Method of Chemometric in Hyperspectral Image Recognition

Xinjun An, Changsheng Zhu

Shandong University of Science and technology, Qingdao 266590, China
xinjunan@163.com

The purpose of this paper is to use chemometric method to analyze the seed production. Based on the hyperspectral image recognition, the carrier of agricultural technology and agricultural production materials is studied. With the wide application of hybrid technology and the influence of many factors in the process of seed production, the phenomenon resulting in seed mixing often occurs, which is a serious threat to the interests of farmers. Therefore, it is of great significance to improve the accuracy and reliability of seed purity testing in order to ensure seed quality and high yield. By analyzing the chemometric methods, a reasonable measurement method is designed and selected to analyze the data, so as to achieve the most effective information in the spectral data. The experiment results show that hyperspectral image technology can reflect the spectral features and image features of the seeds. In addition, a fast and accurate classification model is developed to solve the problems of seed purity detection based on the combination of hyperspectral image technology and chemometric methods. Based on the above findings, it can be concluded that the research has a great significance for the application of hyperspectral imaging technology in the field of non-destructive testing of agricultural products.

1. Introduction

China is a big agricultural country, while seed is the most basic and the most important agricultural means of production. The seed has a special and irreplaceable role in the agricultural means of production, and it is also an important carrier of agricultural technology and production materials. With the development of social economy, the seed has become a special commodity with high technology, which plays an important role in increasing the output of agriculture. High quality seed is an important prerequisite to improve the quality of seed products, but also the key to its variety characteristics. The purity of seed is a good indicator of the quality of seed, which reflects the degree of the typical characteristics of seed varieties. Therefore, it is of great significance to improve the accuracy and purity of seed purity test. Traditional detection techniques, such as morphological identification, protein electrophoresis and DNA molecular detection, have played an important role in the detection of seed purity (Zhang et al., 2015). However, these methods have the disadvantages of time-consuming and long detection period, so it is difficult to get the popularization and promotion.

Therefore, it is important to solve these problems by using chemometric methods to deal with these hyperspectral images more efficiently and to extract the effective information in the best way. It is necessary to reduce the redundancy of the information between the bands and improve the computational efficiency of the model. In addition, the updating of the model should be combined with chemometric methods, so as to improve the robustness of the model. It does not change with other factors such as origin and year, so it is of great significance for the application of hyperspectral imaging technology in the field of non-destructive testing of agricultural products (Pan et al., 2016).

Because of its many advantages, hyperspectral has been widely used in non-destructive testing of agricultural products. The band selection algorithm is used to select the least and most representative subset of bands into a new hyperspectral image space to replace the original hyperspectral image space, and it has little effect on the overall recognition accuracy in the process of dimensionality reduction of hyperspectral image data, which is the ultimate goal of hyperspectral image band selection.

2. Materials and methods

The hyperspectral image recognition system includes two parts: hyperspectral acquisition system and data processing system. Image acquisition system is mainly to achieve the acquisition of hyperspectral images and the preservation of image data. The data processing system includes the following steps: pre-processing, feature extraction, band selection, modelling and model prediction and classification for collected hyperspectral image (Song et al., 2016). The hyperspectral image acquisition system and data processing system are closely introduced in this section.

2.1 Near infrared hyperspectral image acquisition system

Near infrared hyperspectral image acquisition system mainly consists of light source module, image acquisition module, sample delivery platform and computer. The light source module adopts 150W optical fiber halogen lamp (2 Specim, Spectral Imaging Ltd., Oulu, Finland). The light source signal is led out through two branch optical fibers to form a symmetrical linear light source, and the light intensity can be adjusted in the range of 0-100%. The image acquisition module consists of a N 17E-QE line scan imaging spectrometer (Spectral Imaging Ltd. Oulu, Finland) and a CCD camera with a focusing prism. Spectrometer is the core component of near infrared hyperspectral image system, and its function is to decompose complex light into spectral line (Fan et al., 2015). Working principle: When a beam of light passes through a prism-grating-prism assembly and enters into the slit of the monochromator, the light is converged to the parallel light through the optical collimating mirror, and then is dispersed by the diffraction grating according to the different wavelength of the light beam. Finally, the spectrum is formed by focusing mirror according to the different angle of each wavelength. The spectral range of the line scan imaging spectrometer is between 874-1734nm, and the spectral resolution is 5nm. The sample delivery platform is IRCP0076 type electronic control shift platform (Isuzu Optics Corp, Taiwan, China), and the resolution of near infrared hyperspectral image is 320*256 pixels (Ofner et al., 2015).

2.2 Visible-short wave near infrared hyperspectral image acquisition system

The visible-short wave near infrared hyperspectral image acquisition system is composed of imaging spectrometer (1003A-10140 Hyperspec VNIR C-Series, Headwall Photonics Inc., Fitchburg, USA), CCD camera (Pixelfly QE IG285AL, Cooke, USA), lens (10004A-21226 Lens, F/1.4 FL23 mm, Standard Barrel, C-Mount., USA), 150W adjustable power halogen light source (Halogen lamp, EKE, 3250K, Techniquip, USA), electric displacement platform and computer. The slit width of the spectrometer is 25um, and the spectral response range is 400-1000nm. The spectral resolution is 1.29nm/pixel, and the spatial resolution is 0.15mm/pixel. The resolution of visible near infrared hyperspectral image is 1392*1024. Band spacing is 0.64nm/pixel. Image acquisition software is provided by the United States Headwall Co., Ltd. Hyperspec^{TM} software. The resolution of visible near infrared hyperspectral image is 1392*1024, and the band spacing is 0.64nm/pixel. The image acquisition software is Hyperspec^{TM} software provided by the United States Headwall Co., Ltd. 64nm/pixel (Ostendorf, et.al, 2016).

2.3 Calibration of hyperspectral images

Due to the effect of uneven distribution of the intensity of the light source in different bands and the possible dark current in the camera, the noise of the band with weak light intensity is loud. Therefore, the collected hyperspectral image is corrected before data processing. In order to reduce the influence of the change of the light source and the noise of the system, the N standard white board image and the N blackboard image are collected (Cheng et al., 2015). The image of each hyperspectral image is corrected by using the N average white board image and the blackboard image, respectively. The correction formula is as follows:

$$R_c = \frac{R_s \quad R_b}{R_w \quad R_b}$$

(1)

In the Eq, R_c is the corrected image, and R_s is the original image. R_w and R_b are standard whiteboard and blackboard image. Image correction is performed directly by the hyperspectral image acquisition software in the acquisition process. The following image feature extraction is carried out on the R_b software.

2.4 Hyperspectral image data processing system

The data processing technology of seed hyperspectral image is the analysis for the spectral data in the region of interest, so as to achieve the goal of the technology application. The hyperspectral image data processing system includes three parts: preprocessing, feature selection and model building. This part of the operation is implemented on ENVI 4.3 (Research System, Inc, America) and Matlab 2009b (MathWorks, America).

(1) Preprocessing of hyperspectral image

The image preprocessing is performed on the input image feature extraction and modeling before the operation carried out, the main purpose is to enhance the value of image information. At the same time, it can suppress some of the information that may cause unnecessary interference to the feature extraction and data modeling. In the experiment, the final purpose of the preprocessing is to obtain the external contour information of the seed. However, the hyperspectral image data is a three-dimensional data which contains two-dimensional image information and one-dimensional band information. The seed contour information is unique in each band. In order to avoid the complexity of the experimental operation, and the ENVI4.3 software is used to select the clearer image of the corn seed in a certain band to do the next preprocessing (Chen et al., 2016). Finally, the extracted contour can be mapped to all the bands to extract the spectral information of all the bands.

(2) Feature parameter extraction of hyperspectral image

In the process of hyperspectral image recognition, it is very important to determine the characteristics of classification and learning process. The main target of feature selection is to obtain the most effective features for classification. In this paper, the characteristics of the combination of hyperspectral images is used to obtain the profile information of the region of interest according to the image features. Then the average spectral features are extracted as the feature parameters of the seed classification recognition in the region of interest. The formula for calculating the average spectrum in the region of interest is as follows:

$$\mathbf{R}_{mean} = (\lambda)\frac{1}{M\ N}\sum_{i=1}^{M}\ \sum_{j=1}^{N}T_R(\lambda,i,j) \tag{2}$$

In the formula, $T_R(\lambda, i, j)$ is the hyperspectral image in band λ, and there is an equation $\lambda=1,…,K$, $i=1,…M$, $j=1, …N$. K is the total number of bands in the entire spectral range. M and N are the number of transverse and longitudinal pixels of the CCD camera.

3. Selection of near infrared hyperspectral images of maize seeds based on local learning

Because of the large number of bands in hyperspectral data, it can not only increase the storage space of data, but also affect the real-time performance. The main way to solve this problem is to filter the whole wave band, select the most important bands for classification, and realize the dimensionality reduction of hyperspectral image data. This greatly reduces the computational cost and production cost of the actual model, and provides support for the design of multi spectral system. In this paper, the local learning algorithm is introduced to the optimal band selection of the hyperspectral image to achieve the purpose of selecting the best band. The results are stable, so as to realize the rapid selection and identification of maize seeds, and provide a way for the rapid selection and identification of seed purity.

3.1 Hyperspectral imaging system and data acquisition

The hyperspectral imaging system used in this chapter is a near infrared hyperspectral image acquisition system. Image acquisition software is a hyperspectral imaging system acquisition software provided by Taiwan five bell Optical Co., ltd. The movement speed of the platform is 13.8mm/s. Camera exposure time is 3.5ms. In the wavelength range of 874-1734nm, the wavelength interval is 3.36nm, which can be used to obtain near infrared hyperspectral images in 256 bands.

3.2 Hyperspectral image processing

Hyperspectral image processing mainly includes two parts: image preprocessing and feature extraction. In the experiment, firstly, the hyperspectral image is filtered and processed by gray level transformation. The threshold segmentation method is used to extract maize seed profile for the processed hyperspectral image. Therefore, the region of interest can be obtained. Then, the profile of interest is projected to other band images to obtain the region of interest in a total of 256 bands in the 874-1734nm band. At last, the spectral mean of the region of interest of each seed was extracted as the characteristic parameter of the maize seed.

3.3 Experimental results and analysis

(1) The selection of effective bands by local learning

According to the partition of the sample set, the training sample matrix of each set of 130*256 (130 samples, 256 bands) is used as the input of the local learning algorithm. By the local learning algorithm, the initial weights are updated continuously, until the convergence is reached, so that the minimum classification error of the cross validation in the final weight and the transformed feature space is obtained. In the iterative learning process, the minimum weight is sought until convergence, which leads to a small number of irrelevant band

features and is close to 0. Finally, the band weight values are arranged from large to small. According to the actual needs, the effective band is selected as the input, and the PLSDA classification prediction model is established. Figure 1 shows the weight values of the first set of data obtained by local learning. It can be seen from the figure that the weight values of large bands are mostly distributed between 900-950nm and 1700-1734nm.
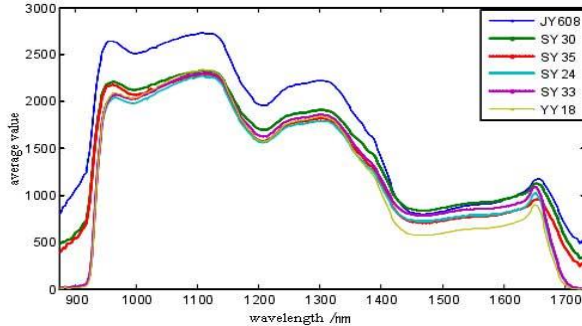

*Figure 1: Average reflective light intensity within class sample*

In the process of learning, it involves the adjustment of the three parameters, which are the kernel width σ, the parameter λ and the learning step η. In a certain range, because of the implied cross validation in logistic regression, their selection has little effect on the experimental results. The final result of the experiment is σ=2 λ=0.5. The learning step η is obtained by linear search, and the final choice is η=0.1.

(2) Modeling and analysis based on PLSDA

In the experiment, according to the weight values of the band, the first 5-15 bands are selected as input to establish PLSDA classification model. In order to simulate the phenomenon of seed production in the actual production, each sample was subjected to 5 random trials (That is, 5 random samples were collected in the test samples). The average accuracy and purity of 5 random samples were used as the final results of each group.
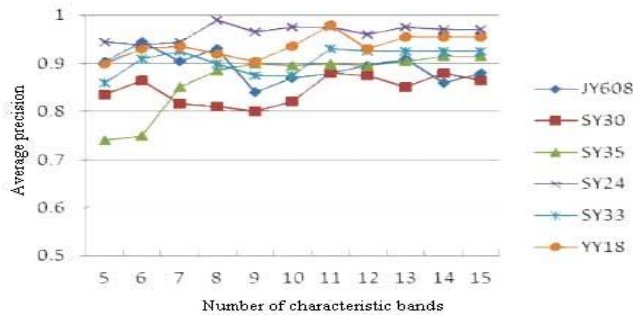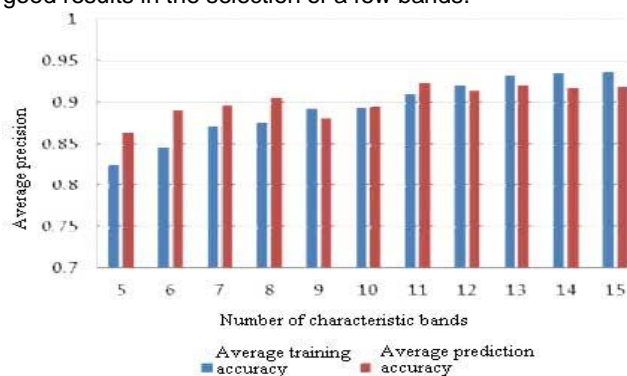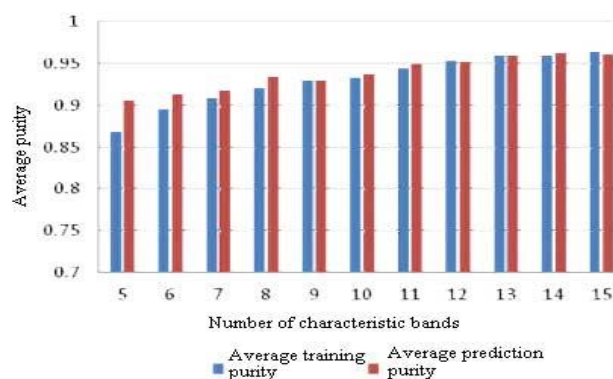

(a) Average prediction accuracy
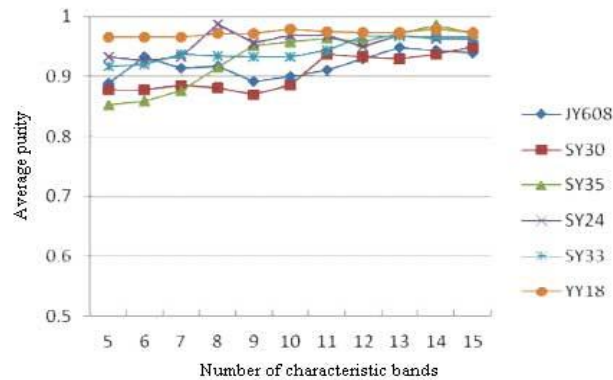
(b) Average prediction purity

Figure 2: 5 random average prediction accuracy and prediction purity of maize seed

Figure 2 shows the average prediction accuracy for each type of maize seed at random after 5 times. Figure 3 shows the average predictive purity of each type of maize seed at random after 5 times. It can be seen from the two images that the prediction accuracy of different types of maize seeds can be achieved between 88.0%-98.0% in the range of 5-15 band, and the predicted purity is between 94.7%-98.7%. The average training purity and the average predictive purity were 96% and 96%, which could meet the current production demand. The experimental results show that the band selected by the local learning algorithm can achieve good results in the selection of a few bands.

(a) Average training accuracy and prediction accuracy

(b) Average training purity and prediction purity

Figure 3: 6 types of average training accuracy and prediction accuracy, average training purity and prediction purity

3.4 The results

Based on the near infrared hyperspectral image in the range of 874-1734nm, the band selection algorithm based on local learning is used to realize the rapid identification of maize seeds. 6 groups of experimental data were designed according to the type of samples. These data are learned by local learning algorithm, and then the effective band is selected according to the size of the weight, and the PLSDA classification prediction model is established. At last, the average training accuracy, the average prediction accuracy, the average training purity and the average prediction purity were used as the indexes of seed purity test. The experimental results show that the band features selected by the local learning algorithm can be used to realize the rapid identification of maize seeds through PLSDA modeling. The result is not affected by the type of maize seed, and has good stability. Therefore, the local learning algorithm is an effective band selection algorithm, which can realize the rapid selection of maize seeds.

4. Conclusions

After the analysis for near infrared hyperspectral image, it can be concluded that there is a serious problem at the diversity of seed market and the wide application of hybrid technology. In addition, the effect of the hyperspectral image on the accuracy and stability of seed purity identification is proved by chemometric methods. Maize seeds are selected as the research object in the experiment process. The local learning algorithm is introduced into the optimum band selection of the near infrared hyperspectral images of seeds, and the partial least squares discriminant analysis (PLSDA) prediction model is established, which realizes

rapid selection of maize seeds under the condition of few bands. In addition, the other result shows that the near infrared hyperspectral images are used to realize the nondestructive testing of agricultural products by the establishment of the mathematical model between the spectral information and the quality parameters of agricultural products.

Reference

Chen Y.N., Sun D.W., Cheng J.H., Gao W.H., 2016, Recent advances for rapid identification of chemical information of muscle foods by hyperspectral imaging analysis. Food Engineering Reviews, 8(3), 336-350.

Cheng J.H., Sun D.W., 2015, Recent applications of spectroscopic and hyperspectral imaging techniques with chemometric analysis for rapid inspection of microbial spoilage in muscle foods. Comprehensive Reviews in Food Science & Food Safety, 14(4), 478–490.

Fan S., Huang W., Guo Z., Zhang B., Zhao C., 2015, Prediction of soluble solids content and firmness of pears using hyperspectral reflectance imaging. Food Analytical Methods, 8(8), 1936-1946.

Ofner J., Kamilli K.A., Eitenberger E., Friedbacher G., Lendl B., Held A., 2015, Chemometric analysis of multisensor hyperspectral images of precipitated atmospheric particulate matter. Analytical Chemistry, 87(18), 9413-20.

Ostendorf R., Butschek L., Hugger S., Fuchs F., Yang Q., Jarvis J., 2016, Recent advances and applications of external cavity-qcls towards hyperspectral imaging for standoff detection and real-time spectroscopic sensing of chemicals. Photonics, 3(2), 28.

Pan T., Sun D., Cheng J., Pu H., 2016, Regression algorithms in hyperspectral data analysis for meat quality detection and evaluation. Comprehensive Reviews in Food Science & Food Safety, 15(3), n/a-n/a.

Song D., Song L., Sun Y., Hu P., Tu K., Pan L., 2016, Black heart detection in white radish by hyperspectral transmittance imaging combined with chemometric analysis and a successive projections algorithm, 6(9), 249.

Zhang B., Fan S., Li J., Huang W., 2015, Detection of early rottenness on apples by using hyperspectral imaging combined with spectral analysis and image processing. Food Analytical Methods, 8(8), 2075-2086.