Journal of Food Engineering 109 (2012) 531–537

Contents lists available at SciVerse ScienceDirect

Journal of Food Engineering

journal homepage: www.elsevier.com/locate/jfoodeng

Application of hybrid image features for fast and non-invasive classification of raisin

Songjing Wang [a], Kangsheng Liu [a], Xinjie Yu [a,b], Di Wu [c], Yong He [c,*]

[a] Department of Mathematics, Zhejiang University, Hangzhou 310027, China
[b] Ningbo Institute of Technology, Zhejiang University, Ningbo 315100, China
[c] College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China

ARTICLE INFO

Article history:
Received 22 March 2011
Received in revised form 19 October 2011
Accepted 22 October 2011
Available online 9 November 2011

Keywords:
Raisin classification
Morphological feature
Color feature
Texture feature
Machine vision

ABSTRACT

A new approach for the non-invasive classification of raisins is presented based on the hybrid image features, namely morphological, color and texture features. A total of 74 features (8 morphological, 30 color, and 36 textural) were extracted from RBG images. Seven kinds of models were established based on different feature sets. They were three kinds of models established based on single feature set, three kinds of models established based on the combination of two feature sets, and one kind of model established based on the combination of all feature sets. Five kinds of classifiers, namely partial least squares (PLS), linear discriminant analysis (LDA), soft independent modeling of class analogy (SIMCA), and least squares support vector machine (LS-SVM) with linear and radial basis function (RBF) kernels were used for the model establishment based on different feature sets. The best correct answer rates (CAR) of 99% was obtained when LDA was used to establish the classification model based on the combination of all feature sets, which was higher than those of the models established based on single feature set or the combination of two feature sets. The results show that the feature combination is helpful to improve the accuracy of raisin classification. It was concluded that the varieties of raisin could be accurately classified based on RGB image features and the combination of morphological, color and texture features was an accurate way to improve the accuracy of classification.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Raisins are important and valuable export products for many countries. The worldwide export volume of raisins is over 695,000 metric tons in 2009 (USDA, 2009). As different raisins are produced with different varieties and different production techniques, there are differences in quality and price among different raisins. Recently, consumption of raisin has increased, and with it, some fraudulent phenomena have increased as well. In China, every year, to make enormous profits, many underground factories adulterate raisins with low price into raisins with high price. These behaviors badly infringe on the rights and interests of consumers. Therefore a fast and accurate detection method is necessary to classify raisin samples from different varieties.

In China, there is no modern evaluation techniques used for the raisin inspection. The raisin quality classification is commonly carried out sorting manually by growers or other hireling employed by dealers, which is costly and inherently unreliable due to its subjective nature (Huxsoll et al., 1995; Satake et al., 2003; Tang et al., 2007). Automatic raisin classification based on machine vision can abolish inconsistent manual evaluation, and reduce dependence on available manpower (Omid et al., 2010). Therefore, it has drawn considerable interest to utilize machine vision technique in the raisin industry.

A typical machine vision method usually includes a process to extract various features from images as the indicators or descriptions of food qualities. Therefore these extracted features are considered as the key or core elements of the machine vision application in food industry (Ballard and Brown, 1982). Du and Sun (2004) categorized these image features into four kinds, i.e., size, shape, color, and texture, which describes the number of pixels, the boundary of food products, the intensity of pixels, and the spatial arrangement of the level of the pixels in a region, respectively (Anon, 1990). These image features have been widely used in raisin industry for the classification. Omid et al. (2010) developed an automated machine for grading raisins based on shape features. Okamura et al. (1993) used a machine vision technique for grading raisin based on texture features with comparable accuracy and precision to human inspector. Li et al. (2009) utilized an image processing technology and a neural network for identifying the level of raisins based on size and color features and achieved a recognition rate of 92%. Both the surface information and geometry information in images play important roles in quality determination of raisin kernel. The combination of different features can provide more information. Therefore it might be able to improve the accuracy of

* Corresponding author. Tel./fax: +86 571 88982143.
E-mail address: yhe@zju.edu.cn (Y. He).

0260-8774/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jfoodeng.2011.10.028

classification, compared with a single feature. However, few papers have focused on the analysis of combinations of surface (color and texture) and geometry (size and shape) information.

The aim of this paper was to investigate an optimal strategy of using hybrid combinations of morphological, color, and texture features for the classification of raisin kernels. In the following section, methods such as linear discriminant analysis (LDA), soft independent modeling of class analogy (SIMCA), partial least squares (PLS) and least squares support vector machine (LS-SVM) were utilized to analyze the classification performances of these combinations. The entire of the study was subdivided into four stages: sample and image acquisition, image pre-processing, feature extraction, and variety classification.

2. Materials and methods

2.1. Sample preparation and image acquisition

Four varieties of seedless raisins provided from Lvguoyuan Co. Ltd. (Jichang City, Xinjiang Uygur Autonomous Region, China) were investigated in this study. They were Ha-mi-wang, Wu-he-bai, Jin-huang-hou, Wang-zhong-wang. Their information is summarized in Table 1. RGB sample images were captured by Microvision MV-VS078FM/FC industrial CCD digital camera (MV-VS series 1394 interface CCD industrial camera, China). The images were 768 pixels vertically by 1024 pixels horizontally with 8-bit depth. In the experiment, the process of white balance in the camera control software was applied to calibrate the system. A Teflon white calibration plate was used for the process of white balance. The plate was put at the position where raisin was taken its image. Then a white paper was used as the background for the image acquisition. There were 5–10 kernels in each image acquisition process. Illumination was provided by two groups of light sources to avoid shadows. Each group contained four fluorescent lamps (Lotmar, Poland) with the characteristic of 45 W, 220 V, and 50 Hz. The number of Lux of the proposed illumination's setup was 4600 ($\pm$4%) Lux. In order to reduce the shadow, both camera and lamp were vertically arranged to the sample. The complete experimental setup is shown in Fig. 1. Moreover the camera was connected with a computer, which is not shown in Fig. 1. The images of typical raisin of four varieties are shown in Fig. 2. Finally, the RGB images of 300 samples were captured. There were 75 samples for each variety. The calibration set included 200 samples (set 1) and remaining 100 samples consisted the prediction set (set 2). The repeatability was examined by taking images of the 100 samples from set 2 after the storage of 2 weeks (set 3).

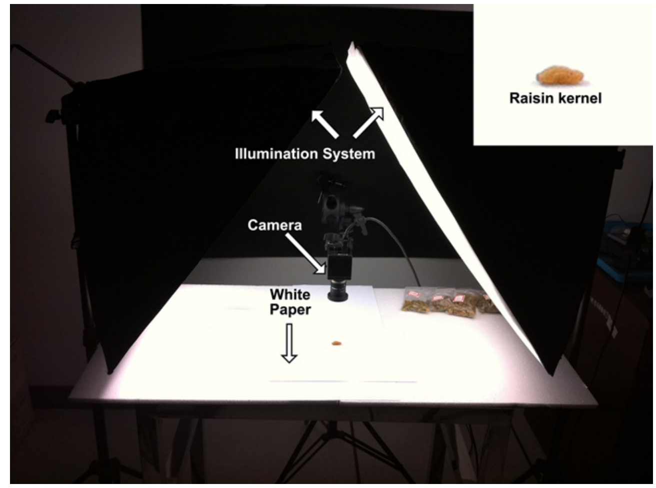

Fig. 1. Experimental setup of image acquisition.

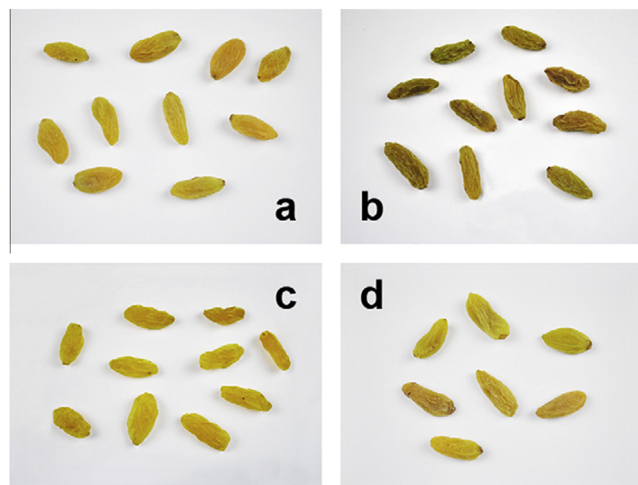

Fig. 2. Images of four varieties of raisin. (a) Ha-mi-wang, (b) Wu-he-bai, (c) Jin-huang-hou, and (d) Wang-zhong-wang.

2.2. Image pre-processing

Images captured by CCD could be influenced by various types of noises. These noises may degrade the qualities of images influent the further image processing (Ballard and Brown, 1982). In order to improve the quality of an original image, image pre-processing algorithms are necessary. In this study, all the image pre-processing and feature extraction were operated by Matlab 7.8 (The MathWorks Inc., Natick, USA).

First, image I (Fig. 3b) is obtained by subtracting background image from original RGB image (Fig. 3a). Then binary image B (Fig. 3c) is obtained from image I by applying a threshold through the following condition:

$$B(i,j) = \begin{cases} 1, & \text{if } I(i,j) > \text{threshold} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

The threshold was obtained using the Matlab function of 'graythresh'.

Segmented image (Fig. 3d) is defined as:

$$B^*(i,j) = B(i,j) \times I(i,j), \quad 1 \leqslant i \leqslant m, \ 1 \leqslant j \leqslant n \tag{2}$$

where m and n are the number of rows and columns, respectively. Further, the center row and column of image B^*, are defined as:

Table 1
Four varieties of raisin kernel.

Varieties	Number	Origin	Color	Size (mm)			Class
				Mean	Max	Min	
Ha-mi-wang	75	Hami, Xinjiang, China	Yellowgreen	15.3	16.9	14.1	Seedless raisin
Wu-he-bai	75	Turfan, Xinjiang, China	Yellowgreen to green	14.5	15.8	13.1	Seedless raisin
Jin-wang-hou	75	Turfan, Xinjiang, Chian	Yellowgreen to yellow	12.4	13.9	11.0	Seedless raisin
Wang-zhong-wang	75	Hami, Xinjiang, China	Yellowgreen to green	16.7	18.1	15.0	Seedless raisin

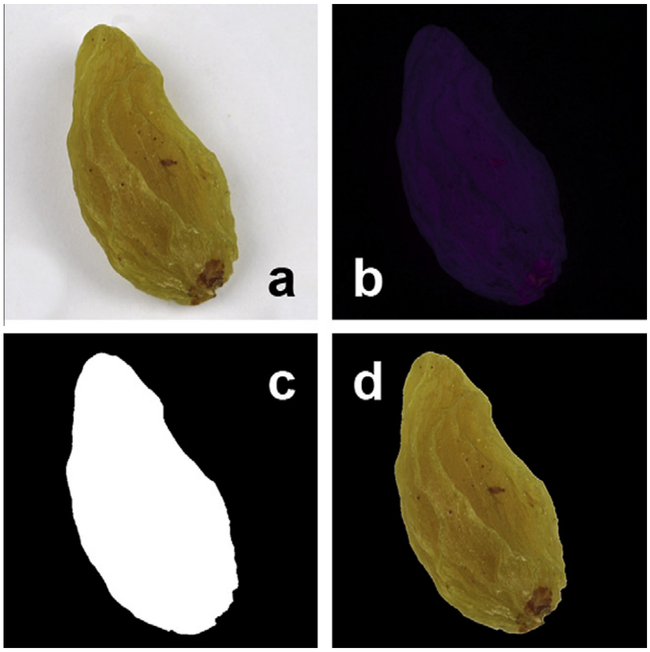

Fig. 3. Image pre-processing: (a) original RGB image, (b) image after being subtracted from background image, (c) binary image, and (d) segmented image.

$$\bar{i} = 1/(m \times n)\sum_{i=1}^{m}\sum_{j=1}^{n} iB^*(i,j) \tag{3}$$

$$\bar{j} = 1/(m \times n)\sum_{i=1}^{m}\sum_{j=1}^{n} jB^*(i,j) \tag{4}$$

Then the covariance matrix of the points in image B^* could be calculated using the formula:

$$C = \sum_{i=1}^{m}\sum_{j=1}^{n} B^*(i,j)[i-\bar{i}, j-\bar{j}]^T[i-\bar{i}, j-\bar{j}] \tag{5}$$

The major axis of image B^* is given by the first eigenvector (the eigenvector with the largest corresponding eigenvalue) of the covariance matrix C. Finally, we rotate the segmented image so that the major axis is the vertical line (Fig. 4). The examples of raisin kernel are shown in Fig. 5a.

2.3. Image feature extraction

2.3.1. Morphological feature extraction

There are several morphological features commonly used, such as area, perimeter, major axis length, minor axis length, maximum radius, minimum radius, mean radius, four invariant shape moments, and 20 harmonics of Fourier descriptors (Gonzalez and Woods, 1992). In this study, morphological features of each sample were extracted from binary image (Fig. 5b) using Matlab function "regionprops". A total of eight morphological features were extracted, which were area, perimeter, major axis length, minor axis length, maximum radius, minimum radius, mean radius, and four invariant shape moments. The description of these eight features is shown in Table 2.

2.3.2. Color feature extraction

Color analysis was based on the histogram from each channel of segmented images (Fig. 5c) in RGB and HSV (Hue, Saturation and Value) color spaces. Color features were extracted from a normalized histogram which was defined as:

$$p(i) = n(i)/N, \quad 0 \leqslant i < l \tag{6}$$

where l is the total number of gray levels in the image, $n(i)$ is number of pixels in ith gray level, N is the total number of pixels. Define the following characteristic coefficients:

$$\mu = \sum_{i=0}^{G-1} ip(i) \tag{7}$$

$$\sigma^2 = \sum_{i=0}^{G-1}(i-\mu)^2 p(i) \tag{8}$$

$$\mu_3 = \sigma^{-3}\sum_{i=0}^{G-1}(i-\mu)^3 p(i) \tag{9}$$

$$\mu_4 = \sigma^{-4}\sum_{i=0}^{G-1}(i-\mu)^4 p(i) - 3 \tag{10}$$

$$e = \sum_{i=0}^{G-1}[p(i)]^2 \tag{11}$$

where μ is the mean value, σ^2 is the variance, μ_3 is the skewness, μ_4 is the kurtosis, e is the energy (Cui and Zeng, 2009). Kurtosis is a measurement of peakedness (Lawrence, 1996). A color feature vector for each color channel of R, G, B, H, S and V is defined as:

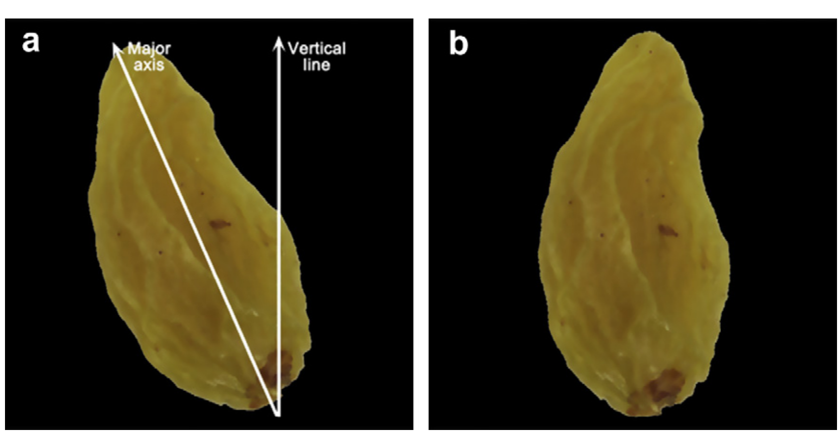

Fig. 4. Rotation of the image using major axis: (a) image before being rotated, and (b) rotated image.

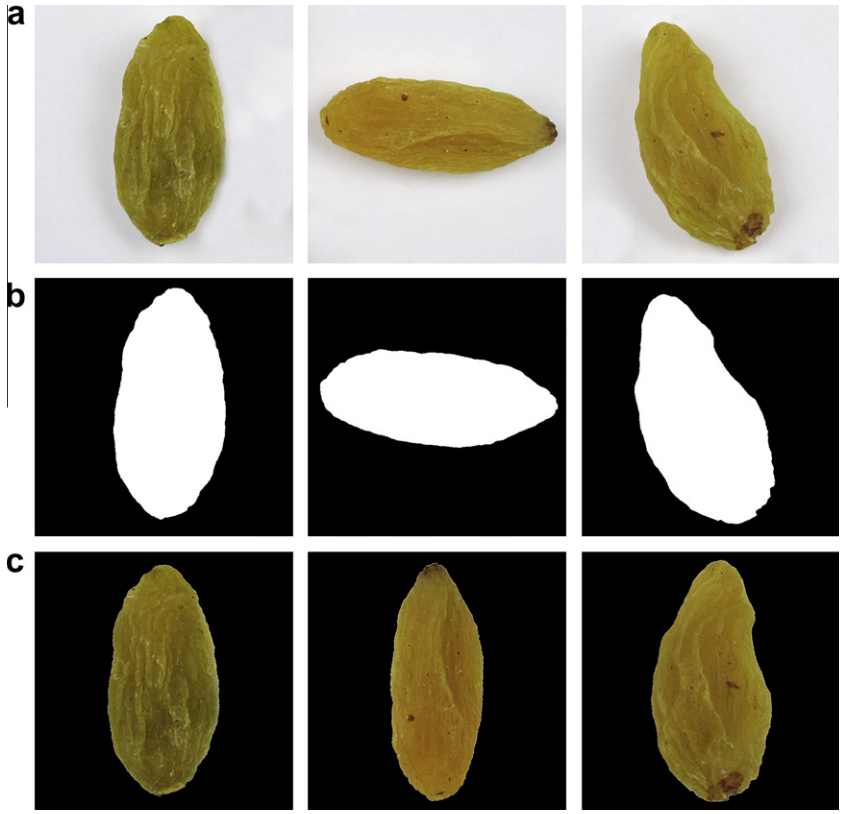

Fig. 5. Examples of raisin kernel: (a) original image of raisin kernel, (b) binary image, and (c) rotated image.

Table 2
Description of eight morphological features.

	Symbol	Properties	Definition or formula
Size features	x_1	Area	The number of pixels in binary image
	x_2	Perimeter	The number of pixels in boundary of binary image
	x_3	Minor axis lenth (*ls*)	The length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region
	x_4	Major axis length (*ll*)	The length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region
Shape features	x_5	Elongation (*el*)	$el = ll/ls$
	x_6	Eccentricity	The eccentricity of the ellipse that has the same second-moments as the region
	x_7	Extent	The proportion of the pixels in the bounding box that are also in the region
	x_8	Solidity (*ss*)	$ss = Area/ConverArea$, where $ConverArea$ is the area of the smallest convex polygon that can contain the raisin

$$\bar{c} = [\mu \quad \sigma^2 \quad \mu_3 \quad \mu_4 \quad e]^T \tag{12}$$

Using color feature vectors defined with Eq. (12), the final 30 color features of RGB and HSV color space were obtained.

2.3.3. Texture feature extraction

Similar to morphological and color, texture is another important image feature that has been found to be an effective tool for pattern recognition in the food industry (Du and Sun, 2006; Zheng et al., 2006; Wu et al., 2008d, 2009a; Chen et al., 2007). Image texture can reflect changes of intensity values of pixels. Texture features could be extracted through using gray level co-occurrence matrix (GLCM) (Haralick et al., 1973). GLCM is a second order statistics method which describes the spatial interrelationships of the gray tones in an image. In this study, texture features were extracted based on GLCM of each color channel (R, G, B, H, S, V) in segmented images (Fig. 5c): contrast (*CO*), dissimilarity (*DI*), homogeneity (*HO*), angular second moment (*ASM*), entropy (*EN*), and correlation (*COR*), which are defined as:

$$CO = \sum_{i,j=0}^{L-1} G_{i,j}(i-j)^2 \tag{13}$$

$$DI = \sum_{i,j=0}^{L-1} G_{i,j}|i-j| \tag{14}$$

$$HO = \sum_{i,j=0}^{L-1} (G_{i,j}/(1+(i-j)^2)) \tag{15}$$

$$ASM = \sum_{i,j=0}^{L-1} G_{i,j}^2 \tag{16}$$

$$EN = \sum_{i,j=0}^{L-1} G_{i,j}(-\ln G_{i,j}) \tag{17}$$

$$COR = \sum_{i,j=0}^{L-1} G_{i,j}\left[(1-\mu_i)(1-\mu_j)/\sqrt[2]{(\sigma_i^2)(\sigma_j^2)}\right] \tag{18}$$

where $G_{i,j}$ is an element of GLCM matrix, μ_i, μ_j are mean values, and σ_i^2, σ_j^2 are standard deviations of matrix rows and columns, respectively (Zheng et al., 2006), defined as:

$$\mu_i = \sum_{i,j=0}^{L-1} i(G_{i,j}) \quad \text{and} \quad \mu_j = \sum_{i,j=0}^{L-1} j(G_{i,j}) \tag{19}$$

$$\sigma_i^2 = \sum_{i,j=0}^{L-1} G_{i,j}(1-\mu_i)^2 \quad \text{and} \quad \sigma_j^2 = \sum_{i,j=0}^{L-1} G_{i,j}(1-\mu_j)^2 \tag{20}$$

A texture feature vector with six features is then defined as:

$$\bar{t} = [CO \quad COR \quad DI \quad ASM \quad EN \quad HO]^T \tag{21}$$

Using texture feature vectors defined with Eq. (21), the final texture feature vector within 36 texture features for RGB and HSV color space was obtained.

2.3.4. Combination of feature extraction

Only using morphological, color, or texture features might not obtain enough accuracy for the raisin kernel quality classification. Therefore, the combination of morphological and color features (morphological–color, containing 8 morphological and 30 color features), the combination of morphological and texture features (morphological–texture, containing 8 morphological and 36 texture features), the combination of color and texture features (color–texture, containing 30 color and 36 texture features), and the combination of morphological, color, and texture features (morphological–color–texture, containing 8 morphological, 30 color and 36 texture features) were analyzed to see if they could improve the classification results. Finally there were seven kinds of models based on different feature sets, three kinds of models established based on single feature set, three kinds of models established based on the combination of two feature sets, and one kind of model established based on the combination of all feature sets.

2.4. Descriptions of multivariate classifier algorithms

PCA is a standard decorrelation technique and following its application one derives an orthogonal projection basis that directly leads to dimensionality reduction and feature extraction (Rao, 1964; Fukunaga, 1991; Zhu, 2007; Wu et al., 2008a). It allows the researcher to reorient the data so that the first few linear combinations known as principal components (PCs) account for as much of the available information as possible.

In order to build a SIMCA model, the samples belonging to each class need to be analyzed by using PCA. For a given class, the resulting model then describes a line, plane or hyper-plane. For each modeled class, the mean orthogonal distance of training data samples from the line, plane or hyper-plane (calculated as the residual standard deviation) is used to determine a critical distance for classification (Wold and Sjostrom, 1977).

Linear discriminant analysis (LDA) is one of the most widely used classification techniques. It aims to find out a linear combination of features that best separates two or more classes of object or event. The resulting combinations could be used as a linear classifier (Belousov et al., 2002).

Partial least squares (PLS) regression is an extension of the multiple linear regression model. PLS algorithm has been found widespread use for multivariate analysis (Wu et al., 2009b, 2010, 2011). PLS analysis was performed to establish a regression model for the prediction of adulterant concentrations (variable matrix Y) based on the spectra (variable matrix X). PLS would be particularly suitable when there were more variables than samples and when there was multicolinearity among the X values.

LS-SVM uses a function called kernel function in order to map the input data from input space to a high-dimensional feature space with fewer training data. In this space, the problem becomes linearly separable. LS-SVM applies least squares error in the training error function (Suykens et al., 2002). LS-SVM has the capability for linear and nonlinear multivariate calibration and solves the multivariate calibration problems in a relatively faster way (Suykens and Vanderwalle, 1999; Lu et al., 2003). It has been found widespread use for multivariate analysis (Wu et al., 2008b, 2009c). In this study, two kernels, namely linear kernel and radial basis function (RBF) kernel were analyzed to choose the best one. The linear kernel type is the simplest and most efficient kernel to perform similarity calculation. RBF kernel is a nonlinear function and a more compact supported kernel, and could reduce the computational complexity of the training procedure while giving good performance under general smoothness assumptions. The optimum γ and σ^2 parameters were selected when they produced the smallest RMSECV (Wu et al., 2008c). The raisin brands were coded as [1,1], [1,−1], [−1,1] and [−1,−1] for variety 1, 2, 3 and 4, respectively.

Unscrambler 9.7 (CAMO PROCESS AS, Oslo, Norway) was used for SIMCA analysis, while Matlab 7.8 was used for the calculation of other methods. The following Matlab toolboxes were used: Statistics Toolbox, and LS-SVM toolbox (LS-SVM v 1.5, Suykens, Leuven, Belgium).

3. Conclusion and discussion

3.1. Unsupervised clustering analysis

After the morphological, color, and texture features were extracted, PCA was executed based on single features (morphological, color, texture) and morphological–color–texture features. Cluster plots of first three PCs of the above feature sets are shown in Fig. 6. Fig. 6a shows that raisin samples of each variety were not well separated based on morphological features. It was noticed that there was a lot of interference between these four varieties. Fig. 6b shows that color features also could not separate raisin samples of different varieties well except variety 3. There was much interference among variety 1, variety 2, and variety 4. When texture features were considered, samples of variety 4 could be partly separated, while samples of other three varieties are mixed (Fig. 6c). When morphological–color–texture features were all considered, samples of four varieties can be separated (Fig. 6d). This was because more information was included in morphological–color–texture features, which were helpful for the variety classification. However, the borders are still not very clear in Fig. 6d. Multivariate analysis algorithms were then further analyzed based on the obtained PCs. For each model, different numbers of PCs were used to establish the model, and the best PC number was determined according to correct answer rates (CAR).

3.2. Analysis of models established based on single feature set

Single feature sets (morphological, color, or texture) were used to establish the classification models based on different classifiers (PLS, LDA, SIMCA, LS-SVM-LIN, LS-SVM-RBF), respectively. The results are shown in first three rows of Table 3. When morphological features were considered, the range of CARs was between 62% and 76%. LDA obtained the best CAR and SIMCA gave lowest CAR. When

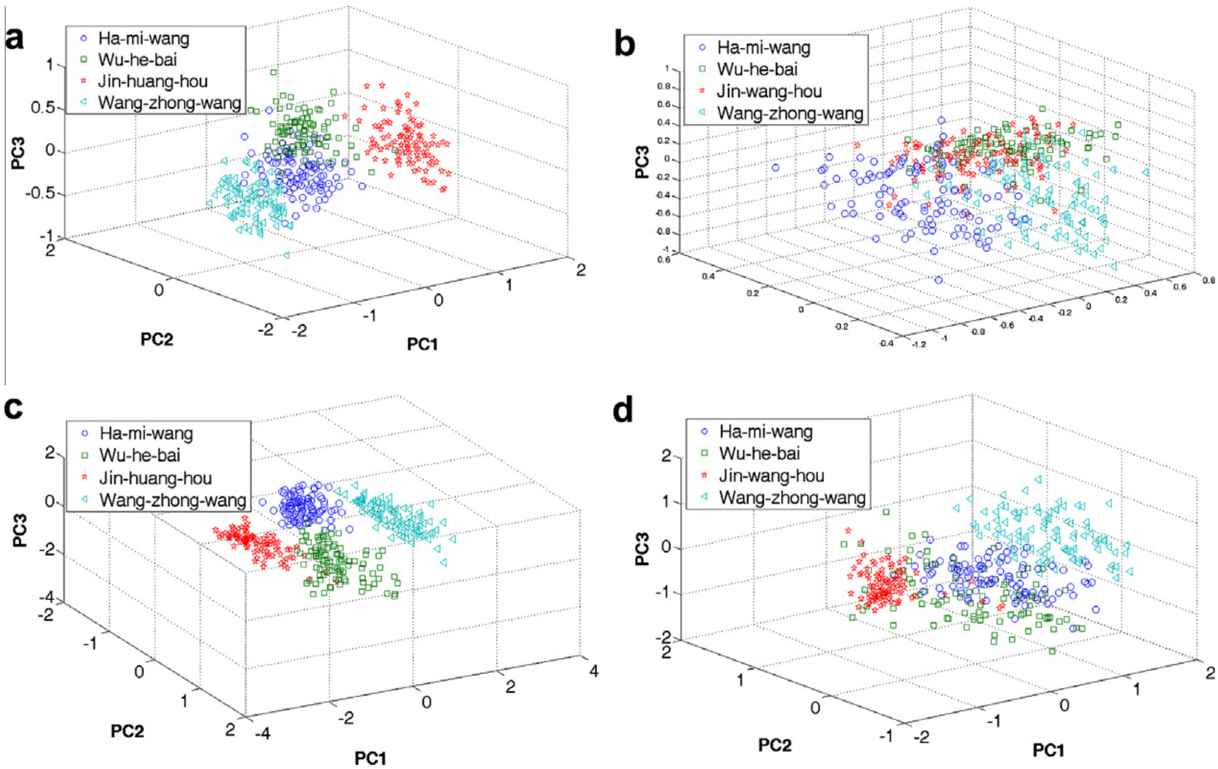

Fig. 6. Scatter plot of first three principle components for four varieties of raisin based on four feature sets: (a) morphological, (b) color, (c) texture, and (d) combination of morphological, color and texture features.

Table 3
Correct answer rates of raisin classification based on different feature sets and different classifiers.

	LS-SVM-LIN (%)	LS-SVM-RBF (%)	PLS (%)	LDA (%)	SIMCA (%)
Morphological	67	73	62	76	60
Color	84	87	87	89	80
Texture	87	86	87	90	81
MC	86	90	87	92	82
MT	89	88	90	92	83
CT	90	89	89	94	83
MCT	91	93	92	99	88
MCT_r	93	92	90	98	90

Morphological: the classification based on morphological features; color: the classification based on color features; texture: the classification based on texture features; MC: the classification based on combination of morphological and color features; MT: the classification based on combination of morphological and texture features; CT: the classification based on combination of color and texture features; MCT: the classification based on combination of morphological, color and texture features; MCT_r: the classification based on combination of morphological, color and texture features with the same raisin samples after 2 weeks.

color features were considered, the range of CARs was between 80% and 89%. SIMCA still had the lowest CAR and LDA the highest CAR. When texture features were considered, similar results were obtained compared to those of color features. The range of CARs was between 81% and 90%. The best CAR (90%) of models established based on single feature set was obtained using texture features with LDA. However the CAR of 90% was not enough for the raisin classification in the industry application.

3.3. Analysis of models established based on the combination of two feature sets

The combinations of two feature sets (morphological–color, morphological–texture, and color–texture) were used to establish the classification models based on five classifiers, respectively. The results are shown in the forth to sixth rows of Table 3. When morphological–color features were considered, the range of CARs was between 82% and 92%. SIMCA obtained the lowest CAR, while the higher CARs were obtained when LDA (92%) and LS-SVM-RBF (90%) were used to establish the classification models. When morphological–texture features were considered, the range of CARs was between 83% and 92%, similar to that of morphological–color features. LDA obtained the best result of the CAR value. The best CAR (94%) of models established based on the combination of two feature sets was obtained using color–texture features with LDA, which is better than that of the models established based on single feature set (90%).

3.4. Analysis of models established based on the combination of all feature sets

All feature set (morphological–color–texture) was used to establish the classification models based on five classifiers, respectively. The results are shown in the seventh row of Table 3. The range of CARs was between 88% and 99%. SIMCA obtained the lowest CAR of 88%. The CARs of LS-SVM-RBF, PLS, LS-SVM-LIN were all above 90%, which was 93%, 92%, and 91%, respectively. The best CAR of 99% was obtained when LDA was used to establish the classification model, which was higher than those of the models established based on single feature set (90%) or the combination of two feature sets (94%).

3.5. Discussion

When the models were established based on single feature set, the average CARs of 67.6%, 85.4%, and 86.2% were obtained based on morphological, color and texture features, respectively. The results show that morphological features had less information related to the varieties of raisin, compared to other two kinds of features. When the models were established based on the combination of two feature sets, the average CARs of 87.4%, 88.4%, and 89.0% were obtained based on morphological–color, morphological–texture and color–texture features, respectively, showing that color–texture features are better than morphological–color and morphological–texture features. When the models were established based on the combination of all feature sets, the average CAR of 92.6% was obtained, which was higher than those of models the models established based on single feature set or the combination of two feature sets. The results show that the feature combination is helpful to improve the accuracy of raisin classification.

By comparing the results of different classifiers, LDA classifier gave the best performance with average CAR of 90.3% followed by LS-SVM-RBF (86.6%), LS-SVM-LIN (84.8%), and PLS (84.8%). The average CAR of SIMCA classifier was lowest (79.6%). Moreover, LDA obtained the best classification results for each kind of feature model. Among the results of all 35 models (7 feature sets $\times$ 5 classifiers), LDA classifier with morphological–color–texture feature obtained the highest CAR of 99% and was considered as the best model for the variety classification of raisin.

3.6. Analysis of repeatability

The repeatability test was examined by taking images of the 100 samples from set 2 after the storage of 2 weeks (considered as set 3), to see if the same results can be obtained with these two image sets. The prediction of samples from set 3 was operated based on the model established with all feature set of 200 samples in set 1. Its results are shown in the eighth rows of Table 3. There was no much difference between the results of set 2 and set 3 based on different classifiers. The average CAR of set 3 was 92.4%, similar to that of set 2. The results indicated that the raw data was robust and repeatable.

4. Conclusions

In this study, a fast, non-destructive and low-cost approach was developed for the classification of raisin kernels using hybrid combinations of morphological, color, and texture features. The results showed that the average CAR of 92.6% of models established based on the combination of all feature sets was better than those of models established based on the combination of two feature sets and models established based on single feature set. The best CAR of 99% was obtained when LDA was used to establish the classification model, which was higher than those of the models established based on single feature set (90%) or the combination of two feature sets (94%). Therefore the combination of morphological, color and texture features can provide more information for the raisin classification. The result of this study was helpful for the raisin quality sorting/grading process. Further study needs to optimize the feature extraction algorithm, and more effort should be put into expanding the variety number and improving the model's robustness and strictness.

Acknowledgements

This study was supported by National Natural Science Foundation of China (10831007) and Ningbo Natural Science Foundation of China (2010A610015).

References

Anon, 1990. IEEE Standard 610.4-1990. In: IEEE Standard Glossary of Image Processing and Pattern Recognition Terminology. IEEE, New York, USA.
Ballard, D.A., Brown, C.M., 1982. Computer Vision. Prentice-Hall, Englewood Cliffs, NJ, USA.
Belousov, A.I., Verzakov, S.A., Von, F.J., 2002. A flexible classification approach with optimal generalisation performance: support vector machines. Chemometrics and Intelligent Laboratory Systems 64, 15–25.
Chen, Q.S., Zhao, J.W., Fang, C.H., Wang, D.M., 2007. Feasibility study on identification of green, black and Oolong teas using near-infrared reflectance spectroscopy based on support vector machine (SVM). Spectrochimica Acta, Part A 66 (3), 568–574.
Cui, Y.Y., Zeng, Z.Y., 2009. Remote sensing image classification based on the HSI transformation and fuzzy support vector machine. In: Proceedings of International Conference on Future Computer and Communication, Kuala Lumpar, Malaysia (CD-ROM).
Du, C.J., Sun, D.W., 2004. Recent development in the applications of image processing techniques for food quality evaluation. Trends in Food Science and Technology 15, 230–249.
Du, C.J., Sun, D.W., 2006. Correlating image texture features extracted by five different methods with the tenderness of cooked pork ham: a feasibility study. Transactions of the ASABE 49 (2), 441–448.
Fukunaga, K., 1991. Introduction to Statistical Pattern Recognition. Academic Press, New York, USA.
Gonzalez, R.C., Woods, R.E., 1992. Digital Image Processing. Addison-Wesley Publishing Co., Reading, MA.
Haralick, R., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics 3, 610–621.
Huxsoll, C.C., Bolin, H.R., Mackey, B.E., 1995. Near-infrared analysis potential for grading raisin quality and moisture. Journal of Food Science 60 (1), 176–180.
Lawrence, T.D., 1996. On the meaning and use of kurtosis. Psychological Methods 2 (3), 292–307.
Li, X.L., Yuan, J., Gu, T.X., Liu, X.Y., 2009. Level detection of raisins based on image analysis and neural network. International Symposium on Neural Networks 6, 343–350.
Lu, C., Van, G.T., Suykens, J.A.K., Van, H.S., Vergote, I., Timmerman, D., 2003. Preoperative prediction of malignancy of ovarian tumors using least squares support vector machines. Artificial Intelligent in Medicine 28, 281–306.
Okamura, N.K., Delwiche, M.J., Thompson, J.F., 1993. Raisin grading by machine vision. Transactions of ASAE 36 (2), 485–492.
Omid, M., Abbasgolipour, M., Keyhani, A., Mohtasebi, S.S., 2010. Implementation of an efficient image processing algorithm for grading raisins. International Journal of Signal and Image Processing 1, 31–34.
Rao, C.R., 1964. The use and interpretation of principal component analysis in applied research. Sankhya 26, 329–358.
Satake, T., Chang, S., Omori, S., Fujioka, O., Sakata, O., 2003. Basic study on grading of Chinese dried green raisin using image information. Journal of the Society of Agricultural Structures 33 (4), 217–224.
Suykens, J.A.K., Vanderwalle, J., 1999. Least squares support vector machine classifiers. Neural Processing Letters 9 (3), 293–300.
Suykens, J.A.K., Van, G.T., De, B.J., De, M.B., Vandewalle, J., 2002. Least Squares Support Vector Machines. World Scientific Publishing, Singapore.
Tang, J.L., Jing, X., He, D.J., 2007. Research of classification of raisin based on BP networks. Journal of Agricultural Mechanization Research 11, 51–53.
USDA, 2009. Raisins: World Markets and Trade. Available at: <http://usda.mannlib.cornell.edu/usda/fas/raiswm//2000s/2009/raiswm-09-04-2009.pdf> (accessed 11 March 2011).
Wold, S., Sjostrom, M., 1977. SIMCA: a method for analyzing chemical data in terms of similarity and analogy. In: Bruce, R., Kowalski, B.R. (Eds.), Chemometrics Theory and Application. American Chemical Society, Washington, pp. 243–282.
Wu, D., Feng, L., Zhang, C., He, Y., 2008a. Early detection of Botrytis cinerea on eggplant leaves based on visible and near-infrared spectroscopy. Transactions of the ASABE 51 (3), 1133–1139.
Wu, D., He, Y., Feng, S., 2008b. Short-wave near-infrared spectroscopy analysis of major compounds in milk powder and wavelength assignment. Analytica Chimica Acta 610 (2), 232–242.
Wu, D., He, Y., Feng, S.J., Sun, D.W., 2008c. Study on infrared spectroscopy technique for fast measurement of protein content in milk powder based on LS-SVM. Journal of Food Engineering 84 (1), 124–131.
Wu, D., Yang, H.Q., Chen, X.J., He, Y., Li, X.L., 2008d. Application of image texture for the sorting of tea categories using multi-spectral imaging technique and support vector machine. Journal of Food Engineering 88 (4), 474–483.
Wu, D., Chen, X.J., He, Y., 2009a. Application of multispectral image texture to discriminating tea categories based on DCT and LS-SVM. Spectroscopy and Spectral Analysis 29 (5), 1382–1385.
Wu, D., Chen, X.J., Shi, P.Y., Wang, S.H., Feng, F.Q., He, Y., 2009b. Determination of alpha-linolenic acid and linoleic acid in edible oils using near-infrared spectroscopy improved by wavelet transform and uninformative variable elimination. Analytica Chimica Acta 634 (2), 166–171.
Wu, D., He, Y., Shi, J.H., Feng, S.J., 2009c. Exploring near and mid infrared spectroscopy to predict trace iron and zinc contents in powdered milk. Journal of Agricultural and Food Chemistry 57 (5), 1697–1704.
Wu, D., He, Y., Nie, P.C., Cao, F., Bao, Y.D., 2010. Hybrid variable selection in visible and near-infrared spectral analysis for non-invasive quality determination of grape juice. Analytica Chimica Acta 659 (1–2), 229–237.
Wu, D., Nie, P.C., Cuello, J., He, Y., Wang, Z.P., Wu, H.X., 2011. Application of visible and near infrared spectroscopy for rapid and non-invasive quantification of common adulterants in Spirulina powder. Journal of Food Engineering 102 (3), 278–286.
Zheng, C.X., Sun, D.W., Zheng, L.Y., 2006. Recent applications of image texture for evaluation of food qualities – a review. Trends in Food Science and Technology 17 (3), 113–128.
Zhu, L., 2007. Face recognition based on two-dimensional image principal component analysis. Journal of Zhejiang University (Engineering Science) 41 (2), 254–267.