLS, which indicates that the redundancy of characteristic spectral bands will decrease the test accuracy of PLS test model.

5. Conclusions
In this study, the test method of kiwifruit hardness based on hyperspectral images is studied. The hyperspectral images of kiwifruit in the spectral band range from 400nm to 1000nm are pre-processed, and the characteristic spectral bands extracted through siPLS+ KPCA are inputted to PLS test model. The test results are compared with those obtained through iPLS and siPLS. The test results show that siPLS+ KPCA + PLS has higher test accuracy than iPLS+ PLS and siPLS+ PLS, indicating that the hyperspectral information of combination intervals is more suitable for testing kiwifruit hardness than that of a single sub-interval. Meanwhile, the redundancy of characteristic spectral bands reduces the test accuracy of PLS test model. The proposed method provides a theoretical basis for the application of hyperspectral technique to the non-destructive test of kiwifruit hardness.

Acknowledgments
Authors wishing to acknowledge assistance or encouragement from colleagues, special work by technical staff or financial support from organizations should do so in an unnumbered Acknowledgments section immediately following the last numbered section of the paper.
References

[1] Fan C. (2014) A hyperspectral band selection method based on information criterion. Electronic Design Engineering, 22: 149–152.

[2] Chu M.H. (2007) Study of hyperspectral feature about remote sensing image. Shandong University of Science and Technology, Qingdao.

[3] Tang J.L, Huang C.H., LenJ.H., et al. (2009) Effects of different harvest dates on fruit quality of Jinkui Kiwifruit. South China Fruits, 41: 110–113.

[4] Liu W.J., Li W.J., Li H.G., et al. (2017) Study of maize haploid identification based on oil content detection with near infrared spectroscopy. Spectroscopy and Spectral, 37: 1089–1094.

[5] Feng X.Y. (2011) NY/T 2009-2011 Determination of Fruit Hardness. China Agricultural Press, Beijing.

[6] Xi R.Y., Huang F.R., Li Y.P., et al. (2016) Quantitative analysis of deep-frying oil adulterated virgin olive oil using vis-NIR spectroscopy with iPLS. Spectroscopy and Spectral, 36: 2462–2467.

[7] Zhao F., Peng Y.K. (2017) Measurement of iodine value of pork’s subcutaneous adipose tissue by interval partial least square and raman spectroscopy. Chinese Journal of Lasers, 44: 243–249.

[8] Peng H.G., Peng Y.F., Zhan Y, et al. (2014) Determination of the sugar content of jujube in south xinjiang by near infrared spectroscopy combined with siPLS methods. Food Science and Technology, 39: 276–280.