http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (22): 6729-6734, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.6729.6734

Determination of Water Content in De-enzyming Green Tea Leaves Based on Hyper-spectral Imaging

[1,2]Hu Yong-Guang and [1,2]Chen Pei-Pei and [3]Li Ping-Ping
[1]Institute of Agricultural Engineering, Jiangsu University, Zhenjiang, Jiangsu 212013, China
[2]Key Laboratory of Modern Agricultural Equipment and Technology, Ministry of Education, Zhenjiang, Jiangsu, 212013, China
[3]College of Forest Resources and Environment, Nanjing Forestry University, Nanjing, Jiangsu, 210037, China

Abstract: To determine water content in de-enzyming green tea leaves rapidly and nondestructively, a prediction model is established based on hyper-spectral imaging technology. Diffuse reflection spectra of 192 samples are collected with hyper-spectral imaging system, among which 144 samples are partitioned to the calibration set and 48 samples are partitioned to the prediction set with a partitioning algorithm based on joint X-Y distance. 3 optimal characteristic wavelengths of 980, 1190 and 1389 nm are selected through principal component analysis. After the preprocessing of image cropping, median filtering and normalization, the eigenvalues of the gray levels and textures are extracted based on gray level co-occurrence matrix. Calibration models of water content are established with principal component regression, back propagation neural network and support vector machine regression based on the eigenvalues above. The results show that the support vector machine regression model with 11 variables is the best and its prediction correlation coefficient is 0.8566 while the root mean square error of prediction is 0.0401. The study provides a rapid and nondestructive way to detect water content in de-enzyming green tea leaves which could be used for online monitoring and feedback control of green tea de-enzyming.

Key words: Water content, hyper-spectral imaging, characteristic wavelength, PCA, SVM

INTRODUCTION

De-enzyming plays a crucial role on the quality formation of green tea products since it is the first step of tea processing (Gong and Du, 2012). Because water content in fresh tea leaves reduces sharply during de-enzyming (Jin *et al*., 2003), it is difficult to accurately detect the water content on line and to give feedback control in real-time. Usually the water content is estimated by experiences such as touching or watching tea leaves and slight judgment mistake results in the waste of de-enzyming tea leaves. The traditional method of measuring water content through drying and weighing is off line and time-consuming and destroys the samples. The measurement of water content by testing the resistance or capacitance can be done continuously and quickly, but its accuracy is low and not stable. Near Infrared Spectroscopy (NIRS) has been widely used to detect water content in agricultural products for its advantages of non-destruction, rapidness and high accuracy (Shiroma and Rodriguez-Saona, 2009; Sun *et al*., 2010; Sun *et al*., 2009). However, the measurement by NIRS reflects the comprehensive information of the scanned material and it cannot detect the water content at a specific site.

In the past decade the hyper-spectral imaging technology is widely applied in the fields of agriculture, forestry, geology, environmental monitoring and resource investigation (Ying *et al*., 2000; Tong *et al*., 2006). At the beginning, hyper-spectral imaging technology focuses on overall change of crop water status (Van Herten and Bontsema, 1995; Ahmad and Reid, 1996; Ushada *et al*., 2007). Some studies are made to detect water content in plant leaves (Sun *et al*., 2008; Zhang *et al*., 2011).

To improve the measurement accuracy of water content in de-enzyming green tea leaves, prediction models are established based on the hyper-spectral imaging technology which could provide technical support for online monitoring and feedback control of green tea de-enzyming.

Corresponding Author: Hu Yong-Guang, Institute of Agricultural Engineering, Jiangsu University, Zhenjiang, Jiangsu, 212013, China

MATERIALS AND METHODS

Sample preparation: The sampling experiment is conducted in a tea processing workshop of Maichun Tea Farm, Danyang City. The variety of the tea plant is Longjing No. 43. From the end of leaf cooling to the end of de-enzyming, the samples of different water content are taken off the tea processing line by wearing rubber gloves. Then the samples are loosely placed into glass dishes for spectral imaging collection. By changing the time of de-enzyming and the heating temperature, the large range of the water contents of the samples are achieved. Meanwhile, the uniform distribution of water content with a certain gratitude should be ensured. Totally 192 samples are prepared for the study.

Hyper-spectral imaging system: It consists of a NIR camera (XEVA1.7, 900-1700 nm, 320×256 pixels), F1.2 LENS (OLENS30), halogen lamp (2900.ER+9596.E), optical fiber (PIN9145+9530), displacement platform (MTS120), a controller (SC100) and a computer.

Imaging collection: The exposure time is set to 20ms and the translational speed of the platform is set to 1.25 mm sec^{-1}. The calibration of black and white boards is conducted before imaging collection. Then the samples of de-enzyming tea leaves are scanned via the hyper-spectral imaging system. Finally all the samples are weighed and stored in sealed dry envelopes for further drying.

Water content measurement: A vacuum oven (DZF-6050 with temperature control precision of ±1°C) and an electronic scale (Sartorius CAV214C-210 with weighing precision of 0.0001 g) are used for water content measurement of de-enzyming tea leaves according to Chinese Standard GB/T 8304-2002 (Tea moisture determination). The de-enyzming tea leaves altogether with the envelopes are put into the oven and dried at 103±2°C for 4 h, then taken out of the oven and kept in a desiccator. The weighing is done after the samples cool. The above operation is repeated for one-hour drying until the weighing error of the last twice is less than 0.005 g.

The calculation of water content is given as follows:

$$M = \frac{M_1 - M_2}{M_0} \times 100\%$$

where, M is water content of the de-enyzming tea leaves, %:

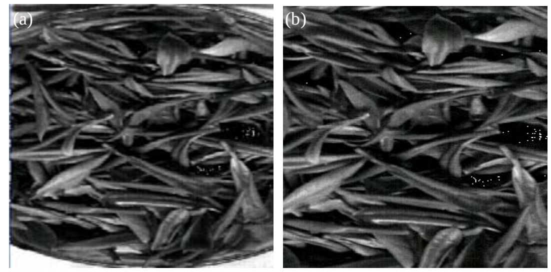

Fig. 1: Raw and cropped hyper-spectral images of the samples (a) Raw image and (b) Cropped image

- M_1 is the quality of the sample and the envelope before drying, g
- M_2 is the quality of the sample and the envelope after drying, g
- M_0 is the quality of the samples before drying, g

The water content of all the samples varies from 41 to 75%.

Preprocessing of Hyper-spectral imaging: Due to imaging conditions, heterogeneous status of the de-enzyming tea leaves and the interference of other random factors, the preprocessing of hyper-spectral imaging must be done.

When scanned, green tea leaves are evenly placed into glass dishes with the diameter of 10cm and the depth of 1 cm. Fig. 1 shows the hyper-spectral images of the samples. To eliminate the background interference, raw images are cropped, so the image size decreases from 1280×1024 pixels (Fig. 1a) to 960×960 pixels (Fig. 1b).

Then further preprocessing is made to the cropped images with the algorithms of smoothing filtering and median filtering. The results are shown in Fig. 2.

Filtering preserves detailed information of the image edges and as much image information of de-enzyming green tea leaves as possible which benefits for the characteristic extraction. Compared with the smooth filtering, the median filtering can more effectively eliminate the interferences from the instrument noises or leaf veins which are irrelevant to the eigenvalues. Therefore, the algorithm of median filter has been adopted in the study. The median filtering of 3×3 is adopted in the study.

In addition, the normalization is done for the eigenvalue of gray levels and textures of every sample to avoid the big differences of order of magnitude for better fitted models (Zhao *et al*., 2007b).

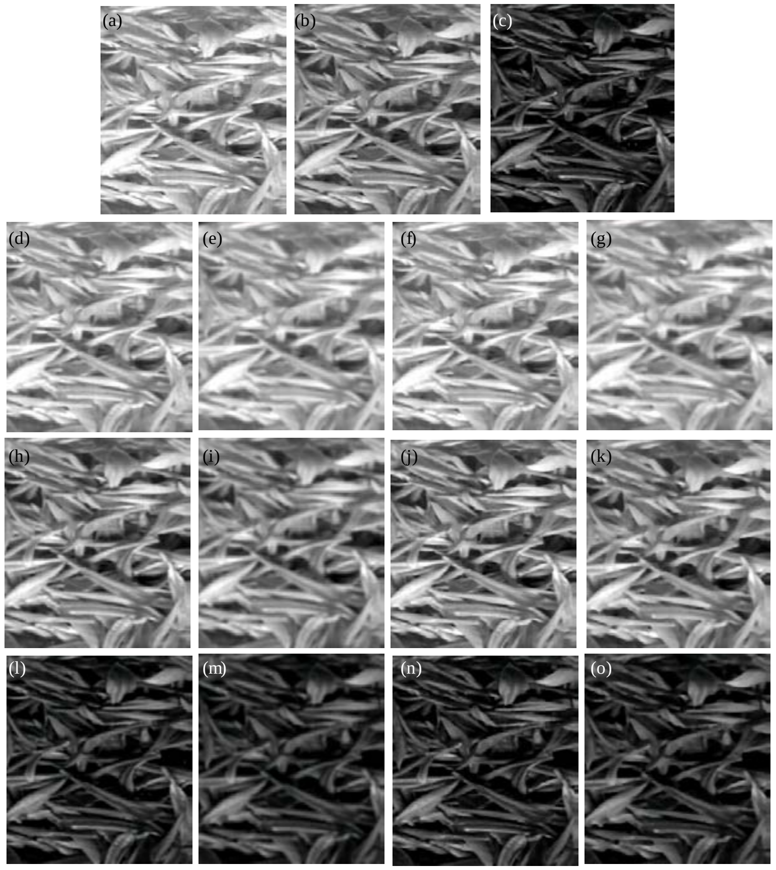

Fig. 2(a-0): Preprocessed images after filtering. Cropped images at different wavelengths (a) 980 nm (b) 1190 nm (c) 1389 nm, Smoothing filtering Median filtering, At the wavelength of 980 nm (d) 3×3 (e) 5×5 (f) 3×3 (g) 5×5, Smoothing filtering Median filtering, At the wavelength of 1190 nm (h) 3×3 (i) 5×5 (j) 3×3 (k) 5×5, Smoothing filtering Median filtering, At the wavelength of 1389 nm (l) 3×3 (m) 5×5 (n) 3×3 and (o) 5×5

Partitioning of sample sets: In order to select the typical samples and to improve the prediction model accuracy, all the samples are partitioned into the calibration set and the prediction set based on joint X-Y distance. According to the proportion of 3:1, 144 samples are in the calibration set and 44 samples are in the prediction set.

RESULTS AND ANALYSIS

Selection of characteristic wavelengths: The images obtained from the hyper-spectral imaging system are 3-D block data with large number of wavebands, high spectral resolution and large amount of data which need dimension reduction through the selection of best characteristic wavelengths for higher efficiency of data processing and the elimination of data redundancy (Zhao *et al*., 2007a; Kim *et al*., 2005; Fang *et al*., 1999).

Principal Component Analysis (PCA) is used to select characteristic wavelengths and the contribution rates of the first 3 components optimization are 97.25%, 2.05 and 0.03%, respectively. Figure 3 shows the first 3 components of PC1, PC2 and PC3. The wavelengths of principal components are the linear combination of the original spectral wavelengths which are not correlative.

The image of PC1 includes the most information of the tea leaves, the image of PC2 includes part of the information and the image of PC1 includes the least information. Therefore, the characteristic wavelengths are selected based on PC1 which is the linear combination of 256 wavelengths described as follows:

$$PC1 = \sum_{i=1}^{256} \alpha_i \lambda_i = \alpha_1 \lambda_1 + \alpha_2 \lambda_2 + \ldots + \alpha_{256} \lambda_{256}$$

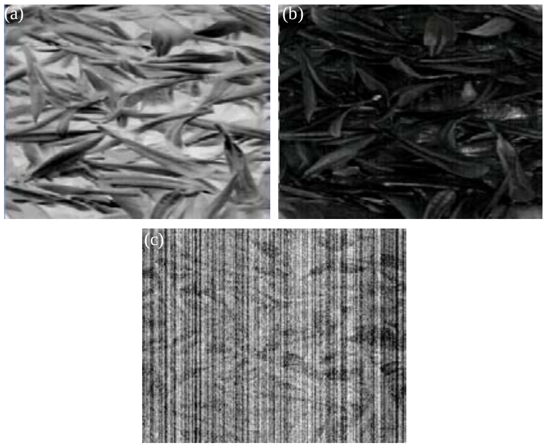

Fig. 3(a-c): The images of the first 3 principal components (a) PC1 (b) PC2 and (c) PC2

where, λ_i is a cetain wavelength and α_i is the coefficient of λ_i.

For every weight coefficient in this combination, the bigger it is, the higher its contribution rate for PC1 is. The top 3 maximum weight coefficients are α_{27}, α_{85} and α_{147} which are corresponding to the wavelengths of 980nm, 1190 and 1389 nm. The images of the 3 characteristic wavelengths are shown in Fig. 2.

Feature extraction: According to the imaging principles, the grey level of one spot on an image at a certain wavelength is corresponding to a spectrum at the same wavelength. The spectrum of de-enzyming green tea leaves includes the information of biochemical components inside. While the water content of the leaves changes, the spectral reflection changes correspondingly.

Figure 4 shows the gray level images of de-enzyming green tea leaves with different water contents. It can be seen that the difference of gray level images is remarkable when the water contents are different. Therefore, the gray level can be used as one of the features for measuring water content in de-enzyming green tea leaves.

As for a specific wavelength, the reflection strength within a sample area can be expressed by the average gray level grades of the images. The average gray level values (AG) of the images are characteristic of the internal information of the leaves at a certain extent, while the Standard Variance (SG) of the gray level reflects the total variability among the gray level values of each pixel and AG. The larger SG is, the more information the images of the leaves contain. For the 3 characteristic wavelengths the corresponding AG and SG are marked as AG_1, AG_2

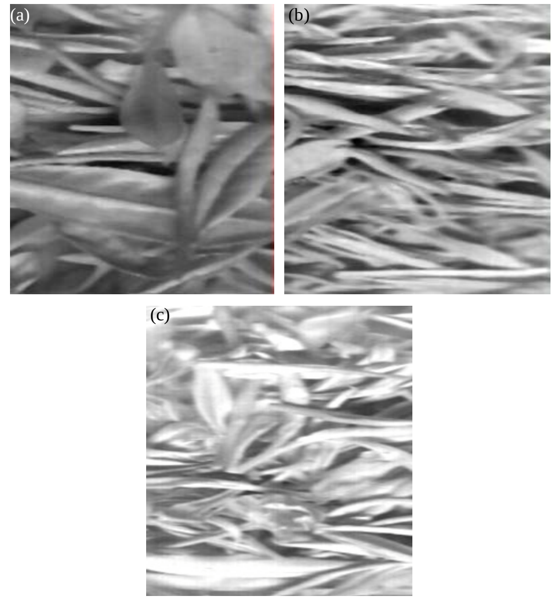

Fig. 4(a-c): Gray level images of de-enzyming tea leaves with different water contents (a) 75% (b) 55% and (c) 45%

Table 1: PCA on gray level images

Principal component number	Eigenvalue	Percentage (%)	Cumulative percentage (%)
1	2.7973	46.6215	46.6215
2	2.2053	36.7551	83.3765
3	0.5803	9.67185	93.0484
4	0.2824	4.7072	97.7556
5	0.09385	1.5642	99.3197
6	0.04082	0.6803	100

AG_3 and SG_1, SG_2, SG_3 for extracting the gray level feature. Textures of tea leaves can be used as a feature for measuring water content since they vary with water contents. The textures are extracted by gray level co-occurrence matrix which covers overall information of the images in direction, neighboring interval and variation amplitude. The second-order moment f_1, contrast ratio f_2, the relevancy f_3, the entropy f_4 and inverse difference moment f_5 are selected for texture extraction (Yang *et al*., 2012).

Model establishment: The principal component analysis is made to the 6 eigenvalues of the gray level image extracted from the 3 characteristic wavelengths. The result is listed in Table 1.

Table 1 shows that the accumulative contribution rate of the first 2 principal components is more than 80%. Therefore, the first 2 principal components are enough to be related with the measurements of water contents for establishing the calibration model. The Principal

Table 2: PCA on texture images

Principal component number	Eigenvalue	Percentage (%)	Cumulative percentage (%)
1	2.9245	58.4895	58.4895
2	1.0624	21.2484	79.7379
3	0.9226	18.4523	98.1902
4	0.0632	1.2650	99.4552
5	0.0272	0.5448	100

Table 3: Prediction result based on the combination of gray level and texture images through several algorithms

	Prediction models	
Algorithms	R	RMSEP
PCR	0.7323	0.0462
BPNN	0.7616	0.0470
SVMR	0.8566	0.0401

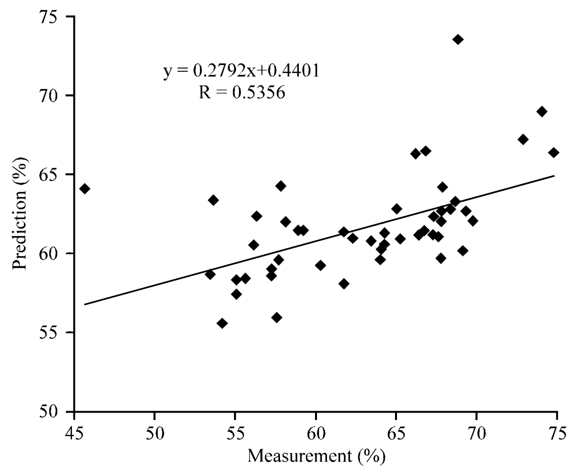

Fig. 5: Prediction result based on gray level feature

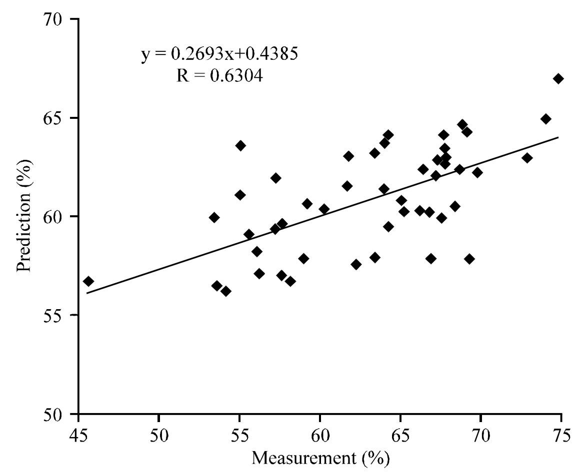

Fig. 6: Prediction result based on texture feature

Component Regression (PCR) model is established with the 6 gray level variables with correlation coefficient (R) of 0.5052 and the Root Mean Square Error (RMSEC) of 0.0616. The model is verified with the 48 unknown samples and the prediction result is shown in Fig. 5. The correlation coefficient of the prediction model is 0.5356 and the Root Mean Square Error (RMSEP) is 0.0564.

The principal component analysis is also made to the 5 eigenvalues of the texture image extracted from the 3 characteristic wavelengths. The result is listed in Table 2.

Table 2 shows that the accumulative contribution rate of the first 2 principal components is less than 80%, but it reaches 98% with the first 3 principal components. Therefore, the first 3 principal components are enough to be related with the measurements of water contents for establishing the calibration model. The PCR model is established with the 5 texture variables with correlation coefficient of 0.4242 and the root mean square error of 0.0684. The model is verified with the 48 unknown samples and the prediction result is shown in Fig. 6. The correlation coefficient of the prediction model is 0.6304 and the root mean square error is 0.0502.

The prediction accuracy of above 2 models is relatively low. Six variables of gray level images and 5 variables of the texture images are combined to rebuild the calibration models. The algorithms of PCR, Back Propagation Neural Network (BPNN) and Support Vector Machine Regression (SVMR) are attempted to improve the model accuracy. The results are listed in Table 3.

Table 3 shows that the new models improve the prediction accuracy by increasing the correlation coefficient between the measurement and the prediction and reducing the root mean square error of the prediction models. Among the models, the SVMR model achieves the best accuracy with the correlation coefficient of 0.8566 and the root mean square error of 0.0401.

CONCLUSION

A rapid and nondestructive way to detect moisture content in de-enzyming green tea leaves is proposed based on hyper-spectral imaging technology.

After the preprocessing of image cropping, median filtering and normalization, 192 samples are partitioned into the calibration set of 144 samples and the prediction set of 48 samples. Three optimal characteristic wavelengths of 980nm, 1190nm and 1389nm are selected with PCA. The features of gray level and texture are extracted to establish a SVM regression model with 11 variables which has achieved the best prediction with correlation coefficient of 0.8566 and the root mean square error of 0.0401.

The model provides the technical support for online monitoring and feedback control of green tea de-enzyming.

ACKNOWLEDGMENTS

The authors would like to express sincere thanks to the financial support by the National High Technology R&.D Program of China under the Grant 2012AA10A508 and the Natural Science Foundation of China under the Grant 31101089.

REFERENCES

Ahmad, I.S. and J.F. Reid, 1996. Evaluation of color representations for maize image. J. Agr. Eng. Res., 3: 185-196.

Fang, R.M., J.R. Cai and L. Xu, 1999. Computer Image Processing Technology and its Applications in Agricultural Engineering. Tsinghua University Press, Beijing.

Gong, X.J. and X. Du, 2012. The variations of main biochemical components during the production process of roasted green tea by automatic production line. J. Sichuan Agric. Univ., 1: 73-77.

Jin, X.Y., J.B. Chen and K.W. Ji, 2003. Tea Processing Engineering. China Agriculture Press, Beijing.

Kim, M.S., A.M. Lefcourt, Y.R. Chen and Y. Tao, 2005. Automated detection of fecal contamination of apples based on multispectral fluorescence image fusion. J. Food Eng., 1: 85-91.

Shiroma, C. and L. Rodriguez-Saona, 2009. Application of NIR and MIR spectroscopy in quality control of potato chips. J. Food Comp. Anal., 22: 596-605.

Sun, H., M.Z. Li, Y.E. Zhang, Y. Zhao and H.H. Wang, 2010. Spectral characteristics of corn under different nitrogen treatments. Spectrosc. Spect. Anal., 3: 715-719.

Sun, J., H.P. Mao, Y.Q. Yang and X.D. Zhang, 2009. Model of moisture content of paddy rice leaf based on canopy spectral reflectance. Trans. Chinese Soc. Agric. Eng., 9: 133-136.

Sun, R.D., H.Y. Yu, C.L. Yu and L. Zhang, 2008. Research on non-destructive detecting of cucumber leave water content based on image procession. J. Agric. Mech. Res., 7: 87-89.

Tong, Q.X., B. Zhang and L.F. Zheng, 2006. Hyperspectral Remote Sensing Principle, Technology and Application. Higher Education Press, Beijing.

Ushada, M., H. Murase and H. Fukuda, 2007. Non-destructive sensing and its inverse model for canopy parameters using texture analysis and artificial neural network. Comput. Electron. Agric., 57: 149-165.

Van Herten, E.J. and J. Bontsema, 1995. Non-destructive crop measurement by image processing for crop growth control. J. Agric. Eng. Res., 61: 97-105.

Yang, D.K., D.W. Hou and Y.D. Bu, 2012. Development of GLCM with image textural feature extraction. Inform. Syst. Eng., 1: 136-136.

Ying, Y.B., X.Q. Rao, Y. Zhao and Y.Y. Jiang, 2000. Advance on application of machine vision technique to automatic quality identification of agricultural products (I). Trans. Chinese Soc. Agric. Eng., 16: 103-107.

Zhang, X.D., H.P. Mao, Y. Zhou Z.Y. Zuo and H.Y. Gao, 2011. Study on detection of moisture content in lettuce leaves based on hyperspectral imaging technology. J. Anhui Agric. Sci., 33: 20329-20331.

Zhao, C.H., W.H. Chen and L. Yang, 2007. Research advances and analysis of hyperspectral remote sensing image band selection. J. Nat. Sci. Heilongjiang Univ., 5: 592-602.

Zhao, W.J., F.Z. Duan, X.M. Liu and Z.Y. Xu, 2007. Special and Practice of ENVI Remote Sensing Image Processing. China Environmental Science Press, Beijing.