Food Bioprocess Technol (2012) 5:1552–1563
DOI 10.1007/s11947-011-0531-9

ORIGINAL PAPER

Raisin Quality Classification Using Least Squares Support Vector Machine (LSSVM) Based on Combined Color and Texture Features

Xinjie Yu · Kangsheng Liu · Di Wu · Yong He

Received: 9 June 2010 / Accepted: 31 January 2011 / Published online: 22 February 2011

Abstract In this paper, an approach based on combined color and texture features to classify raisins is presented. Least squares support vector machine (LSSVM), linear discriminant analysis, and soft independent modeling of class analogy were used to construct classification models. A total of 480 images were captured from four grades of raisin samples by a Basler 601 fc IEEE1394 digital camera, 200 images were randomly selected to create calibration model (training set), and remaining images were used to verify the model (prediction set). Color features and texture features were obtained from two color spaces: red–green–blue and hue–saturation–intensity using histogram method and gray level co-occurrence matrix method, respectively. Our results indicate that the best performance with about 95% of average correct answer rate is achieved by LSSVM using combined color and texture features from HSI color space. This result is significantly higher than the performance of solely used color or texture features. The combined color and texture features coupled with a LSSVM classifier are a highly accurate way for raisin quality classification.

Keywords Raisin · Classification · Color features · Texture features · Least squares support vector machine

X. Yu · K. Liu
Department of Mathematics, Zhejiang University,
Hangzhou 310027, China

X. Yu
Ningbo Institute of Technology, Zhejiang University,
Ningbo 315100, China

D. Wu · Y. He (✉)
College of Biosystems Engineering and Food Science,
Zhejiang University,
Hangzhou 310029, China
e-mail: yhe@zju.edu.cn

Introduction

Raisins are produced in most geographic regions of the world, and consumption occurs in all cultures. The USA is the world's leading raisin producer. Other important raisin-producing countries are Turkey, China, Iran, Chile, South Africa, Greece, Australia, and Uzbekistan (Williamson and Carughi 2010). According to USDA's Foreign Agricultural Service, world raisin exports in 2009 totaled more than 695,000 metric tons (USDA 2009). Raisins are produced with different grapes and different production techniques, resulting in huge differences in quality and price among different raisins. Raisin quality classification is currently carried out sorting manually at packing houses. Manual evaluation and classification of raisins 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).

In recent years, some machine vision techniques have been applied to raisin industry for quality evaluation and classification. Okamura et al. (1993) used a machine vision technique for grading raisins. They used visual surface features of wrinkle edge density, average gradient magnitude, angularity, and elongation for predicting raisin grade with comparable accuracy and precision to human inspector. Li et al. (2009) utilized image processing technology and neural network to identify the quality of raisins by chroma characteristics and achieved an average recognition rate of 92%. Abbasgolipour et al. (2010) designed and fabricated a machine vision system to sort desired and undesired raisins (two classes) according to their hue, saturation, and intensity color features with a sorting accuracy (correct classification rate) of 93%.

However, in all these works, each color and texture information was studied independently, and their performance could potentially be improved by using combined color and texture features for highly accurate industry application. The quality of a raisin is mainly based on visual features such as color, degree of wrinkles, and shape (Christensen 2000). Therefore, both color information and texture information are important in grading of raisins. Although much research has been done on raisin classification using separate color/texture features, less attention has been paid to the performance of combined color and texture features.

Support vector machine (SVM) is a powerful methodology for solving problems in nonlinear classification, function, and density estimation (Cristianini and Shawe 2000), with a good theoretical basis in statistical learning theory (Vapnik 1998). It solves binary classification problem by finding maximal margin hyperplanes in terms of a subset of the input data (support vectors) between different classes. If the input data are not linearly separable, SVM firstly maps data into a high-dimensional feature space, and then classifies data by maximal margin hyperplanes (Chen et al. 2007). Suykens and Vandewalle (1999) have modified Vapnik's standard SVM classifier into least squares support vector machine (LSSVM). LSSVM is the improvement and amelioration of standard SVM. It not only encompasses similar advantages as SVM, but also has additional advantages because it solves regression problem using a set of linear equations instead of quadratic programming (Suykens et al. 2002). Therefore, LSSVM is much easier to use and has a shorter computing time compared to SVM.

It is a new idea that raisin would be classified according to its quality level using combined color and texture features coupled with a LSSVM classifier. The objectives of this study were (1) to compare the discrimination performances of combined color and texture features to that of separate color features or separate texture features, (2) to find optimal set of color and texture features which gives the best result for raisin quality classification, and (3) to develop a model for raisin quality classification by LSSVM.

Materials and Methods

Samples and Image Acquisition

Four grades (grades 1–4) of golden seedless raisins from various sources were investigated in this work. They are Ha-mi-wang (grade 1, originating from Hami), Tian-shan no. 1 (grade 2, originating from Wulumuqi), Jin-huang-hou (grade 3, originating from Turpan), and Wang-zhong-wang (grade 4, originating from Laiyan). All raisins were purchased from a local raisin sales company in Zhejiang, China and their information is summarized in Table 1. In our experiment, grade levels were evaluated by three trained raisin experts according to total raisin quality, including its taste, color, degree of wrinkles, and shape. Each expert's score was dependent on taste (25%), color (25%), degree of wrinkles (25%), and shape (25%). The average of three experts' scores was used to determine the raisin grade level.

Glass containers (90 mm in diameter and 15 mm in height) were filled with raisins, each container representing a sample. Each grade had 120 samples for a total of 480 samples. Original RGB sample images, 480 pixels vertically by 640 pixels horizontally with 24-bit depth, were captured from containers filled with raisins using a Basler 601fc IEEE1394 "FireWire" complementary metal oxide semiconductor camera (Germany) with a computer M3514-MP C-mount lens $F/1.4$ with f=35 mm (Japan). Automatic exposure control and gamma correction were implemented by the software BCAMViewer V1.8 (Basler vision technologies, Germany). A tungsten halogen lamp (Lowell pro-lamp 14.5 V) was used. In order to reduce the shadow, both camera and lamp were vertically arranged with respect to raisin sample with a distance of 500 mm between the camera (lamp) and the sample. Then a total of 480 images were obtained.

Color Feature Extraction

Color analysis was based on the histogram from each channel of color image in RGB and HSI color space (Cui and Zeng 2009). Color features were extracted from a normalized histogram, which was calculated using the expression:

$$p(i) = \frac{1}{NM} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \delta(f(x,y),i), \quad 0 \leq i < G \tag{1}$$

where G being the total number of gray levels in the image, $f(x,y)$ represents the gray level at location of coordinate (x,y), N and M are image width and height, and δ is a Kronecker delta function defined like:

$$\delta(j,k) = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases} \tag{2}$$

Color features provide information on the distribution of pixels on a digital image, but do not give any information on their relative positions. Thus, these are the features characterizing contrast (strength of the patterns) and the color distributions of the image, but

Table 1 Characteristics of the four grades of golden raisins used in this research

Grade level	Origin	Color	Shape	Production time
Grade 1	Hami, Xinjiang, China	Golden	Homogenous in size, fine and uniform wrinkling	February 2010
Grade 2	Wulumuqi, Xinjiang, China	Bright yellow	Fine and uniform wrinkling	October 2009
Grade 3	Turpan, Xinjiang, China	Brown yellow	Less homogenous in size	January 2010
Grade 4	Laiyan, Shandong, China	Yellowish green	Misshapen and less uniform wrinkling	November 2009

not its spatial structure. It is possible to define the following characteristic coefficients:

$$\mu = \sum_{i=0}^{G-1} ip(i) \tag{3}$$

$$\sigma^2 = \sum_{i=0}^{G-1} (i-\mu)^2 p(i) \tag{4}$$

$$\mu_3 = \sigma^{-3} \sum_{i=0}^{G-1} (i-\mu)^3 p(i) \tag{5}$$

$$\mu_4 = \sigma^{-4} \sum_{i=0}^{G-1} (i-\mu)^4 p(i) - 3 \tag{6}$$

$$e = \sum_{i=0}^{G-1} [p(i)]^2 \tag{7}$$

$$h = -\sum_{i=0}^{G-1} p(i) \ln[p(i)] \tag{8}$$

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, and h is the entropy of the analyzed image histogram (Haralick et al. 1973). After calculating given color features, a color feature vector is obtained for each color channel of R, G, B, H, S, and I defined like:

$$\overline{c} = \begin{bmatrix} \mu & \sigma^2 & \mu_3 & \mu_4 & e & h \end{bmatrix}^T \tag{9}$$

Using feature vectors defined with Eq. 9, the final color feature vector within 18 color features for RGB or HSI color space can be defined like:

$$\overline{C_k} = \begin{bmatrix} \overline{c_{k,1}}^T & \overline{c_{k,2}}^T & \overline{c_{k,3}}^T \end{bmatrix}^T \tag{10}$$

where $\overline{C_k}$ is color feature vector, k represents the RGB or HSI color space, and subscripts 1, 2, and 3 denote R, G, and B channels or H, S, and I channels.

Texture Feature Extraction

Texture is an important image feature and has been applied greatly in the food industry for quality evaluation and inspection (Du and Sun 2006; Zheng et al. 2006; Wu et al. 2009; Milde et al. 2010; Yi et al. 2010; Kumar and Mittal 2010). In this study, we used gray level co-occurrence matrix (GLCM) as a statistical texture analysis technique. The GLCM is a second order statistics method which describes the spatial interrelationships of the gray tones in an image. It contains elements that are counts of the number of pixel pairs which are separated by certain distance and at some angular direction. In this work, GLCM was computed based on two parameters, which were the distance between the pixel pair d and their angular relation θ. The distance d is measured by number of pixels (d=1 for neighboring pixels, etc.). The θ can be any angle between 0° and 360°. For image I, defined square window $N \times N$, brightness levels i and j, the non-normalized GLCM P_{ij} are defined by:

$$P_{ij,\theta} = \sum_{x=1}^{N} \sum_{y=1}^{N} C\{(I(x,y) = i) \wedge (I(x \pm d\theta_0, y \mp d\theta_1) = j)\} \tag{11}$$

where $C\{\cdot\}$=1 if the argument is true and $C\{\cdot\}$=0 otherwise. The ± and ∓ signs in Eq. 11 mean that each pixel pair is counted twice: once forward and once backward in order to make the GLCM diagonal symmetric. In this work, sample image was filled with raisin pixel information, and orientation of raisin in the sample image was randomly distributed. Therefore, we can use any one direction (such as 0°, 45°, 90°, and 135°) to calculate texture features from the GLCM. The texture features were calculated when the θ equals to 0° (with θ_0=0, and θ_1=1) and the d equals to 5.

Before texture calculation, we normalized the GLCM and let it represent probabilities instead of counts. Normalization involves dividing by the total number of counted pixel pairs. Haralick et al. (1973) studied six texture features based on GLCM: contrast (CO), dissimilarity

(DI), homogeneity (HO), angular second moment (ASM), entropy (EN), and correlation (COR). For normalized GLCM P_{ij} and L gray levels, these features are defined as:

$$\mathrm{CO} = \sum_{i,j=0}^{L-1} P_{i,j}(i-j)^2 \quad (12)$$

$$\mathrm{DI} = \sum_{i,j=0}^{L-1} P_{i,j}|i-j| \quad (13)$$

$$\mathrm{HO} = \sum_{i,j=0}^{L-1} \frac{P_{i,j}}{1+(i-j)^2} \quad (14)$$

$$\mathrm{ASM} = \sum_{i,j=0}^{L-1} {P_{i,j}}^2 \quad (15)$$

$$\mathrm{EN} = \sum_{i,j=0}^{L-1} P_{i,j}\left(-\ln P_{i,j}\right) \quad (16)$$

$$\mathrm{COR} = \sum_{i,j=0}^{L-1} P_{i,j}\left[\frac{(1-\mu_i)\left(1-\mu_j\right)}{\sqrt[2]{\left(\sigma_i^2\right)\left(\sigma_j^2\right)}}\right] \quad (17)$$

where μ_i and μ_j are mean values and σ_i^2 and σ_j^2 are standard deviations of matrix rows and columns, respectively, defined like:

$$\mu_i = \sum_{i,j=0}^{L-1} i\left(P_{i,j}\right) \quad \text{and} \quad \mu_j = \sum_{i,j=0}^{L-1} j\left(P_{i,j}\right) \quad (18)$$

$$\sigma_i^2 = \sum_{i,j=0}^{L-1} P_{i,j}(1-\mu_i)^2 \quad \text{and} \quad \sigma_j^2 = \sum_{i,j=0}^{L-1} P_{i,j}\left(1-\mu_j\right)^2 \quad (19)$$

After calculating given texture features, a texture feature vector was obtained from GLCM for every color channel of R, G, B, H, S, and I defined like:

$$\bar{t} = [\,CO \quad COR \quad DI \quad ASM \quad EN \quad HO\,]^T \quad (20)$$

Using texture feature vectors defined with Eq. 20, the final texture feature vector within 18 texture features for RGB or HSI color space can be defined like:

$$\overline{T_k} = \left[\,\overline{t_{k,1}}^T \quad \overline{t_{k,2}}^T \quad \overline{t_{k,3}}^T\,\right]^T \quad (21)$$

where $\overline{T_k}$ is texture feature vector, k represents the RGB or HSI color space, and subscripts 1, 2, and 3 denote R, G, and B channels or H, S, and I channels.

Combined Color and Texture Feature Extraction

As shown later in this paper, using only color features or texture features for raisin quality classification does not give sufficient results, therefore different combinations of color features and texture features were calculated in order to obtain better classification results. The final combined color and texture feature vector is obtained by combining the vectors defined with Eqs. 10 and 21, using the expression:

$$\overline{CT_k} = \left[\,\overline{C_k}^T \quad \overline{T_k}^T\,\right]^T \quad (22)$$

where $\overline{CT_k}$ is combined feature vector, $\overline{C_k}$ is the color feature vector, and $\overline{T_k}$ is the texture feature vector. Subscripts k represent the RGB or HSI color space.

Principal Component Analysis

Principal component analysis (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. 2008b). Let $X \in R^n$ be a random vector, where n is the dimensionality of the vector and $\sum X \in R^{n\times n}$ be the covariance matrix of X. The PCA of a random vector X factorizes the covariance matrix $\sum X$ into the form, $\sum X = \Phi\,\Lambda\Phi^T$, where Φ is an orthogonal eigenvector matrix and Λ is a diagonal eigenvalue matrix with diagonal elements in decreasing order.

An important property of PCA is its optimal signal reconstruction in the sense of minimum mean square error when only a subset of principal components is used to represent the original signal. Following this property, an immediate application of PCA is the dimensionality reduction by projecting a random vector X onto the eigenvectors, $Z = P^T X$ ($X \in R^n$, and $Z \in R^m$), where $P \in R^{n\times m}$ is a subset of eigenvector matrix Φ and $m<n$. The lower dimensional vector Z (with a dimensionality of m) captures the most expressive features of the original data X (Moon and Phillips 1998). The original data X can thus be largely reduced by observing few principal components (PCs) without significant loss of useful information, and each variable in X has its own weighting or loading value on each PC, which usually can be used to identify important variables that are responsible for the specific features appeared in corresponding PCs (Purcell et al. 2007).

In this study, the obtained feature vectors are high-dimensionality data, so that the cluster performances by these high-dimensionality data are difficult to be described directly. After dimensionality reduction, the cluster performances can be compared in two-dimensional PCs plot using the first two primary principal components, which makes it easy to explain the cluster results of raisin samples by different feature vectors. Furthermore, loadings in principal component explain the contribution of each variable in feature vector. They can be used to determine the most discriminant features in the feature vector.

Linear Discriminant Analysis and Soft Independent Modeling of Class Analogy

Among many possible techniques for data classification linear discriminant analysis (LDA) is a commonly used one. LDA maximizes the ratio of between class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability (Michie et al. 1994; Belousov et al. 2002). LDA is considered a "supervised" or modeling method because the mean observation for each of a number of predefined groups is calculated and new observations are assigned to the group whose mean is closest (Burks et al. 2000). In this study, LDA was carried out using DISCRIM procedure in SAS (SAS Institute Inc., USA). The DISCRIM procedure was used to generate discriminant functions for classifying raisins into one of the four grades based on different features.

Soft independent modeling of class analogy (SIMCA) is a method for supervised data classification that requires a training data set consisting of samples (or objects) with a set of attributes and their class membership. In order to build classification models, the samples belonging to each class need to be analyzed using principal components analysis (PCA). For a given class, the resulting model then describes a line, plane, or hyperplane. For each modeled class, the mean orthogonal distance of training data samples from the line, plane, or hyperplane (calculated as the residual standard deviation) is used to determine a critical distance for classification. New observations are projected into each principal component model, and the residual distances are calculated. An observation is assigned to the model class when its residual distance from the model is below the statistical limit for the class (Wold & Sjostrom 1977). In this study, SIMCA was carried out using Unscrambler ver 9.8 (Camo Process AS, Oslo, Norway).

Least Squares Support Vector Machine

The standard LSSVM algorithm is showing as follows. Assume a set of training samples defined as input data vectors $X = \{x_1, ..., x_i\}$, with $x_i \in R^n$ (i=1,…, N), and a set of corresponding class labels $Y = \{y_1, ..., y_i\}$, with $y_i \in \{1, 2, ..., l\}$ (i=1,…, N), where N is the number of training samples, n is the dimensionality of input data vector, and l is the total number of classes. The following model is constructed by using nonlinear mapping function $\varphi(\cdot)$, which maps the input data to a higher dimensional feature space:

$$f(x) = w^T \varphi(x) + b \tag{23}$$

where $w \in R^n$ is the weight vector and b is the bias. When the least squares support vector is used as a soft testing tool, a new optimization problem is formulated in the case of structural risk minimization (SRM):

$$\min_{w,b,e} J(w, b, e) = \frac{1}{2}\|w\|^2 + \gamma \frac{1}{2} \sum_{i=1}^{N} e_i^2 \tag{24}$$

subject to the constraints:

$$y_i\left[w^T \varphi(x_i) + b\right] = 1 - e_i,\ i = 1, 2, ..., N \tag{25}$$

where e_i is the classification error and γ is the relative weight of the classification error (regularization parameter).

Lagrangian function of the convex optimization program could be established by introducing a dual set of positivity Lagrange multipliers $a_i \in R$ and was expressed as:

$$L(w, b, e, a) = J(w, b, e) - \sum_{i=1}^{N} a_i\left\{y_i\left[w^T \varphi(x_i) + b\right] - 1 + e_i\right\} \tag{26}$$

The corresponding condition of Karush–Kuhn–Tucker (Karush 1939; Kuhn and Tucker 1951) was given as:

$$\begin{cases} \dfrac{\partial L}{\partial w} = 0 \to w = \sum_{i=1}^{N} a_i y_i \phi(x_i) \\ \dfrac{\partial L}{\partial b} = 0 \to \sum_{i=1}^{N} a_i y_i = 0 \\ \dfrac{\partial L}{\partial e_i} = 0 \to a_i = \gamma e_i \\ \dfrac{\partial L}{\partial a_i} = 0 \to y_i\left[w^T \phi(x_i) + b\right] - 1 + e_i = 0, i = 1, ..., N \end{cases} \tag{27}$$

Also, Eq. 27 is shown as a matrix:

$$\left[\begin{array}{ccc|c} I & 0 & 0 & -Z^T \\ 0 & 0 & 0 & -Y^T \\ 0 & 0 & \gamma I & -I \\ \hline Z & Y & I & 0 \end{array}\right] \left[\begin{array}{c} w \\ b \\ e \\ \hline a \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ \hline \vec{1} \end{array}\right] \tag{28}$$

where $Z = \left[\varphi(x_1)^T y_1, ..., \varphi(x_N)^T y_N\right]$, $Y = [y_1, ..., y_N]$, $\vec{1} = [1, ..., 1]$, $e = [e_1, ..., e_N]$, $a = [a_1, ..., a_N]$, and I is the identity matrix. The value of w and e are obtained from the solution of Eq. 27, and then Eq. 28 is presented as:

$$\left[\begin{array}{c|c} 0 & Y^T \\ \hline Y & \Omega+\gamma^{-1}I \end{array}\right]\left[\begin{array}{c} b \\ \hline a \end{array}\right]=\left[\begin{array}{c} 0 \\ \hline \vec{1} \end{array}\right] \tag{29}$$

where $\Omega = ZZ^T$. According to the Mercer rule, the kernel function ψ was introduced (Vapnik 1995; Chapelle et al. 1999). Thus, the function Ω was shown as:

$$\Omega_{ij} = y_i y_j \varphi(x_i)^T \varphi(x_j) = y_i y_j \psi(x_i, x_j) \tag{30}$$

Thus, the classifier was developed using the solution of Eqs. 29 and 30, instead of the solution of the relative complex quadratic programming in the traditional SVM (Yu and Cheng 2006). In this study, LSSVM was chosen as classifier because of performance reasons, as it has low complexity but still gives satisfying results on high-dimension feature spaces.

Kernel function is an important factor in LSSVM and influences directly the classification rate. The common examples of kernel function contain linear, polynomial, radial basis function (RBF) kernel, and multi-layer perceptron. RBF kernel is a nonlinear function and a more compacted supported kernel that can reduce the computational complexity of the training procedure compared to other kernels while giving good performance under general smoothness assumptions (Lua et al. 2003). Thus, RBF kernel was used in this study.

Because LSSVM classification model can only classify two classes and there were four grade levels of raisins here, the LSSVM classification model needs to encode and decode a multi-class classification task into multiple binary classifiers using minimum output coding (Van et al. 2002; Wu et al. 2008a). The four grade levels with original labels [1 2 3 4] were encoded in the following codebook (Table 2).

In LSSVM, two important parameters need to be considered. Parameter γ is a regularization parameter for RBF kernel. It determines the trade-off between SRM principle and empirical risk minimization and is important to improve the generalization performance of LSSVM. Parameter σ^2 represents the bandwidth of the RBF kernel. It controls the value of function regression error and influences directly the number of initial eigenvectors. In this study, these two parameters were optimized with values of γ in the range of 2^{-1}–2^{15} and σ^2 in the range of 2^0–2^6 with adequate increments by leave one out cross validation technique (Pelckmans et al. 2003). The optimization was performed on the LSSVM Matlab software, which is downloaded from the website http://www.esat.kuleuven.be/sista/lssvmlab/.

Table 2 Multiple binary classifiers for input set of LSSVM for four grades

Grade level	1	2	3	4
Grade classifiers	−1	−1	1	1
	−1	1	−1	1

Model Efficiency Estimation

In this study, after unsupervised clustering analysis by PCA, supervised pattern recognition methods of LDA, SIMCA, and LSSVM were used to evaluate the classification ability of different feature vectors. Raisin samples were divided randomly into two subsets. One subset was called training set and was used to build a calibration (discriminant) model and the other was called prediction set and was used to test the robustness of the model. The training set contains 200 samples, and each grade has 50 samples. The remaining 280 samples constitute the prediction set, and each grade has 70 samples. To characterize prediction ability (efficiency) of created classification models, correct answer rate (CAR) was used

$$\mathrm{CAR} = \frac{N_{\mathrm{right}}}{N_0} \times 100\% \tag{31}$$

where N_{right} refers to the number of rightly classified samples; N_0 is the total number of samples in prediction set.

Results and Discussion

Feature Extraction and Unsupervised Clustering Analysis

Experiments were conducted with the 480 original raisin images. First, a feature library was created by extracting six color features from the histogram of each channel in both RGB and HSI color spaces and six texture features from the gray level co-occurrence matrix of each channel in both RGB and HSI color spaces. Then, for each color space, we defined three feature vectors from the feature library: color feature vector C_{rgb} (or C_{hsi}), texture feature vector T_{rgb} (or T_{hsi}), and combined color and texture vector CT_{rgb} (or CT_{hsi}), as listed in Table 3. All these six feature vectors were used respectively for unsupervised cluster analysis based on PCA for the 480 raisin samples. The results of PCA clustering are visualized by the scores of the first two PCs in Fig.1a–f. For each feature vector, the accumulative variation of first two principal components was above 92%, which meant that the first two PCs can explain mass variation of the feature vector.

From clustering plots Fig.1a–c, grades 1, 3, and 4 samples were scattered in a wide range and could not be separated clearly; only grade 2 was clustered closely and separated from the others. Therefore, it was difficult to identify all of the samples by using only color feature vector (C_{rgb}), texture feature vector (T_{rgb}) in RGB color space, or color feature vector (C_{hsi}) in HSI color space.

Table 3 Derived feature vectors

Feature vector	Property
C_{rgb}	Color features[a] obtained from R, G, and B channels (defined by Eq. 10)
C_{hsi}	Color features[a] obtained H, S, and I channels (defined by Eq. 10)
T_{rgb}	Texture features[b] obtained from R, G, and B channels (defined by Eq. 21)
T_{hsi}	Texture features[b] obtained from H, S, and I channels (defined by Eq. 21)
CT_{rgb}	Combination of color and texture features obtained from R, G, and B channels (C_{rgb} + T_{rgb}, defined by Eq. 22)
CT_{hsi}	Combination of color and texture features obtained from H, S, and I channels (C_{hsi} + T_{hsi}, defined by Eq. 22)

[a] Six color features such as mean value, variance, skewness, kurtosis, energy, and entropy derived from histogram

[b] Six texture features such as contrast, dissimilarity, homogeneity, angular second moment, entropy, and correlation derived from gray level co-occurrence matrix

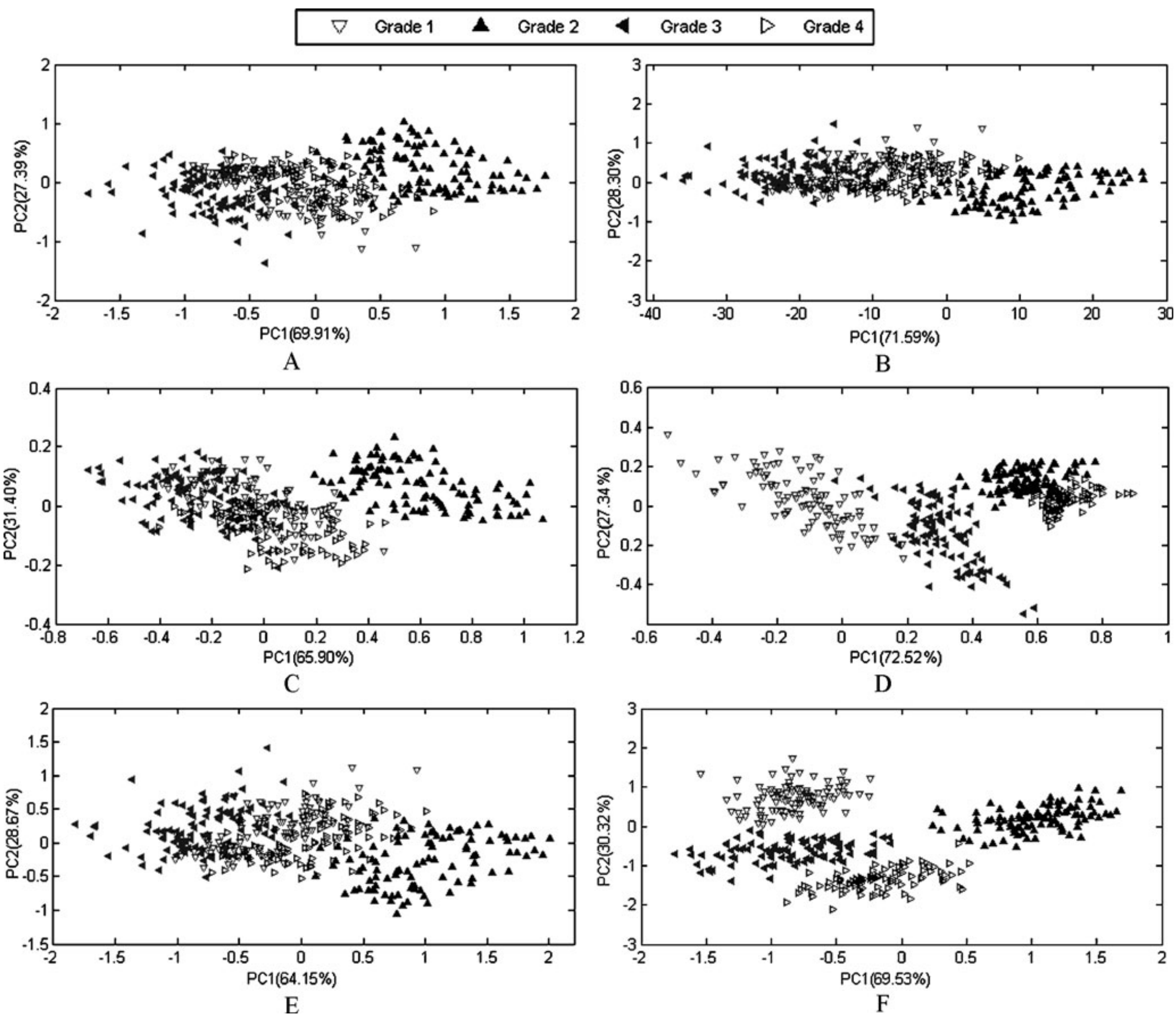

Fig. 1 PCA clustering results using different feature vectors. **a** Result using color features in RGB space(C_{rgb}). **b** Result using color features in HSI space (C_{hsi}). **c** Result using texture features in RGB space (T_{rgb}). **d** Result using texture features in HSI space (T_{hsi}). **e** Result using combined color and texture features in RGB space (CT_{rgb}). **f** Result using combined color and texture features in HSI space (CT_{hsi})

Figure 1d indicates that change has taken place on clustering performance. Samples from different grades of raisins located in concentrative area, and most of them could be separated clearly; however, the boundaries of different grades were not clear and several samples overlapped with samples of other grades. It can be concluded that texture feature vector (T_{hsi}) from HSI color space can improve the clustering performance, but cannot adequately discriminate these raisin grades.

In Fig.1e, grades 1, 3, and 4 samples gathered more together comparing with that in Fig.1a and c. However, the boundaries of different grades were not clear. Thus, combined color and feature vector (CT_{rgb}) in RGB color space provided a better clustering performance than any of the separate color feature vector (C_{rgb}) or texture feature vector (T_{rgb}), but this result was not sufficient.

In Fig.1f, each of four grades was clustered closely and separated apparently from the others. Comparing Fig.1a–e, it can be found that samples of each grade gather more together and the boundaries of different raisin grades are clearer in Fig.1f. Examination of PC loadings can be used to identify the important features underlying clustering or discrimination. According to the loadings of first principal component (PC1) for feature vector CT_{hsi} (Fig.2), the most discriminant features in CT_{hsi} were found to be the texture features of homogeneity (HO), angular second moment (ASM), dissimilarity (DI) in channel I, the color feature of skewness (μ_3) in channel S, and the color feature of entropy (h) in channel H.

Based on the results of unsupervised clustering analysis, we may safely come to the conclusion that by using combined color and texture feature vector (CT_{hsi}) in HSI color space; significant performance increase can be achieved compared to any of the other feature vectors.

Classification by LDA and SIMCA Using Different Features

In this work, LDA and SIMCA were used to evaluate the classification ability of different feature vectors. The first step was to compare classification results by LDA based on different feature vectors. Feature vectors C_{rgb}, C_{hsi}, T_{rgb}, T_{hsi}, CT_{rgb}, and CT_{hsi} of raisin samples in training set were set as the input variables for LDA to build discrimination models. Classification performance of each discrimination model was evaluated by the other samples in prediction set. Classification results described as correct answer rate are shown in Table 4.

In the next step, SIMCA was used to classify raisins. For each feature vector, four SIMCA calibration models were established based on four different grades of raisin samples in training set. The optimal number of principal components (PC) ranged from 1 to 12 was found for SIMCA method. The classification performance of each calibration model was evaluated by the other samples in prediction set. Classification results described as correct answer rate by SIMCA are shown in Table 4.

The LDA and SIMCA classification results (Table 4) suggest that combined color and texture features have obtained the best results of classification (with about 92.87% of average correct answer rate by LDA using CT_{hsi}, and about 92.50% of average correct answer rate by SIMCA using CT_{rgb}), which were better than that of solely used color or texture features, but the CAR values were not high enough for highly accurate industry application. As further analysis, LSSVM was used to find optimal set of color and texture features, which gave the best result for raisin quality classification, and to evaluate if it can improve classification accuracy.

Classification by LSSVM Using Separate Color Features

LSSVM models were constructed for raisin quality classification. In the first step, color feature vector C_{rgb} and C_{hsi} were respectively set as the input variables for LSSVM models. The optimal combination of (γ, σ^2) was achieved with γ=180.572, σ^2=27.833 for LSSVM model of C_{rgb} and γ=1.005, σ^2=5.840 for LSSVM model of C_{hsi}, respectively. The prediction results by color features described as correction answer rate in percents are shown in Table 5.

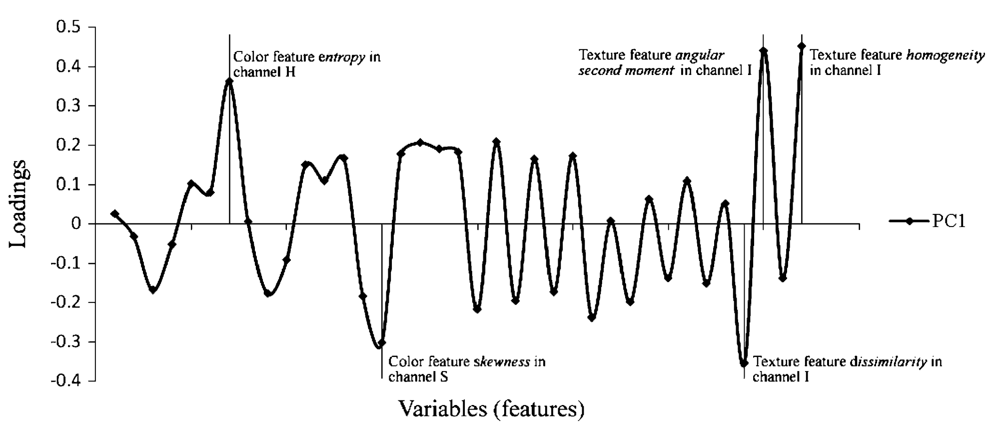

Fig. 2 Loadings of first principal component (PC1) for feature vector CT_{hsi}

Table 4 Prediction results of LDA and SIMCA models for raisin classification based on different features

Features	Classification results (CAR, %)									
	LDA					SIMCA				
	Grade 1	Grade 2	Grade 3	Grade 4	Average	Grade 1	Grade 2	Grade 3	Grade 4	Average
C_{rgb}	78.57	95.71	90.00	91.43	88.93	87.14	88.57	85.71	87.14	87.14
C_{hsi}	90.00	91.43	97.14	78.57	89.29	92.86	90.00	91.43	78.57	88.21
T_{rgb}	94.29	91.43	85.71	75.71	86.79	88.57	94.29	88.57	87.14	89.64
T_{hsi}	94.29	94.29	90.00	88.57	91.79	92.86	94.29	87.14	90.00	91.07
CT_{rgb}	85.71	92.86	90.00	88.57	89.29	91.43	94.29	92.86	91.43	92.50
CT_{hsi}	92.90	94.29	97.14	87.14	92.87	94.29	94.29	90.00	90.00	92.15

Table 5 indicates that no remarkable change has taken place on the classification performances between color features computed from RGB color space (C_{rgb}) and from HSI color space (C_{hsi}). Their average correct answer rates are both under 90% (86.79% of C_{rgb} and 88.93% of C_{hsi}), showing that separate color feature did not give sufficient classification result.

Classification by LSSVM Using Separate Texture Features

The second step was to compare classification results based on different texture features in order to obtain the best feature set. Here we set texture feature vectors T_{rgb} and T_{hsi} respectively as the input data for LSSVM. The optimal combination of (γ, σ^2) was achieved with γ=37.118, σ^2= 10.445 for LSSVM model of T_{rgb} and γ=35.061, σ^2= 12.461 for LSSVM model of T_{hsi}, respectively. The prediction results by texture features described as correct answer rate in percents are shown in Table 6.

From Table 6, it can be found that the best discrimination performance by texture feature was achieved with the texture feature in HSI color space (T_{hsi}, 92.86%). The performance increased compared to the T_{rgb} (88.57%) over 4 percentage units, the C_{rgb} (86.79%) over 5 percentage units, and the C_{hsi} (88.93%) over 3 percentage units, showing that separate texture feature in HSI color space can give comparatively better result for raisin classification.

Classification by LSSVM Using Combined Color and Texture Features

The separate texture feature in HSI color space (T_{hsi}) can give reasonable result for raisin classification. However, it ignored the contribution of color information to the classification. In the third step, we attempted combined color and texture feature vectors (CT_{rgb}, CT_{hsi}) to find out whether or not the performance of separate color or texture feature could be improved by combining them. The optimal combination of (γ, σ^2) was achieved with γ=275.906, σ^2= 23.488 for LSSVM model of CT_{rgb} and γ=18.654, σ^2= 25.063 for LSSVM model of CT_{hsi}, respectively. The prediction results by combined color and texture feature vectors described as correct answer rate in percents are shown in Table 7.

Table 7 indicates that the best result of classification (with about 95% of accuracy) by LSSVM has been

Table 5 Prediction results of LSSVM models for raisin classification based on different color features (C_{rgb}, C_{hsi})

Features	Grade level	Sample number	Classification results by LSSVM				CAR (%)
			Grade 1	Grade 2	Grade 3	Grade 4	
C_{rgb}	Grade 1	70	59	1	2	8	84.29
	Grade 2	70	2	66	0	2	94.29
	Grade 3	70	5	0	62	3	88.57
	Grade 4	70	7	4	3	56	80.00
	Average						86.79
C_{hsi}	Grade 1	70	58	0	7	5	82.86
	Grade 2	70	0	68	0	2	97.14
	Grade 3	70	4	0	63	3	90.00
	Grade 4	70	5	1	4	60	85.71
	Average						88.93

Table 6 Prediction results of LSSVM models for raisin classification based on different texture features (T_{rgb}, T_{hsi})

Features	Grade level	Sample number	Classification results by LSSVM				CAR (%)
			Grade 1	Grade 2	Grade 3	Grade 4	
T_{rgb}	Grade 1	70	59	1	4	6	84.29
	Grade 2	70	1	67	0	2	95.71
	Grade 3	70	6	2	58	4	82.86
	Grade 4	70	2	1	3	64	91.43
	Average						88.57
T_{hsi}	Grade 1	70	68	0	2	0	97.14
	Grade 2	70	0	63	2	5	90.00
	Grade 3	70	3	2	65	0	92.86
	Grade 4	70	0	6	0	64	91.43
	Average						92.86

obtained with the combination of color and texture features extracted from HSI color space (CT_{hsi}). By comparing the performance with that of color features (Table 5) or texture features (Table 6) used alone, it could be found that combined color and texture features increase accuracy of classification in both RGB color space and HSI color space. Moreover, the best classification result with average correct answer rate of 95% was reached by LSSVM using CT_{hsi}, better than that of LDA (with about 92.87% of average correct answer rate using CT_{hsi}, seen in Table 4) and SIMCA (with about 92.50% of average correct answer rate using CT_{rgb}, seen in Table 4), showing that LSSVM can improve classification accuracy for raisin classification.

Conclusions

The results indicated that color or texture information alone was not sufficient for accurate raisin quality classification. But with the combination of color and texture information (CT_{hsi}) we have achieved high classifying process accuracy by LSSVM classifier. The best classification result (with about 95% of average correct answer rate) by LSSVM was better than that of LDA (with about 92.87% of average correct answer rate) and SIMCA (with about 92.50% of average correct answer rate). It can be concluded that combined color and texture features in HSI color space coupled with a LSSVM classifier can be an accurate and efficient raisin quality classification tool. The result of this study is helpful for online monitoring of the raisin quality sorting/grading process. LSSVM supplied a highly accurate way for raisin classification, whereas its complex calculation should be considered for further application in raisin industry. Therefore, further investigations are required to build LSSVM model based on effective variables (the most discriminant features in the feature vector CT_{hsi}) instead of full variables, which can be used to reduce the complexity and time cost of LSSVM calculation and meet industry requirements, as well as expand the grades number of raisin, optimize the image process algorithm, and improve the model's robustness and strictness before its online application.

Table 7 Prediction results of LSSVM models for raisin classification based on different combined color and texture features (CT_{rgb}, CT_{hsi})

Features	Grade level	Sample number	Classification results by LSSVM				CAR (%)
			Grade 1	Grade 2	Grade 3	Grade 4	
CT_{rgb}	Grade 1	70	60	2	3	5	85.71
	Grade 2	70	1	67	0	2	95.71
	Grade 3	70	5	0	61	4	87.14
	Grade 4	70	2	2	4	62	88.57
	Average						89.29
CT_{hsi}	Grade 1	70	68	0	2	0	97.14
	Grade 2	70	0	68	0	2	97.14
	Grade 3	70	2	0	65	3	92.86
	Grade 4	70	0	2	3	65	92.86
	Average						95.00

Acknowledgments This study was supported by 863 National High-Tech Research and Development Plan (2007AA10Z210), National Agricultural Science and Technology Achievements Transformation Fund Programs (2009 GB23600517), Zhejiang Provincial Natural Science Foundation of China (Z3090295), National Natural Science Foundation of China (10831007), and Ningbo Natural Science Foundation of China (2010A610015).

References

Abbasgolipour, M., Omid, M., Keyhani, A., & Mohtasebi, S. S. (2010). Sorting raisins by machine vision system. *Modern Applied Science, 4*(2), 49–60.

Belousov, A. I., Verzakov, S. A., & Von Frese, J. (2002). A flexible classification approach with optimal generalisation performance: Support vector machines. *Chemometrics and Intelligent Laboratory Systems, 64*(1), 15–25.

Burks, T. F., Shearer, S. A., & Payne, F. A. (2000). Classification of weed species using color texture features and discriminant analysis. *Transactions of the ASABE, 43*(2), 441–448.

Chapelle, O., Haffner, P., & Vapnik, V. (1999). SVMs for histogram-based image classification. *IEEE Transactions on Neural Networks, 10*(5), 1055–1064.

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). *Spectrochim Acta Part A, 66*(3), 568–574.

Christensen, L. P. (2000). Raisin Quality. In Christensen (Ed.), *Raisin production manual* (pp. 228–235). Oakland: University of California, Agricultural and Natural Resources Publication 3393.

Cristianini, N., & Shawe, T. J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. New York: Cambridge University Press.

Cui YY, & Zeng ZY (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, 03–05 April 2009, Kuala Lumpar, Malaysia (CD-ROM).

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*. New York: Academic.

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.

Karush, W. (1939). *Minima of functions of several variables with inequalities as side constraints*. Chicago: Department of Mathematics, University of Chicago. M.Sc. Dissertation.

Kuhn, H. W., & Tucker, A. W. (1951). Nonlinear programming. In N. Jerzy et al. (Eds.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (pp. 481–492). Berkeley: University of California Press.

Kumar, S., & Mittal, G. S. (2010). Rapid detection of microorganisms using image processing parameters and neural network. *Food and Bioprocess Technology, 3*(5), 741–751.

Li X.L., Yuan J., Gu T.X., & Liu X.Y. (2009) Level detection of raisins based on image analysis and neural network. In: Proceedings of the 6th International Symposium on Neural Networks (pp 343–350). Berlin, Heidelberg: Springer

Lua, C., VanGestel, T., Suykens, J. A. K., VanHuffel, S., Vergote, I., & Timmerman, D. (2003). Preoperative prediction of malignancy of ovarian tumors using least squares support vector machines. *Artificial Intelligent in Medicine, 28*(3), 28–306.

Michie, D., Spiegelhalter, D.J., & Taylor, C.C. (1994) Machine learning, neural and statistical classification. Ellis Horwood, Chichester.

Milde, L. B., Ramallo, L. A., & Puppo, M. C. (2010). Gluten-free bread based on tapioca starch: texture and sensory studies. *Food and Bioprocess Technology*. doi:10.1007/s11947-010-0381-x.

Moon, H., & Phillips P. (1998). Analysis of PCA-based face recognition algorithms. In Bowyer KW, & Phillips PJ (eds.), Empirical Evaluation Techniques in Computer Vision (pp. 57–71). Wiley-IEEE Computer Society.

Okamura, N. K., Delwiche, M. J., & Thompson, J. F. (1993). Raisin Grading by Machine Vision. *Transactions of the American Society of Agricultural Engineers, 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.

Pelckmans, K., Suykens, J.A.K., & Van Gestel, T. (2003). LS-SVM lab Toolbox User's Guide (version 1.5), Available at: http://www.est.kuleuve n.ac.be/sista/lssvmlab/tutorial. Accessed April 13, 2010.

Purcell, D. E., O'Shea, M. G., & Kokot, S. (2007). Role of chemometrics for At-field application of NIR spectroscopy to predict sugarcane clonal performance. *Chemometrics and Intelligent Laboratory Systems, 87*(1), 113–124.

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 Gestel, T., De Brabanter, J., De Moor, B., & Vandewalle, J. (2002). *Least squares support vector machines*. Singapore: World Scientific Publishing.

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: usda.mannlib.cornell.edu/usda/fas/raiswm//2000s/2009/. Accessed April 13, 2010.

Van, G. T., Suykens, J., Lanckriet, G., Lambrechts, A., De, M. B., & Vandewalle, J. (2002). Multiclass LS-SVMs: Moderated outputs and coding–decoding schemes. *Neural Processing Letters, 15*(1), 45–48.

Vapnik, V. (1995). *The native of statistical learning theory*. New York: Springer.

Vapnik, V. (1998). *Statistical learning theory*. New York: John Wiley and Sons Inc.

Williamson, G., & Carughi, A. (2010). Polyphenol content and health benefits of raisins. *Nutrition Research, 30*(8), 511–519.

Wold, S., & Sjostrom, M. (1977). SIMCA: A method for analyzing chemical data in terms of similarity and analogy. In B. R. Kowalski (Ed.), *Chemometrics Theory and Application, American Chemical Society Symposium Series 52* (pp. 243–282). Washington, DC: American Chemical Society.

Wu, D., Feng, L., He, Y., & Bao, Y. (2008a). Variety identification of Chinese cabbage seeds using visible and near-infrared spectroscopy. *Transactions of the ASABE, 51*(6), 2193–2199.

Wu, D., Feng, L., Zhang, C., & He, Y. (2008b). 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., Chen, X. J., & He, Y. (2009). Application of multispectral image texture to discriminating tea categories based on DCT and LS-SVM. *Spectroscopy and Spectral Analysis, 29*(5), 1382–1385.

Yi, H., Zhang, L., Hua, C., Sun, K., & Zhang, L. (2010). Extraction and enzymatic hydrolysis of inulin from jerusalem artichoke and their effects on textural and sensorial characteristics of yogurt. *Food and Bioprocess Technology, 3*(2), 315–319.

Yu, K., & Cheng, Y. Y. (2006). Discriminating the genuineness of Chinese medicines using least squares support vector machines. *Chinese Journal of Analytical Chemistry, 34*(4), 561–564.

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.