Advanced Science and Technology Letters
Vol.111 (NGCIT 2015), pp.42-48
http://dx.doi.org/10.14257/astl.2015.111.09

Hyperspectral Detection Method for Starch Content of Potato

Wei Jiang[12], Shuwen Wang[1],Yongcun Fan[1], Junlong Fang[1]*

[1]Electrical and Information Institute, Northeast Agricultural University , 59 Mucai Street, Harbin, 150030,China

[2]Department of Computing ,Harbin Finance University, 65 Diantan Street, Harbin, 150030,China
Email:jwhancg@126.com

Abstract. The purpose of this research is to explore to detect and predict the starch content of potato by the hyperspectral imaging technique. The original images of 118 samples were acquired by the hyperspectral imaging system, and then the suitable pretreatment was selected. The prediction model for starch content of potato was constructed by the 100 hyperspectral images randomly selected which was dealt with smoothing, PCA and PLS. The determination coefficient of this model was 0.8234, and the root mean square error of 0.5633. The other 18 sample images were used to validate the accuracy of model, and the determination coefficient R2 was 0.9031 and the root mean square error (RMSEP) was 0.5025. The results shown that it is feasible to detect and predict the starch content of potato by hyperspectral image technology.

Keywords: Hyperspectral imaging, Potato, Starch, Partial least square

1 Introduction

Potato has high economic value and is one of the most grain crops in the world. The level of potato starch content is one of the main standard to measure the quality of potato [1]. It is important significance for potato breeding and deep processing to accurately and rapidly determine the content of potato starch. At present, it is wide to determine starch content by the traditional methods, but it is difficult to promote in the analysis and test for the large qualities of samples because of the long test time, high cost, not easy to master and inaccuracy result of determination [2].

Hyperspectral imaging technology is the combination of image technology and spectrum technology, It not only can be used to characterize image features of measured sample in the spatial distribution, but also to obtain its spectral properties from a pixel or groups of pixels [3]. Hyperspectral imaging technology has been gradually applied to the quality detection field of fruit and vegetable due to the advantage acquiring spatial spectrum and image information of fruits and vegetables [4-5]. It is applied at the detection of internal quality for the moisture content of potatoes, total sugar, moisture content and hardness of bananas, sugar content and moisture content of Xuehua pear [6]. In this study, potato was used as experimental

ISSN: 2287-1233 ASTL

samples and hyperspectral imaging technology was utilized to obtain spectral curve information acquisition of potato, and the spectral information were extracted by the principal component analysis(PCA) method and then hyperspectral mathematical model about starch content was constructed according to PLS method, and finally this model was validated by a validation set. This model can provide efficient and accurate analysis for detection of potato starch.

2 Material and Method

2.1 Sample and Equipment

Potato was bought from farm produce market in Harbin. There are no surface defects to be used as experimental samples and the samples were cleaned by water for further experiment. The images of 118 samples were acquired, and the 100 images were randomly selected to construct model, and the other were employed to predict model.

Measurement of starch content: homogenate precipitation method [7]. The statistical distribution of potato starch obtained by the homogenate precipitation method is listed in Table 1.

Table. 1. Distribution statistics of starch content of potato

Sample classification	Number	Min (%)	Max (%)	Average(%)
Calibration set	100	11.84	18.06	14.89
Validation set	18	12.16	18.15	15.18

2.2 Hyperspectral image acquisition

The image were acquired by hyperspectral image acquisition system made in HeadWall Company, USA, shown as Fig.1. The system consists of three parts, image acquisition unit, a light source and a sample conveying platform. The image acquisition unit includes an image spectrometer, CCD camera and lens. The light source is a fiber halogen lamp with 150W adjustable power. The slit width of hyperspectral image spectral is 25μm and the spectral range is from 400 to 1000nm. The spectral resolution is 1.29nm, and spacing of image acquisition band is 3nm, and the spatial resolution is 0.15mm.

2.3 Hyperspectral image correction

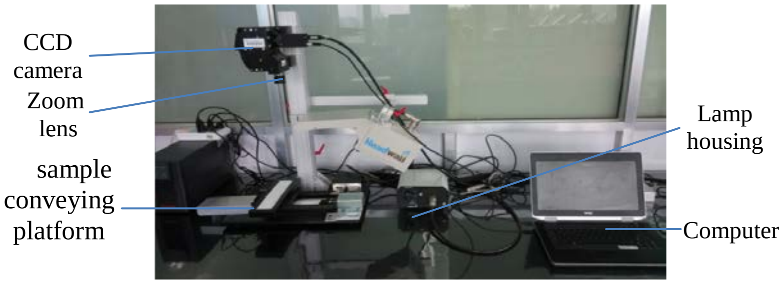

Fig.1. Hyperspectral image system

In order to eliminate the influence of the part of the noise, it must be to correct the hyperspectral images before acquiring the images. The hyperspectral image corrected can be obtained by the eqn.(1) [8-9].

$$R = \frac{R_s - R_d}{R_w - R_d} \tag{1}$$

Which, R is a calibrated image; Rs is the original image;Rw is the white calibration image; and Rd is the full black calibration image.The original spectral image is shown as Fig.2.

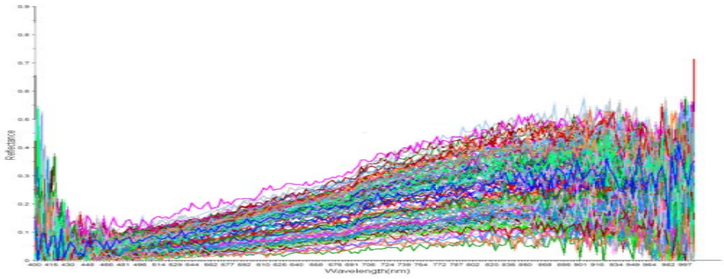

Fig.2. Original spectrum

2.4 Data preprocessing

In order to weaken or eliminate the effect of baseline drift, scattering and other kinds of non-objective factors on the spectra, it is necessary to pretreat the spectra acquired hyperspectral imaging spectrometer [10].The original images were pretreated by moving smooth, the second derivative and multiplicative scatter correction (MSC) method.

2.5 Principal component analysis

Principal component analysis(PCA) was used to reduce the dimensionality of hyperspectral data. Principal of component analysis principle is as follows [17]:

A matrix $X_{n\times p}$ is a spectral matrix of samples, *n* is the number of samples, *p* is the number of variables (band). x_i (*i*=1,2,..., n) express the sample *i*, X_k (*k*=1,2,... , *p*) express the band *k*. *V* is the covariance matrix of sample.

Eigenvalue $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$ is obtained by calculating the characteristic matrix V, the corresponding unit eigenvector is $u_1, u_2, \ldots, u_p$, and the principal component of sample *k* is shown as eqn. (2):

$$y_k = u_k{}^I X, k = 1,2,\ldots, p \tag{2}$$

Which, y_1 is the first principal component, and y_2 is the second principal component, … , by analogy.

For PCA, the proportion of the principal components *k* (y_k) to the total variance is shown as eqn. (3):

$$\lambda_k \Big/ \sum_{i=1}^{p} \lambda_i \quad (k = 1,2,\ldots, p) \tag{3}$$

Eqn. (3) is called the contribution rate of principal component y_k. The contribution rate of first principal component y_1 is maximum, and it is indicated that y_1 has the strongest ability to explains the original variable X_k (*k*=1,2,… , *p*), while the explain ability of y_2, ..., y_p decrease successively.

2.6 Partial Least Squares

Partial least squares regression (PLSR) is the most commonly modeling for the spectral analysis[12] and it can solve the problem which sample number is smaller than variable number. The principle is [13]:

First of all, the matrix X and Y are decomposed at the same time, following as eqn. (4) and (5):

$$X = T \bullet P + E \tag{4}$$

$$Y = U \bullet Q + F \tag{5}$$

Which, T and U are the score of matrix X and Y, respectively. P and Q are load matrix X and Y. E and F are the error matrix.

Then, T and U are established a linear relationship.

$$U = T \bullet B \tag{6}$$

Where B is the regression coefficient matrix.

Finally, the score T_{un} of an unknown sample X_{un} can be obtain by calibration set matrix P, and then Y_{un} can be calculated during the forecast process.

$$Y_{un} = T_{un} BQ + F \tag{7}$$

3 Result and Analysis

3.1 Selection of pretreatment method

In this test, the band range of the hyperspectral image for potato is from 400 to 1000nm, and too large data could result in reducing the processing speed of data. In addition, a large amount of redundant information contained by spectral image data was produced due to the heavy correlation of the images of adjacent bands. The result of smoothing -PCA treatment and explanation variance was shown in Fig. 3. It can be seen from Fig. 3 that the contribution rate for the first principal component was 82.4% according to the principle of principal component extraction, and the cumulative contribution rate is over 85% for the selected principal components, and the number of principal component could be from 6 to 30. PLS regression models were constructed for the different number of principal component instead of whole band data. The results showed that it is the best accurate to select the first 15 principal components to instead whole band data.

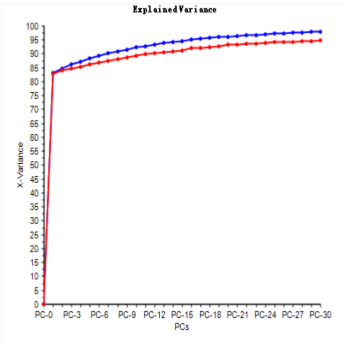

Fig.3. Explained Variance

3.2 Model analysis and validation by PLSR

At first, the different spectral pretreatment methods were used to treat the original whole band, and PCA method was employed, and then PLS model was established according to determining and choosing the optimal number of principal components by the cross validation method.

The result of the calibration model for the spectra of potato on can be seen in Table 2. The determination coefficient R2 for the set of modeling and cross validation are high and root mean square error (RMSE) is low, so the models are accurate. The analysis model of spectrum is the best established by the combination with the treatment of smooth-moving and PCA. The determination coefficient R2 of calibration set is 0.9031 and RMSE is 0.5025. Therefore, according to the pretreatment method, the model is not only accurate, but also the spectral dimension can be reduced. It is advantage for rapid detection.

Table 2. Partial least squares regression (PLSR) modele

method	Spectral points	Factors	RMSE	R^2
None	203	10	0.7482	0.7228
Smooth-PLS	203	9	0.6258	0.7910
Smoot-PCA-PLS	15	8	0.56329	0.82341

4 Conclusions

Potato spectral data were obtained by hyperspectral imaging technology. In the whole band range from 400 to 1000nm, the starch content of potato could be predicted according to the PLS regression model established by the hyperspectral image of potato which was treated by smoothed and PCA. The determination coefficient R2 of the prediction set reached 0.9031 and the root mean square error (RMSE) is 0.5025 for this model. Therefore, there is the better accuracy of prediction for validation model established by the hyperspectral image.

Acknowledgements. Foundation project: Special fund for system construction of a modern agricultural and industry technology (CARS-10 P22); the National Science & Technology Pillar Program during the Twelfth Five-year Plan Period " Research and demonstration for modern agriculture and machinery equipment " project (2014BAD06B03)

References

1. Wang Chunying, Chen Yili, Shi Ying. Research progress of Influence the quality for Fried potato [J]. Chineese Potato.2003.3:23~24
2. Ying Chen, Junming Yue, Ying Ding, et al. Determination of Starch Content in Potato with Acid-hydrolysis[J].Seed.2009.28(9):109-120
3. Zhou Zhu, Li Xiaoyu, Tao Hailong, Gao Hailong.Detection of potato external defects based on hyperspectral imaging technology[J].Transactions of the Chinese Sociely of Agricultural Engineering.2012.28(21):221-229
4. Peng Y,Lu R.Journal of Food Engineering,2006,82:142.
5. Lu Renfu,Peng Yankun.Biosystems Engineering,2006,93(2):161.
6. Ruoyu Zhang.Detection of soluble solids content and firmness of tomato using hyperspectral imaging [D].Zhejiang University.2014.63-75
7. Tian Feng,Li Yuzhen,Ma Zhanlan. The comparison of detection method for Starch Content in Potato.Science and Technology of Qinghai Agriculture and Forestry.1997.1
8. Zhang Xiaoyan,Liu Wei,Xing Li,Zhao Fengmin,Yang Yanchen,Yang Bingnan.An Near-infrared Prediction Model for Quality Indexes of Potato Processing. Infrared. 2012.33(12):33-38
9. ElMasry G,Wang N,Vigneault,C.Detecting chilling injury in Red Delicious apple using hyperspectral imaging and neural networks[J].Postharvest Biol.Technol.2009,52(1):1-8

10.Lu Wanzhen, Yuan Hongfu, Xu Guangtong, et al. Modern analysis technique of near infrared spectroscopy[M]. Beijing: China Petrochemical Press
11. Li Bonian, Wu Libin. Data Analysis Method using MATLAB [M]. Beijing: China Machine Press.106-116.
12. Geladi P.Kowalski B R.Analytica Chimica Acta.1986,185:1.
13. Zhu Fengle,Zhang Hailiang,Shao Yongni,He Yong.Visualization of the Chilling Storage Time for Turbot Flesh Based on Hyperspectral Imaging Technique.2014.34(7):1938-1942.

