Available online at www.sciencedirect.com
ScienceDirect
Journal of Food Engineering 81 (2007) 98–107
JOURNAL OF FOOD ENGINEERING
www.elsevier.com/locate/jfoodeng

Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry

Gamal ElMasry [a,b], Ning Wang [a,*], Adel ElSayed [b], Michael Ngadi [a]

[a] Department of Bioresource Engineering, McGill University, Macdonald Campus, 21,111 Lakeshore Road, Ste-Anne-de-Bellevue, Que, Canada H9X 3V9
[b] Faculty of Agriculture, Suez Canal University, Ismailia, Egypt

Received 15 June 2006; received in revised form 18 October 2006; accepted 19 October 2006
Available online 5 December 2006

Abstract

Hyperspectral imaging in the visible and near-infrared (400–1000 nm) regions was tested for nondestructive determination of moisture content (MC), total soluble solids (TSS), and acidity (expressed as pH) in strawberry. The spectral data were analyzed using the partial least squares (PLS) analysis, a multivariate calibration technique. The correlation coefficients (r) with the whole spectral range (400–1000 nm) for predicting MC, TSS, and pH were 0.90, 0.80, and 0.87 with SEC of 6.085, 0.233, and 0.105 and SEP of 3.874, 0.184, and 0.129, respectively. Optimal wavelengths were selected using β-coefficients from PLS models. Multiple linear regression (MLR) models were established using only the optimal wavelengths to predict the quality attributes. The correlation coefficients (r) for predicting MC, TSS, and pH using MLR models were 0.87, 0.80, and 0.92 with SEC of 6.72, 0.220, and 0.084 and SEP of 5.786, 0.211, and 0.091, respectively. Moreover, for classifying strawberry based on ripeness stage, a texture analysis was conducted on the images based on grey-level co-occurrence matrix (GLCM). The higher classification accuracy of 89.61% was achieved using the GLCM parameters at horizontal direction at angle of 0°.

Keywords: Hyperspectral imaging; Strawberry; Partial least square; Texture analysis; Grey-level co-occurrence matrix

1. Introduction

Fruit quality is defined as a measure of characters or attributes that determine the suitability of the fruit to be eaten as fresh or stored for reasonable period without deterioration. Fruit quality could be considered as a multiple concept encompassing the physical, physiological, nutritional, and pathological attributes that affect fruit shelf life. Quality of a fresh produce includes appearance (size, shape, colour, gloss, and freedom from defects and decay), texture (firmness, crispness, and toughness), flavor (sweetness, sourness, aroma, and off-flavors), and nutritive value (vitamins, minerals, nutrients and carbohydrates). Strawberry (*Fragaria* sp.) is one of the economically important fruits which are more popularly eaten fresh, used for garnishing cakes and pastries, flavored for juices and milk products, and processed into jams and others. Thus, together with the recent concern for food quality and safety, automatic technologies for judging the fresh quality of strawberry are being sought.

Technologies that can sort fruit for appearance, texture, taste, flavor and/or nutritive value would assure fruit quality and consistency, increase consumer confidence and satisfaction, and enhance the competitiveness and profitability of the fruit industry (Lu & Ariana, 2002). Currently, fruits are sorted manually or automatically on the basis of their external quality features. However, internal quality attributes such as dry matter content, total soluble solids content, sugar content, and juice acidity are very important in the modern quality evaluation industries. Most instrumental techniques to measure these properties are destructive and involve a considerable amount of manual work. The development of nondestructive measurements of these

* Corresponding author. Tel.: +1 514 398 7781; fax: +1 514 583 8711.
E-mail address: ning.wang@mcgill.ca (N. Wang).

0260-8774/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jfoodeng.2006.10.016

quality attributes will be very useful for producers, processors, and distributors to ascertain fast evaluation. This trend put a great challenge to utilize spectroscopic and hyperspectral imaging for these tasks.

Spectroscopic and hyperspectral imaging systems have many advantages compared to classical chemical and physical analytical methods. It has a short measuring time with limited sample preparation, is chemical-free, and can be applied to estimate more than one attribute at the same time (Lammertyn, Nicolaï, Ooms, De Smedt, & De Baerdemaeker, 1998). All these factors reduce energy requirements and costs of process, and provide more consistent fruit to consumers. The spectroscopic method has a great drawback compared with the hyperspectral imaging because it acquires the spectral data from a single point or from a small portion of the tested fruit. The hyperspectral imaging, on the contrary, has advantages of receiving spatially distributed spectral responses at each pixel of a fruit image.

The hyperspectral imaging technique has been implemented in several applications such as inspection of poultry carcasses (Chao, Chen, Hruschka, & Park, 2001; Park, Windham, Lawrence, & Smith, 2004), defect detection or quality determination on apples, eggplants, pears, cucumber and tomatoes (Cheng et al., 2004; Kim et al., 2002; Liu, Chen, Wang, Chan, & Kim, 2006; Li, Wang, & Gu, 2002; Polder, Van der Heijden, & Young, 2002) as well as physical, chemical and mechanical properties estimation in various commodities (Lu, 2004; Nagata, Tallada, Kobayashi, & Toyoda, 2005). In addition, a significant amount of research has been done in the area of spectroscopy and hyperspectral imaging applied specially to fruit analysis. The successful attempts to evaluate internal properties nondestructively were accomplished using spectral technology for prediction sugar content (Bellon, Vigneau, & Leclercq, 1993), soluble solids (Park, Abbott, Lee, Choi, & Choi, 2003; Peiris, Dull, Leffler, & Kays, 1999), firmness (Park et al., 2003; Peirs, Scheerlinck, Touchant, & Nicolaï, 2002), moisture content (Katayama, Komaki, & Tamiya, 1996), acidity (Lammertyn et al., 1998; Peirs et al., 2002) and so many other applications. In strawberries, however, there had been very limited literature on the use of spectroscopic technique for quality estimation. A preliminary work was done by Ito (2002) to determine soluble solids in the NIR range of 750–1100 nm. Nagata, Tallada, Kobayashi, Cui, and Gejima (2004) conducted a hyperspectral investigation for soluble solids and firmness prediction in strawberry in the visible range (450–650 nm).

The great power of hyperspectral imaging resides in its capability to deal with spectral as well as spatial information. To exploit this capability, researchers tend to implement their algorithms to deal with both information at the same level of interest. One of the most important applications of this possibility is the classification processes according to certain external characteristics, which has been one of the targets in image processing applications. However, it is important to differentiate between the external texture which used to express roughness or smoothness of fruit surface in an image and the real texture of the fruit which used to express the mechanical properties and toughness or firmness of the fruit. External texture has been used for the analysis of different types of images, such as microscopic, aerial and satellite. Textural features represent the spatial distribution of tonal variations in an image at various wavelengths such as the visible and infrared portions of the spectrum (Kavdır & Guyer, 2004). The grey-level of co-occurrence matrix (GLCM) is an important expression to deal with external textural change in agricultural produces. The co-occurrence matrix has been successfully used to classify texture of agricultural produces (Létal, Jirák, Šuderlová, & Hájek, 2003; Bharati, Liu, & MacGregor, 2004; Pydipati, Burks, & Lee, 2006). The texture analysis for identifying roughness and ripeness stage of strawberry will be one of the scopes of this work.

2. Objectives

The main aim of this research was to develop calibration models based on hyperspectral imaging to estimate some quality attributes of strawberry. This outmost goal was achieved by meeting the following specific objectives:

(i) Developing partial least square (PLS) models to quantitatively predict moisture content (MC), total soluble solids (TSS), and pH in strawberry.
(ii) Selecting the lowest number of optimal wavelengths which give the highest correlation between the spectral data and the three quality parameters.
(iii) Developing multiple regression models using spectral responses from only the optimal wavelengths and then test and validate the prediction accuracy of the developed calibration models.
(iv) Identifying ripeness stage of strawberry by texture analysis using grey-level co-occurrence matrix (GLCM) parameters.

3. Material and methods

3.1. Fruit samples

Strawberry was obtained from local retail stores. Good appearance of the tested fruits is essential for the experiments. All abnormal fruits were discarded. Seventy-seven fruits free from any abnormal features such as defects, bruises, diseases, and contaminations were selected. All green calyxes were completely removed from the tested fruits. To generate fruits with wide variations in their properties, some fruits were kept in room temperature for 2 days and other for 3 days, while other fruits were stored in cold temperature (5 °C) for 3 days and other for 5 days. The resulting fruits were varied in their ripeness and their quality attributes. The tested fruits are randomly divided into two major classes. Class 1 consists of 60 fruits used

as a calibration/training set for developing the PLS model. Class 2 consists of 16 fruits used for model validation and to verify the prediction power of the calibration models.

3.2. Hyperspectral imaging system

A laboratory hyperspectral imaging system was constructed as shown in Fig. 1. It is composed of the following four components: (1) an illumination unit which consists of two 50 W halogen lamps adjusted at angle of 45° to illuminate the camera's field of view, (2) a fruit holder surrounded by a cube tent made from white nylon to diffuse the light and provide a optimum lighting condition, (3) a spectrograph (ImSpector V10E, Optikon Co., Canada) coupled with a standard C-mount zoom lens, and (4) a CCD camera (PCO-1600, Pco. Imaging, Germany). The assembly disperses the incoming line of light into the spectral and spatial matrices and then projects them onto the CCD. The optics, spectrograph and the camera, has high sensitivity from 400 to 1000 nm and the exposure time was adjusted at 200 ms throughout the whole test. The distance between the lens and the surface of the strawberry being imaged was fixed at 40 cm. After finishing the scans on a fruit, a three-dimensional (x, y, λ) spatial and spectral data space was constructed. Images were binned during acquisition in spatial direction to provide images with spatial dimension of 400×400 pixels with 826 spectral bands from 400 to 1000 nm. The hyperspectral imaging system was controlled by a PC with a software called Hypervisual Imaging Analyzer® (ProVision Technologies, Stennis Space Center, USA) for image acquisition.

3.3. Preprocessing of hyperspectral images

All the acquired hyperspectral images were processed and analyzed using Environment for Visualizing Images (ENVI 4.2) software (Research Systems Inc., Boulder Co., USA).

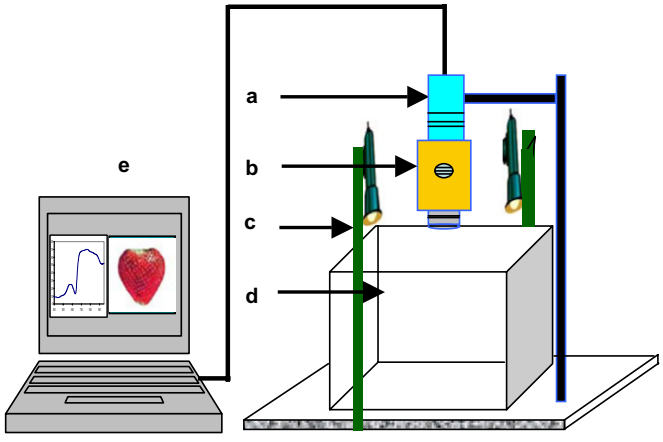

Fig. 1. Hyperspectral imaging system: (a) a CCD camera, (b) a spectrograph with a standard C-mount zoom lens, (c) a halogen lighting unit, (d) a white nylon tent, and (e) a PC supported with an image acquisition software.

The hyperspectral images were firstly corrected with a white and a dark reference. The dark reference was used to remove the effect of dark current of the CCD detectors, which are thermally sensitive. The corrected image (R) is estimated using Eq. (1):

$$R = \frac{R_0 - D}{W - D} \times 100 \quad (1)$$

where R_0 is the recorded hyperspectral image, D the dark image (with 0% reflectance) recorded by turning off the lighting source with the lens of the camera completely closed, and W is the white reference image (Teflon white board with 99% reflectance).

The corrected images will be the basis for the subsequent image analysis to extract information about the spectral properties of each fruit for optimizing surface characteristics identification, selection of effective wavelengths and texture analysis purposes.

4. Data analysis

4.1. Spectral analysis for predicting quality attributes

4.1.1. Recording spectral data

Due to physiological and chemical variations among tested fruits, each fruit reflects, absorbs, and emits electromagnetic energy in distinctive patterns at specific wavelengths. In essence, these spectral responses can be used to uniquely characterize and identify the fruits. To collect the spectral response of each fruit, a binary mask was first created to produce an image containing only the fruit in the middle of the image, avoiding any interference from the background. Image at 500 nm-bands was picked up for this task because the fruit appeared opaque compared with the background and can be segmented easily by simple thresholding. All active pixels in the segmented image were used as a mask and act as an area of interest (AOI). At each pixel of AOI, the relative reflectance was recorded at each wavelength from 400 to 1000 nm. Then, an average reflectance spectrum was determined by averaging the relative reflectance spectra of all pixels in the AOI. In total, 76 average spectra (400–1000 nm) representing all the tested 76 fruits were recorded and stored in the PC for wavelengths selection and calibration model development.

4.1.2. Measurement of quality parameters

Three attributes of each fruit (moisture content, total soluble solids and pH) were measured and used as indicators of fruit quality. After acquiring the spectral images, each fruit was divided into two equal halves. One half was used for moisture content determination and the other half was juiced using a juicing machine to determine pH and total soluble solids.

The moisture content was determined in the first half by oven-drying method using forced-air convection oven

(Model no. OV-5206-2, Blue M Electric Company, Blue Island, IL, USA); meanwhile, the pH of the strawberry juice was determined with a pH meter (Orion, model 250A, Beverly, MA, USA), and total soluble solids (°Brix) was determined using a refractometer (Model no. ATC-1090, Atago Co., Ltd., Japan).

4.1.3. Spectral data analysis and building the calibration models

To develop a model between spectral responses of the tested fruits and their quality attributes, partial least squares (PLS) analysis was applied to build the model of prediction. PLS analysis between one attribute (moisture content, soluble solids or pH) and the spectral data (average spectra with 826 wavelengths in the range from 400 to 1000 nm) was conducted using GRAMS/AI (Thermo Electron Corp., Salem, New Hampshire, USA). Generally, PLS is implemented to transfer a large set of highly correlated and often collinear experimental data (826 wavelengths or variables) into independent latent variables or factors. When applied to spectra of the calibration set (60 fruits), the aim of PLS analysis is to find a mathematical relationship between a set of independent variables, X matrix ($N_{60\,\text{fruits}} \times K_{826\,\text{wavelengths}}$), and the dependent variable, Y matrix ($N_{60\,\text{fruits}} \times 1$). The values of one attribute (moisture content, soluble solids or pH) of the calibration set were used to represent the dependent variable (Y). Meanwhile, the reflectance values at 826 wavelengths of the 60 fruits represented the independent variables or the predictors (X). Typically, most of the variations can be captured within the first few latent variables/factors while the remaining latent variables describe random noise or linear dependencies between the wavelengths/predictors.

The PLS algorithm according to Geladi and Kowalski (1986a, 1986b), Haaland and Thomas (1988a, 1988b), Osborne, Jordan, and Künnemeyera (1997) determines a set of orthogonal projection axes W, called PLS-weights, and wavelength scores T. Then, regression coefficients β are obtained by regressing Y onto the wavelength scores T as follow:

$$\widehat{Y} = XW^*_a\beta = T_a\beta \tag{2}$$

$$W^* = (W(P' * W)^{-1}) \tag{3}$$

where $\widehat{Y}$ is the predicted value of the attribute of interest, a is the number of PLS factors, and P' is the wavelength loadings. PLS analysis yielded the calibration model, the optimal number of latent factors, and the predicted value of the attribute under concern in each sample in the calibration set.

The input spectra were preprocessed using mean-centering and automatic baseline correction. Other preprocessing treatments such as Savitzky–Golay smoothing, multiplicative scatter correction (MSC), and first and second derivatives were tested, but did not enhance the calibration model. Mean-centering is a process of calculating the average spectrum of the training set spectra (60 spectra) and then subtracting the average spectrum from each individual spectrum. Same procedure was performed on the measured values of each attribute. Automatic baseline correction was chosen to remove the baseline effects from the spectra occurred during spectral collection. No outliers were detected using both spectral and concentration residuals. The optimal number of latent factors for establishing the calibration model was determined using the minimum value of predicted residual error sum of squares (PRESS). The quality of the calibration model was evaluated by the standard error of calibration (SEC), standard error of prediction (SEP) and the correlation coefficient (r) between the predicted and measured value of the attribute. A good model should have a low SEC, a low SEP, a high correlation coefficient, and a small difference between SEC and SEP. These criteria are defined as follows:

$$\text{SEC} = \sqrt{\frac{1}{I_C - 1}\sum_{i=1}^{I_C}(\hat{y}_i - y_i)^2} \tag{4}$$

$$\text{SEP} = \sqrt{\frac{1}{I_p - 1}\sum_{i=1}^{I_p}(\hat{y}_i - y_i - \text{bias})^2} \tag{5}$$

$$\text{bias} = \frac{1}{I_p}\sum_{i=1}^{I_p}(\hat{y}_i - y_i) \tag{6}$$

where $\hat{y}_i$, predicted value of an attribute in fruit number i; y_i, measured value of an attribute in fruit number i; I_C, number of fruits (spectra) in the calibration set (60); and I_p, number of fruits (spectra) in the validation set (17).

4.1.4. Selection of optimal wavelengths and multiple regression model development

Researchers are often interested in finding vital few wavelengths that would be most influential on the quality evaluation of the product and eliminate wavelengths having no discrimination power. The selected wavelengths reduce the data dimensionality while preserving the most important information contained in the lower dimensional data space. These selected wavelengths depend on the behaviour of spectral responses of the fruits under study and the differences among them. Therefore, to establish consistent multispectral imaging systems, several essential spectral bands are first sought through a variety of strategies, such as general visual inspection of the spectral curves and correlation coefficients (Keskin, Dodd, Han, & Khalilian, 2004), analysis of spectral differences from the average spectrum (Liu, Windham, Lawrence, & Park, 2003), stepwise regression (Chong & Jun, 2005), principal component analysis (Kim et al., 2002; Mehl, Chen, Kim, & Chan, 2004; Xing & De Baerdemaeker, 2005), and other techniques (Hruschka, 2001).

After proving the good performance of the PLS models on predicting the attributes of interest, the next step is to reduce the dimensionality of the spectral data to its lowest level by selecting only the essential wavelengths that carry

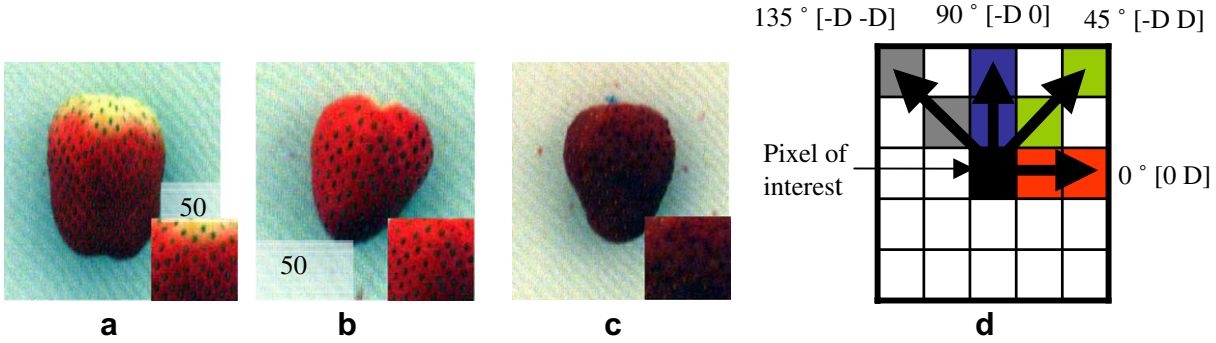

Fig. 2. RGB image for (a) unripe, (b) ripe, and (c) overripe strawberries. The cropped small square for texture analysis is depicted at the bottom left corner of each image. (d) Extraction of GLCM at different directions (0°, 45°, 90°, and 135°) and distance (D) for each pixel in the cropped square image.

the maximum spectral information. In this study, β-coefficients resulting from the best PLS calibration model were used for identifying the optimal wavelengths. The wavelengths that correspond to the highest absolute values of β-coefficients were considered optimal wavelengths. These optimal wavelengths that carry the maximum spectral information could be implemented in multispectral imaging in further studies for on-line applications. Wavelengths which correspond to the lowest absolute values of β-coefficients were completely neglected because they have no or very little contribution in prediction.

Only the selected optimal wavelengths were used to establish multiple linear regression (MLR) models instead of using the whole spectral range. This step was conducted using MATLAB (MATLAB 7.0, Release 14, The MathWorks Inc., Natick, MA, USA) to produce a model as

$$\widehat{Y} = a_0 + \sum_{k=1}^{K} a_k R_{\lambda k}$$

where $\widehat{Y}$, predicted value of the attribute; K, number of optimal wavelengths (number of X-variables or predictors); a_0, a_k, regression coefficients, and $R_{\lambda k}$, reflectance at a wavelength λ corresponding to the kth term in the model.

Once the linear regression model was determined, the equations were used to predict the attributes of samples in the calibration and validation sets. Actual values of the attribute were plotted to visually evaluate the performance of the models. To quantify the predictive ability of the models, the correlation coefficient (r), SEC, and SEP were determined.

4.2. Texture analysis for identifying ripeness stage

Due to large variations found in fruits in terms of surface textures, texture analysis was conducted to identify the ripeness stage of strawberry. As shown in Fig. 2a–c, the tested fruits contained three ripeness categories: unripe, ripe and overripe. In this work, the fruits containing equal or more than 30% green area were classified as unripe fruits. RGB image was constructed for each fruit by picking up the Red (650 nm), Green (500 nm) and Blue (450 nm) bands from the corrected hyperspectral image to form a colour image.[1] Small window (50 × 50 pixels) as shown in the bottom left corner of Fig. 2 was cropped from the middle of each colour image to discover the texture change from fruit to fruit. Texture analysis characterizes regions in an image by their texture content in terms of smoothness, roughness, silkiness, or bumpiness in the context of an image. In this case, these characteristics refer to variations in the brightness values, or grey levels. The most commonly used texture measures are derived from the grey-level co-occurrence matrix (GLCM). A co-occurrence matrix is a square matrix with elements corresponding to the relative frequency ($P_{i,j}$) of occurrence of pairs of grey level of pixels separated by a certain distance (D) in a given direction (0°, 45°, 90°, or 135°) as shown in Fig. 2d. Each entry (i,j) in GLCM corresponds to the number of occurrences of the pair of grey levels i and j which are a distance D apart in the image. The following GLCM parameters were calculated using a program developed by MATLAB to express texture:

$$\text{Contrast} = \sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2$$

$$\text{Homogenity} = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2}$$

$$\text{Angular second moment} = \sum_{i,j=0}^{N-1} P_{i,j}^2$$

$$\text{Correlation} = \frac{\sum_{i,j=0}^{N-1}(ij)P_{i,j} - \mu_x\mu_y}{\sigma_x\sigma_y}$$

where μ_x, μ_y, σ_x, and σ_y are the means and standard deviations, respectively, of the sums of rows and columns in the GLCM matrix.

The textural parameters mentioned above were first described by Haralick and Shanmugam (1974) used in spectral and spatial image analysis for remote sensing

[1] For interpretation of colour in Fig. 2, the reader is referred to the web version of this article.

applications, and then these parameters were widely used in texture classification in agricultural produces by Tsheko (2002), Kavdır and Guyer (2004), and Pydipati et al. (2006). Generally, contrast is used to express the local variations in the GLCM. Homogeneity usually measures the closeness of the distribution of elements in the GLCM to its diagonal. Correlation is a measure of image linearity among pixels and the lower the values, the less linearly correlation. Angular second moment (ASM) provides the sum of squared elements in the GLCM. Energy is another expression that usually used instead of ASM, and it equals the square roots of ASM. The above-mentioned parameters were determined at different distances (from $D = 1$ to $D = 10$) for each pixels in the GLCM and then averaged to give only one value at each direction.

In addition, discriminant analysis was conducted by a MATLAB program using the four texture features as discrimination parameters (X-variables) and the ripeness stage as Y-variable. The discriminant analysis was compared under four directions (0°, 45°, 90°, and 135°). The overall classification efficiency for each direction was calculated as the number of fruits which correctly classified to the total number of fruits. The cross-validation by leaving one fruits out each time was considered to validate the discrimination.

5. Results and discussion

5.1. Spectral analysis for predicting quality attributes

5.1.1. Reflectance spectra

Fig. 3 shows the average reflectance spectra in a range of 400–1000 nm of strawberries collected at different ripeness stages. The presence of water in the fruit gave a rise to the characteristic absorption bands that appear as localized minima. The samples containing higher moisture contents had lower reflectivity across their spectra. In spite of ripeness stage, the reflectance curves of strawberry were rather smooth across the entire spectral region. In case of unripe fruit, the reflectance curve had three broadband absorption regions around 500, 680, and 960 nm in addition to small absorption region at 840 nm. The regions around 500 and 680 represent anthocyanin and chlorophyll pigments which represent the colour characteristics in the fruit (Abbott, Lu, Upchurch, & Stroshine, 1997; Seeram, Lee, Scheuller, & Heber, 2006). The absorption regions in the NIR at 840 and 960 nm represent sugars and water absorption bands.

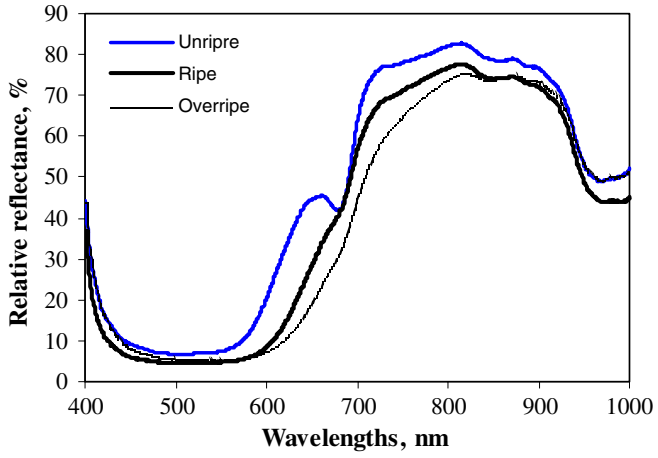

Fig. 3. A typical reflectance of VIS/NIR spectrum (400–1000 nm) of strawberries at different ripeness stages.

The reflectance from both ripe and overripe fruits was consistently lower than that from the unripe one over the entire spectral region. The chlorophyll absorption band at 680 nm was absent in case of ripe and unripe fruits due to the degradation of chlorophyll in these fruits. Meanwhile, the relative reflectance at anthocyanin and sugar absorbance bands (at 500 and 840 nm, respectively) were much lower than those in unripe ones, meaning that the anthocyanin and sugars in the ripe and overripe fruits are much higher.

5.1.2. The PLS models using the whole spectral range of 400–1000 nm

The PLS calibration models were established using the average spectra from 60 fruits in the calibration/training set utilizing the whole spectral range consisting of 826 wavebands. The models were validated using the average spectra of 17 fruits in the validation set. The number of latent factors for PLS model of each attribute is determined at the lowest value of predicted residual errors sum of squares (PRESS) as shown in Fig. 4a. As depicted in Fig. 4a, PRESS had high values at the beginning and decreased rapidly with the increase of the number of latent factors in the model until its lowest value which corresponds to the ideal number of latent factors. With the increase of the number of latent factors, the PRESS starts to increase again and the performance of the model decreases accordingly. The number of latent factors to predict MC, TSS, and pH were nine, four, and six factors, respectively. Measured values of such attributes from chemical test (destructive) and its predicted values resulting from PLS models (nondestructive) are shown in Fig. 4b–d.

Table 1 shows that the model was very accurate for predicting moisture content with r of 0.90 and 0.96 for training and validation sets, respectively. The SEC and SEP were 6.085 and 3.874 for calibration and validation sets, respectively. TSS was predicted with r of 0.8 and SEC of 0.233 for the training set. The accuracy of the model in the validation set for predicting TSS was with r of 0.85 and SEP of 0.184. The pH was predicted with r of 0.87 and SEC of 0.105 for the training set and r of 0.87 and SEP of 0.129 resulted from the validation sets. It is obvious for the three attributes under study that the validation tests gave similar results as the calibration set indicating good performance of the models for predicting these quality components nondestructively.

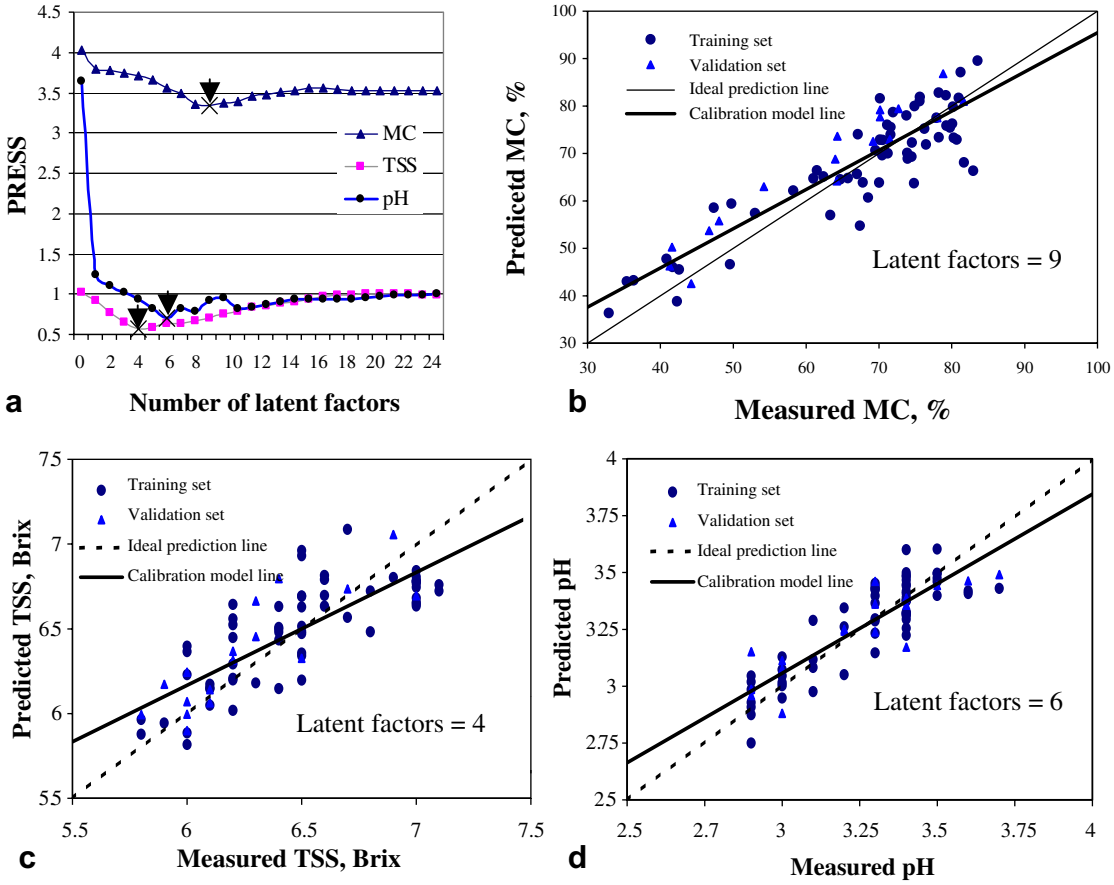

Fig. 4. Prediction of MC, TSS, and pH using PLS models: (a) predicted residual error sum of squares (PRESS) for predicting MC, TSS, and pH as a function of number of factors, (b) measured and predicted MC values for training and validation sets using nine factors, (c) measured and predicted TSS values for training and validation sets using four factors, and (d) measured and predicted pH values for training and validation sets using six factors.

Table 1
PLS models for predicting MC, TSS and pH in strawberry

Attribute	No. of latent factor	Calibration		Validation	
		SEC	*r*	SEP	*r*
MC	9	6.085	0.90	3.874	0.96
TSS	4	0.233	0.80	0.184	0.85
pH	6	0.105	0.87	0.129	0.87

5.1.3. Multiple linear regression models using only the optimal wavelengths

Based on the β-coefficients of PLS, the optimal wavelengths for moisture content prediction as explained in Table 2 were found to be 480, 528, 608, 685, 753, 817, 939, 977 nm. It is clear that, the major absorbance bands of water at 760 and 970 nm were found in the selected range. For predicting TSS, common spectral regions were found in six optimal wavelengths at 421, 520, 581, 683, 847, 950 nm. This result is comparable with the finding of Ito (2002) who found that the optimal wavelengths for soluble solids prediction in strawberry were 882 and 907 nm. In addition, eight optimal wavelengths of 421, 521, 585, 646, 681, 840, 950, 990 nm were found to be responsible for discrimination between fruits in terms of their pH (Table 2).

Fig. 5a shows an example for the procedure of extracting the optimal wavelengths from β-coefficient plot of PLS calibration model of pH. The extracted wavelengths

Table 2
Performance of MLR models for predicting MC, TSS, and pH using only the optimal wavelengths extracted from β-coefficients of PLS analysis

Attribute	Optimal wavelengths	Calibration		Validation	
		SEC	*r*	SEP	*r*
MC	480, 528, 608, 685, 753, 817, 939, 977	6.72	0.87	5.786	0.91
TSS	421, 520, 581, 683, 847, 950	0.220	0.80	0.211	0.80
pH	421, 521, 585, 646, 681, 840, 950, 990	0.084	0.92	0.091	0.94

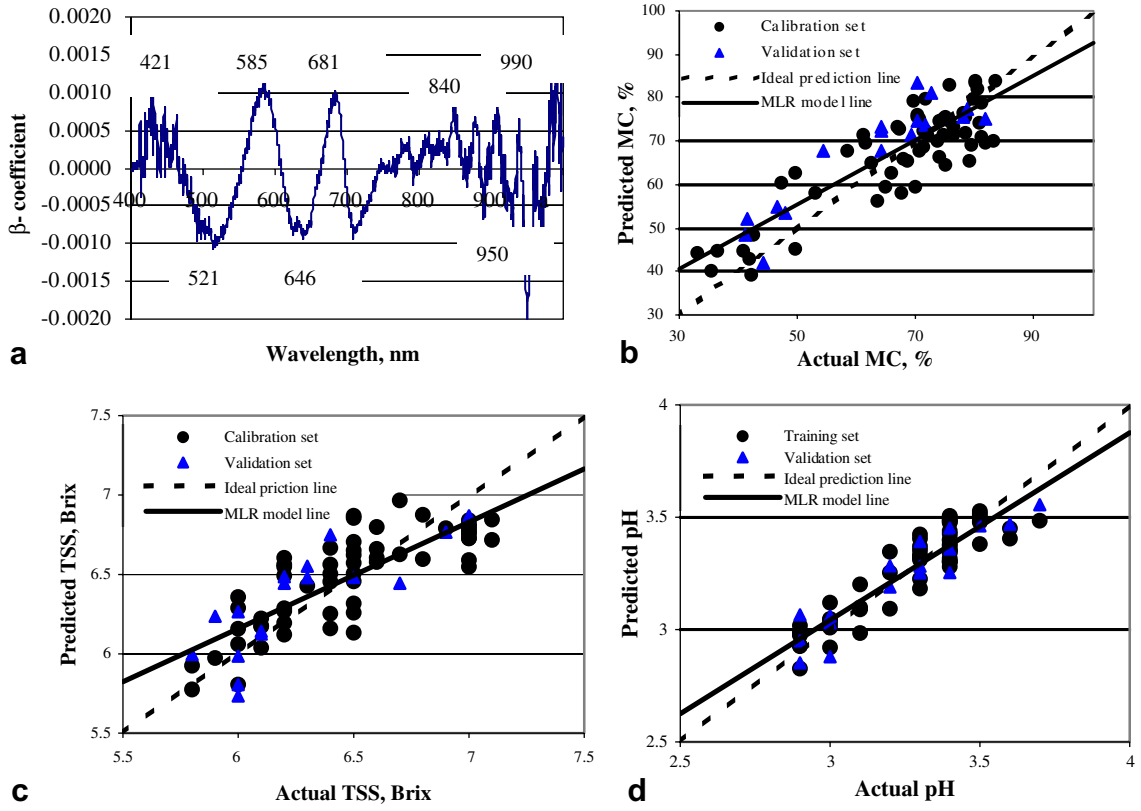

Fig. 5. Optimal wavelengths for predicting quality attributes in strawberry: (a) selection of optimal wavelengths that correspond to the highest absolute value of β-coefficients, (b) measured and predicted MC values for training and validation sets using only eight wavelengths 480, 528, 608, 685, 753, 817, 939, 977 nm, (c) measured and predicted TSS values for training and validation sets using six wavelengths 421, 520, 581, 683, 847, 950 nm, and (d) measured and predicted pH values for training and validation sets using eight wavelengths 421, 521, 585, 646, 681, 840, 950, 990 nm.

were corresponded to the highest absolute values of β-coefficients in the plot in spite of its sign.

For each attribute, the optimal wavelengths were used to build a multiple linear regression (MLR) model between the reflectance at these wavelengths (as *X*-variables) and the measured values of this attribute (as *Y*-variables). Same as above, the calibration set was used to develop the MLR models for predicting the attributes, while the validation set was used to validate these models.

The accuracy of the MLR models for predicting MC, TSS, and pH are shown in Fig 5. The performance of these models is evaluated by SEC, SEP and correlation coefficient (r) as shown in Table 2.

In Table 2, the MLR models produced high performance not only in the training set but also in the validation set. The MLR model of the training set in predicting MC, TSS, and pH had correlation coefficient of 0.87, 0.80 and 0.92 with SEC of 6.72, 0.220, and 0.084, respectively. On the other hand, the correlation coefficient of the validation sets for MC, TSS, and pH prediction were 0.91, 0.80, and 0.94 with SEP of 5.786, 0.211, and 0.091, respectively.

Compared with PLS models, MLR models had higher performance in prediction in terms of SEC, SEP and correlation coefficient although only few numbers of wavelengths were utilized. This is attributed to the fact that the problems of colinearity and overfitting were alleviated in MLR models that utilize only the essential wavelengths and neglect the useless wavelengths that do not carry much spectral information.

5.2. Texture analysis for identifying ripeness stage

5.2.1. Texture parameters extracting from GLCM

Fig. 6 shows the means of the four texture parameters (contrast, homogeneity, energy, and correlation) of strawberry at different ripeness stages. Contrast of unripe fruits was much higher compared with ripe and overripe fruits, which means that unripe fruits contain high local variations in all directions (0°, 45°, 90°, and 135°). The same conclusion obtained when looking to the homogeneity values. The unripe fruits were less homogeneous, while the overripe fruits had the highest homogeneity for all directions. The low correlation values of overripe fruits indicate that this stage had rough texture compared with unripe fruits (highest correlation values) which are smooth in their texture. Generally, no intersections were noticed among all parameters for all direction under all ripeness stages. That means the interferences among ripeness classes would be small, which leads to the possibility for discrimination between these stages using the texture features.

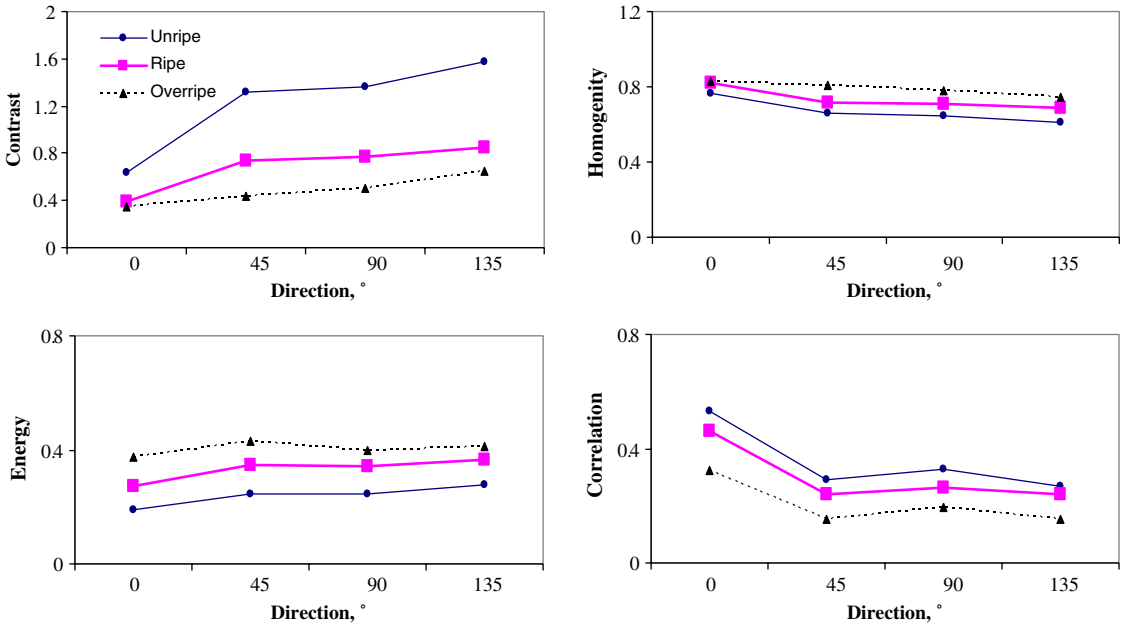

Fig. 6. Strawberry texture parameters at different ripeness stage using GLCM.

Table 3
Confusion matrix for ripeness classification using discriminant analysis with cross-validation at 0° direction

From/to	Unripe	Ripe	Overripe	Total	% Correct
Unripe	0	2	7	9	77.78
Ripe	5	14	0	19	73.68
Overripe	48	1	0	49	97.96
Total	53	17	7	77	89.61

5.2.2. Classification of fruit ripeness based on texture parameters

The discrimination efficiency at different directions of 0°, 45°, 90°, and 135° were found to be 89.61%, 77.62%, 81.82%, and 79.22%, respectively. It is clear that, the horizontal direction at angle 0° had the highest discrimination efficiency (89.61%) compared with the other directions. Only the confusion matrix at this direction for fruit classification is shown in Table 3.

6. Conclusions

The above results indicated the possibility of developing a nondestructive technique using hyperspectral imaging for measuring strawberry quality parameters. The PLS models were established between reflectance spectra and the quality parameters. The models for the moisture content ($r = 0.97$, SEC = 6.085, SEP = 3.874, and nine latent factors), the total soluble solids ($r = 0.85$, SEC = 0.233, SEP = 0.184, and four latent factors) and pH ($r = 0.87$, SEC = 0.105, SEP = 0.129, and six latent factors) show good prediction performance.

The optimal wavelengths were extracted using the β-coefficients from PLS models. Multiple linear regression (MLR) models using only the reflectance at these optimal wavelengths were established. The MLR models for the moisture content ($r = 0.91$, SEC = 6.72, SEP = 5.786, and eight optimal wavelengths), the total soluble solids (SEC = 0.220, SEP = 0.211, $r = 0.80$, and six optimal wavelengths) and pH ($r = 0.94$, SEC = 0.084, SEP = 0.091, and eight optimal wavelengths) demonstrated good prediction performance.

Texture measures were derived from the grey-level co-occurrence matrix (GLCM) of strawberry images to identify its ripeness stage. High classification accuracy of 89.61% for correctly identifying strawberry ripeness stage was achieved using the GLCM parameters at horizontal direction at angle of 0°.

Acknowledgements

The authors gratefully acknowledge the financial support of Egyptian Ministry of Higher Education through the Channel Program Fellowship offered to Mr. Gamal ElMasry.

References

Abbott, J. A., Lu, R., Upchurch, B. L., Stroshine, R.L. (1997). Technologies for non-destructive quality evaluation of fruits and vegetables. In *Horticultural reviews — Technologies for nondestructive quality evaluation of fruits and vegetables* (Vol. 20, pp. 1–120). John Wiley & Sons, Inc. ISBN 0-471-18906-5.

Bharati, M. H., Liu, J. J., & MacGregor, J. F. (2004). Image texture analysis: methods and comparisons. *Chemometrics and Intelligent Laboratory Systems, 72*(1), 57–71.

Bellon, V., Vigneau, J. L., & Leclercq, M. (1993). Feasibility and performance of a new, multiplexed, fast and low-cost fiber-optic NIR spectrometer for the on-line measurement of sugars in fruits. *Applied Spectroscopy, 47*(7), 1079–1083.

Chao, K., Chen, Y. R., Hruschka, W. R., & Park, B. (2001). Chicken heart disease characterization by multi-spectral imaging. *Transactions of the ASAE, 17*(1), 99–106.

Cheng, X., Chen, Y. R., Tao, Y., Wang, C. Y., Kim, M. S., & Lefcourt, A. M. (2004). A novel integrated PCA and FLD method on hyperspectral image feature extraction for cucumber chilling damage inspection. *Transactions of the ASAE, 47*(4), 1313–1320.

Chong, L.-G., & Jun, C.-H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems, 78*(1), 103–112.

Geladi, P., & Kowalski, B. P. (1986a). Partial least-squares regression: a tutorial. *Analytica Chimica Acta, 185*(1), 1–17.

Geladi, P., & Kowalski, B. P. (1986b). An example of 2-block predictive partial least-squares regression with simulated data. *Analytica Chimica Acta, 185*(1), 19–32.

Haaland, D. M., & Thomas, E. V. (1988a). Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Analytical Chemistry, 60*, 1193–1202.

Haaland, D. M., & Thomas, E. V. (1988b). least-squares methods for spectral analyses. 2. Application to simulated and glass spectral data. *Analytical Chemistry, 60*, 1202–1208.

Haralick, R. M., & Shanmugam, K. S. (1974). Combined spectral and spatial processing of ERTS imagery data. *Remote Sensing of Environment, 3*(1), 3–13.

Hruschka, W. R. (2001). Near infrared technology in the agricultural and food industries. In Phil Williams & Karl Norris (Eds.), *Data analysis: Wavelength selection methods*. American Association of Cereal Chemists.

Ito, H. (2002). Potential of near infrared spectroscopy for nondestructive estimation of Brix in strawberries. In *The proceedings of IV international strawberry symposium, January 31, 2002, Tampere, Finland*. The International Society for Horticultural Science (ISHS) Acta Horticulturae (Vol. 567, pp. 751–754). ISBN: 9-066-05775-0.

Katayama, K., Komaki, K., & Tamiya, S. (1996). Prediction of starch, moisture, and sugar in sweetpotato by near infrared transmittance. *HortScience, 31*(6), 1003–1006.

Kavdır, I., & Guyer, D. E. (2004). Comparison of artificial neural networks and statistical classifiers in apple sorting using textural features. *Biosystems Engineering, 89*(3), 331–344.

Keskin, M., Dodd, R. B., Han, Y. J., & Khalilian, A. (2004). Assessing nitrogen content of golf course turfgrass clippings using spectral reflectance. *Applied Engineering in Agriculture, 20*(6), 851–860.

Kim, M. S., Lefcourt, A. M., Chao, K., Chen, Y. R., Kim, I., & Chan, D. E. (2002). Multispectral detection of fecal contamination on apples based on hyperspectral imagery: part I. Application of visible and near-infrared reflectance imaging. *Transactions of the ASAE, 45*(6), 2027–2037.

Lammertyn, J., Nicolaï, B., Ooms, K., De Smedt, V., & De Baerdemaeker, J. (1998). Non-destructive measurement of acidity, soluble solids, and firmness of Jonagold apples using NIR-spectroscopy. *Transactions of the ASAE, 41*(4), 1089–1094.

Létal, J., Jirák, D., Šuderlová, L., & Hájek, M. (2003). MRI 'texture' analysis of MR images of apples during ripening and storage. *Lebensmittel-Wissenschaft und-Technologie, 36*(7), 719–727.

Li, Q., Wang, M., & Gu, W. (2002). Computer vision based system for apple surface defect detection. *Computers and Electronics in Agriculture, 36*(2), 215–223.

Liu, Y., Chen, Y. R., Wang, C. Y., Chan, D. E., & Kim, M. S. (2006). Development of hyperspectral imaging technique for the detection of chilling injury in cucumbers; spectral and image analysis. *Applied Engineering in Agriculture, 22*(1), 101–111.

Liu, Y., Windham, W. R., Lawrence, K. C., & Park, B. (2003). Simple algorithms for the classification of visible/near-infrared and hyperspectral imaging spectra of chicken skins, feces, and fecal contaminated skins. *Applied Spectroscopy, 57*(12), 1609–1612.

Lu, R. (2004). Multispectral imaging for predicting firmness and soluble solids content of apple fruit. *Postharvest Biology and Technology, 31*(1), 147–157.

Lu, R., & Ariana, D. (2002). A near-infrared sensing technique for measuring internal quality of apple fruit. *Applied Engineering in Agriculture, 18*(5), 585–590.

Mehl, P. M., Chen, Y.-R., Kim, M. S., & Chan, D. E. (2004). Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations. *Journal of Food Engineering, 61*(1), 67–81.

Nagata, M., Tallada, J. G., Kobayashi, T., & Toyoda, H. (2005). NIR hyperspectral imaging for measurement of internal quality in strawberries. *ASAE meeting*, Tampa, Florida, ASAE Paper No. 053131.

Nagata, M., Tallada, J. G., Kobayashi, T., Cui, Y., & Gejima, Y. (2004). Predicting maturity quality parameters of strawberries using hyperspectral imaging. *ASAE/CSAE annual international meeting*, Ottawa, Ontario, Canada, Paper No. 043033.

Osborne, S. D., Jordan, R. B., & Künnemeyera, R. (1997). Method of wavelength selection for partial least squares. *Analyst, 122*, 1531–1537.

Park, B., Abbott, J. A., Lee, K. J., Choi, C. H., & Choi, K. H. (2003). Near-infrared diffuse reflectance for quantitative and qualitative measurement of soluble solids and firmness of Delicious and Gala apples. *Transactions of the ASAE, 46*(6), 1721–1731.

Park, B., Windham, W. R., Lawrence, K. C., & Smith, D. P. (2004). Hyperspectral image classification for fecal and ingesta identification by spectral angle mapper. *ASAE/CSAE meeting*, Ottawa, Ontario, Canada, ASAE Paper No. 043032.

Peiris, K. H. S., Dull, G. G., Leffler, R. G., & Kays, S. J. (1999). Spatial variability of soluble solids or dry-matter content within individual fruits, bulbs, or tubers: Implications for the development and use of NIR spectrometric technique. *HortScience, 34*(1), 114–118.

Peirs, A., Scheerlinck, N., Touchant, K., & Nicolaï, B. M. (2002). Comparison of fourier transform and dispersive near-infrared reflectance spectroscopy for apple quality measurements. *Biosystems Engineering, 81*(3), 305–311.

Polder, G., Van der Heijden, G. W. A. M., & Young, I. T. (2002). Spectral image analysis for measuring ripeness of tomatoes. *Transactions of the ASAE, 45*, 1155–1161.

Pydipati, R., Burks, T. F., & Lee, W. S. (2006). Identification of citrus disease using color texture features and discriminant analysis. *Computers and Electronics in Agriculture, 52*(1), 49–59.

Seeram, N. P., Lee, R., Scheuller, H. S., & Heber, D. (2006). Identification of phenolic compounds in strawberries by liquid chromatography electrospray ionization mass spectroscopy. *Food Chemistry, 97*(1), 1–11.

Tsheko, R. (2002). Discrimination of plant species using co occurrence matrix of leaves. *Agricultural Engineering International: the CIGR Journal of Scientific Research and Development, IV*(May), Manuscript IT 01 004.

Xing, J., & De Baerdemaeker, J. (2005). Bruise detection on 'Jonagold' apples using hyperspectral imaging. *Postharvest Biology and Technology, 37*(1), 152–162.