

Research Article

Computer Vision Estimation of the Volume and Weight of Apples by Using 3D Reconstruction and Noncontact Measuring Methods

Baohua Zhang, Ning Guo, Jichao Huang, Baoxing Gu, and Jun Zhou 

College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 210031, Jiangsu, China

Correspondence should be addressed to Jun Zhou; zhoujunnjau@126.com

Received 19 January 2020; Accepted 8 October 2020; Published 18 November 2020

Academic Editor: Xavier Vilanova

Copyright © 2020 Baohua Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A computer vision system for the estimation of apple volume and weight by using 3D reconstruction and noncontact measuring methods was investigated. The 3D surface of the apples could be reconstructed by using a single multispectral camera and near-infrared linear-array structured light. Both the traditional image feature and height information were extracted from the height maps. Two different type height features (Type I and II) were extracted, and both of them were fused with a projection area to form combination features (Combination Feature I and II). Partial least squares analysis and least squares-support vector machine were implemented for calibration models with projection area and combination features as inputs. Grid-Search Technique and Leave-One-Out Cross-Validation were also investigated to find out the optimal parameter values of the RBF kernel. The optimal LS-SVM models with Combination Feature II outperformed PLS models. The coefficient and root mean square error of prediction for the best prediction by LS-SVM were 0.9032 and 10.1155 for volume, whereas 0.8602 and 9.9556 for weight, respectively. The overall results indicated that height information can improve the prediction performance, and the proposed system could be applied as an alternative to the traditional methods for noncontract measurement of the volume and weight of apple fruits.

1. Introduction

Volume and weight are two important parameters in the external quality of apples. Apple volume and weight parameters not only have an impact on consumers' preferences, thereby influencing their marketing values, but are also considered indicators of apple quality. Therefore, apple volume and weight estimation during postharvest handling and processing stages are really important and necessary for producers and it is a goal of some research [1, 2].

Computer vision systems are being widely used for quality monitoring and inspection in agricultural products and food processing. The traditional computer vision systems imitate the vision of the human eyes by capturing images using three filters centered at red (R), green (G), and blue (B) wavelengths [3]. Nowadays, computer vision systems are playing an indispensable role for external quality inspection in the automatic grading and sorting systems. Their

applications about fruits and vegetables include defect detection such as common defect detection on citrus [4, 5], defect detection on apples [6, 7], defect detection on bananas [8], size assessment of berries [9], size estimation of sweet onions [10], automatic classification of fruits [11], and color inspection of various fruits and vegetables [12, 13].

Various computer vision methods have been investigated to estimate the volume and weight of agricultural products by using noncontact measurement methods. Since the 2D digital images captured by the computer vision systems are composed of pixels, the projected area, perimeter, or length and width features can be measured in the images by using image processing algorithms [14]. The most commonly used image feature and basic convenient measurement for volume and weight evaluation is the projected area. Ellipsoid approximation and image processing in the projected area were used to estimate the volume of watermelons of varying sizes by Koc [15]. Teoh and Syaifudin [16] measured the projected area

of mango by image processing and analysis against the actual weight of mango in a graph, the results showed that the projected area measured has a high correlation with the actual weight of mango with $R^2 = 0.934$. Estimation of the volume and weight of spherical or quasi-spherical objects is relatively easy due to they have a strong correlation with some dimensional parameters of the 2D projected area, but it becomes more complex to fruits and vegetables due to their natural irregularities [17].

In order to extract the height information (the third dimension) of the agricultural products for a more accurate evaluation of the volume and weight, three-dimensional (3D) computer vision techniques have been increasingly investigated and applied to measure the volume of agricultural products [10]. In order to acquire the 3D images (X dimension and Y dimension represent the spatial information, Z represents the height information) of the objects, various sensors and techniques could be used. Binocular stereo vision techniques, which are based on binocular CCD cameras, are the most common ways to generate the 3D images in the 3D detection. 3D measurement by using binocular stereo vision systems is the process of obtaining depth information from a pair of cameras [18]. Chalidabhongse et al. described a vision system that can reconstruct 3D mango volume by using volumetric caving on multiple silhouette images [19]. After craving all silhouettes, the coarse 3D shape of the fruit could be obtained, and then, the volume and surface area would be computed. Omid et al. developed an image processing-based technique to measure the volume and mass of citrus fruits [20]. Their technique used two cameras to give perpendicular views of the fruits. The volumes of the fruits were calculated by dividing the fruit image into a number of elementary elliptical frustums. The volume is calculated as the sum of the volumes of individual frustums. However, 3D techniques based on binocular stereo vision systems are time-consuming and not suitable to be used for online or real-time volume and weight evaluation of agricultural products due to the complexity of the CCD camera calibration, feature point extraction, and matching. A laser-based vision system, which is based on a monocular camera coupled with a laser, is a classic active 3D computer vision system. Distance information from the surface of the object to the camera could be measured by the laser-based vision systems by using the time of flight (TOF) technique [21]. However, applications about the quality inspection by using laser-based vision systems are not found yet. An RGB-depth (RGB-D) vision system, which is based on an RGB-D sensor, is another type of active 3D computer vision system. RGB-D vision systems can simultaneously capture the depth and color images of the scene and automatically map the depth and color data, resulting in a colored point cloud in a 3D spatial domain [10]. Wang and Li measured the size of the sweet onions using nondestructive imaging methods based on the RGB-D sensor [10]; the results demonstrated that it is promising to estimate the onion size based on its depth image. Coded structured light-based vision systems are widely used systems for 3D surface reconstruction and 3D size measurement in industry and precision inspection fields [22, 23]. Storbeck and Daan developed a coded structured light vision

system to evaluate the volume of fish; the accuracy was 95% [24]. Other applications in the industry have been stated the coded structured light vision systems are suitable and efficient for size measurement of moving objects on the conveyor belt. However, the applications of coded structured light vision systems for volume and weight measurement in the food and fruit processing industry are scarce.

2. Objectives

The primary objective of our research was to develop a near-infrared linear-array structured light vision system to estimate the volume and weight of apples using 3D reconstruction and noncontact measurement methods. In order to achieve the primary objective, several subobjectives as follows have to be fulfilled: (1) developing a near-infrared linear-array structured light vision system, (2) acquiring the 3D reconstruction and height map images of the inspected apple, (3) extracting the 2D and 3D image features and selecting the most relevant features or feature combination, (4) establishing the multivariate calibration and prediction models using PLS and LS-SVM, and (5) evaluating the prediction performance of the prediction models and features.

3. Materials and Methods

3.1. Samples Used in Our Research. There are many cultivars of apples grown in China. “Fuji” apple is one of the most popular cultivars, favored by consumers due to its rich nutrition and healthcare benefits. Fuji apples were purchased from a local fruit market in Beijing, China, on October 15, 2014. A total of 100 Fuji apples of various sizes and shapes were selected as the experimental samples in our research. The diameters of the samples vary from 60 mm to 95 mm. Two images were captured for every sample in a random position by our system. A total of 200 images (3D reconstruction and height map images) were acquired. Seventy samples were used for training the regression models, and the rest were used for validating the models.

3.2. Measuring the Actual Volume and Weight of Apples. The actual volume and weight of the apples should be measured for training, calibrating, and validating the noncontract estimation models.

The actual volumes of the apples were measured using the water displacement method (WDM). In our research, the apple is dipped into the water with a sinker rod. The weight of the displaced water is then calculated by subtracting the weight of the water-filled container from the weight of the container when it contains the fruit [20]. The resulting value is then used to calculate the actual volume of the apple by using the following equation [25]:

$$\text{Volume (cm}^3\text{)} = \frac{\text{Weight of displaced water (kg)}}{\text{Water density (kg/cm}^3\text{)}}. \quad (1)$$

It is noted that the sinker rod used in our research is very thin; considering that the volume of the apple is far greater than that of the sinker rod, the volume of the sinker rod

can be considered negligible and therefore can be neglected in the volume measurement.

The actual gross weights of the apples were measured by using a digital balance (Shuangquan, China) with an accuracy of ± 0.01 g.

In our experiment, the volume of the samples vary from 180 cm^3 to 360 cm^3 , and the weight of the samples vary from 160 g to 280 g. The large variations of volumes and weights could cover almost all the apples in the markets, and this can ensure the universality and practicability of the system and algorithm.

3.3. Near-Infrared Linear-Array Structured Light Vision System. The near-infrared linear-array structured light vision system developed in our research consists of a computer (Dell, Intel® Core(TM) i5-2400 CPU @3.10GHz, RAM 4.0GB), a CCD camera (JAI AD-080GE2CCD multispectral camera, Japan) with a high spatial resolution (1024×768 pixels, RGB image), and a high sensitivity in the near-infrared area (800 nm, NIR image), a conveyor belt, and a lighting system composed of a pair of visible Light Emitting Diode (LED) light source and a near-infrared (800 nm) linear-array structured light. The pair of the visible LED light source was placed symmetrically above and to each side of the sample. The near-infrared linear-array structured light was fixed at the upper left side of the sample with the same plane and height as the camera. The entire system was housed in a black box. Both the RGB image and NIR image at 800 nm wavelength could be acquired by the multispectral camera simultaneously through the same optical path. The schematic diagram of the main components of our near-infrared linear-array structured light vision system is illustrated in Figure 1.

The image acquisition and feature extraction algorithms, as well as the control panel, were integrated into the hand-coded software, implemented in Visual C++ and Open Source Computer Vision (OpenCV).

3.4. 3D Reconstruction and Height Map Image Acquisition. The 3D reconstruction and height map image of the inspected apples were acquired by using the proposed system in our research; the detailed processes were as follows. The height information for one pixel in the light strip was measured by triangulation (Figure 2). The reference plane (conveyor belt) is configured to be parallel to the baseline of the camera and the laser projector. If there is no object under the camera and projector, the projection light strip will be a straight line without distortion. However, if there is an object with a certain height, a light strip with distortion will be present in the scene and imaged by the camera. The principle can be explained by the similar triangle ΔAPB and ΔCPD , and the relative height h from the surface point P to the reference plane can be calculated by using the equation:

$$\frac{h}{L-h} = \frac{d}{s}, \quad (2)$$

where s represents the baseline distance from the CCD camera to the laser projector, L represents the stand-off dis-

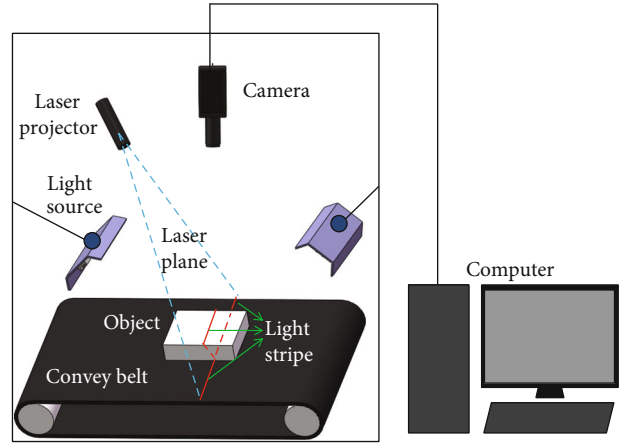


FIGURE 1: Schematic illustration of the computer vision system used in our research.

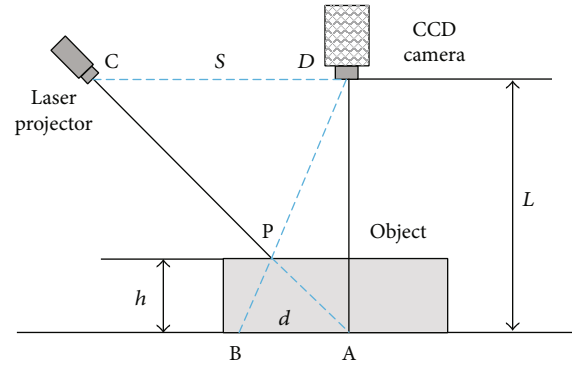


FIGURE 2: Measurement principle of the height information by using a linear-array structured light method.

tance from the optical center of the CCD camera to the reference plane, and d represents the distance between two corresponding points A and B, which can be extracted by image processing. Equation (2) can be transformed into:

$$h = \frac{L-h}{s}d. \quad (3)$$

Considering the value of L is far greater than that of h , equation (3) can be approximated as:

$$h = \frac{L}{s}d. \quad (4)$$

Therefore, the first step of height measurement is to identify points A and B in the images acquired by the camera. It is obvious that the x coordinate of point A (x, y) can be obtained before measurement (actually, the x -coordinate of all the original pixels in the light strip is the same because the light strip is perpendicular to the x -axis of the camera image plane, so the original x -coordinate will be recorded in the program) and the y of point A (x, y) is the same as the point P and can be easily obtained by image processing.

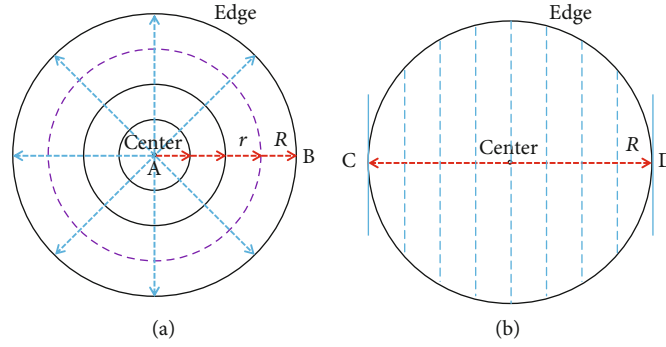


FIGURE 3: Feature extraction methods: (a) Type I feature; (b) Type II feature.

As the points B, P, and D are collinear, it is obvious that point B is easily detected because it is the same pixel for both inspected part point P and reference plane point B in the camera image plane. The height of each pixel in the central line of the light strip will be conducted by equation (3). After calculating, the height profile would be obtained. The near-infrared linear-array structured light vision system would scan the whole apple and calculate the height profiles in the light strip during the noncontact measurement with the adjustable motor speed of the convey belt. After the scanning, the 3D surface of the upper half of the inspected apples would be reconstructed. In order to make it more intuitive and visual, the height map would be also present in pseudocolor height map and gray level height map images in this paper.

3.5. Feature Extraction and Selection. In order to establish the prediction models for volume and weight estimation, relevant image features should be extracted and selected. Since the 3D height map images (X and Y represent the spatial information, Z represents the height information) are acquired by our vision system, the features relevant to the volume and weight are not the same as the image features extracted from traditional 2D images. Not only the features relative to boundary shape and projection area as in 2D images could be extracted, but the features relative to the height and surface conditions in 3D images could also be extracted. This makes it more reliable and accurate to estimate the volume and weight of apples in 3D reconstruction images than that in the 2D images.

In this paper, both the commonly used traditional 2D image feature (projection area) and height features would be extracted from the 3D reconstruction images. Two different types of height features were extracted from the 3D height maps of apples.

The projection area can be extracted from the projection area obtained by thresholding the gray level height map according to the height information.

Height features can be extracted directly from the 3D height map images. Two different types of height features would be extracted in our research. The first type of height features (marked as Type I feature) was extracted from the 50 concentric annuli equally distributed in the height maps with an adaptive distance to the size of the inspected apples

by averaging the height values of all pixels as shown in Figure 3(a). The second type of height features (marked as Type II feature) was extracted from the 50 vertical lines equally distributed in the height maps with an adaptive distance to the size of the inspected apples by averaging the height values of all pixels as shown in Figure 3(b).

In real-world applications, relevant features are not generally known beforehand, which results in the extraction of several features that also include irrelevant ones [14, 26]. In order to find out the best prediction model and efficient features or feature combination, a single projection area and each type of height feature combined with the projection area will be fed to the regression models. It should be noted that, in order to make it clear to the readers, we labeled the combination of Type I feature and projection area as Combination Feature I, and the combination of Type II feature and projection area as Combination Feature II.

3.6. Partial Least Squares (PLS). PLS is a bilinear modeling method where the original independent information (X -data) is projected onto a small number of latent variables (LVs) to simplify the relationship between X and Y for predicting with the smallest number of LVs [4, 27, 28]. The first step in PLS is to decompose the matrix, and the model is given:

$$X = TP + E, \quad (5)$$

$$Y = UQ + F. \quad (6)$$

In those equations, T and U are the score matrices of X matrix and Y matrix, P and Q are the loading matrices of the X matrix and Y matrix, and E and F are the errors which come from the process of PLS regression.

The second step is that T and U are the processes by linear regression. It must build the following linear correlation:

$$U = BT + E, \quad (7)$$

where B represents the internal relations between U and T ; in order to reach this object, the coordinate of T is rotated. In PLS analysis, the optimal number of PLS components that optimize the predictive ability of the model should be determined. This choice is typically made with the use of cross-validation. Prediction residual sum of squares (PRESS) or

total residual variance (RV) for the test samples is used as a function to determine the number of LVs that optimizes the predictive ability of the model.

3.7. Least Squares-Support Vector Machine (LS-SVM). LS-SVM, a state-of-the-art statistical learning method, is capable of dealing with linear and nonlinear multivariate analysis and resolving these problems in a relatively fast way. Moreover, the support vector machine (SVM) is capable of learning in high-dimensional feature space with fewer training data. It employs a set of linear equations instead of quadratic programming problems to obtain the support vectors. SVM embodies the structural risk minimization principle instead of the traditional empirical risk minimization principle to avoid over-fitting problems. The LS-SVM regression model can be expressed as [4]:

$$y(x) = \sum_{k=1}^N \alpha_k K(x, x_k) + b, \quad (8)$$

where $K(x, x_k)$ is the kernel function, x_k is the input vector, α_k is the Lagrange multiplier called support value, and b is the bias. The frequently used kernel function $K(x, x_k)$ includes linear kernel, nonlinear kernel, and radial basis function (RBF) kernel. $K(x, x_k)$ must follow Mercer's condition and perform the linear and nonlinear mapping [29, 30], considering that the RBF kernel is a nonlinear function and a more compact supported kernel and could reduce the computational complexity of the training procedure while giving good performance under general smoothness assumptions. In our study, the RBF kernel function was used and the RBF kernel function is defined as follows:

$$K(x, x_k) = \exp\left(-\frac{\|x_k - x\|^2}{2\sigma^2}\right), \quad (9)$$

where $\|x_k - x\|$ represents the distance between the input vector and threshold vector, and σ is a width vector.

The proper kernel parameter setting plays a crucial role in building a good LS-SVM regression model with high prediction accuracy and stability. In this study, we used Grid-Search Technique and Leave-One-Out Cross-Validation to find out the optimal parameter values, including the regularization parameter gam (γ) and the RBF kernel function parameter $sig2$ (σ^2). Grid-Search is a two-dimensional minimization procedure based on exhaustive search in a limited range. Detailed information about Grid-Search Technique and Leave-One-Out Cross-Validation can be found in the literature [28].

3.8. Evaluation of the Performance of the Methods. The performance of the model calibration and prediction was assessed in terms of correlation coefficient (r), root mean square error of calibration (RMSEC), and root mean square error of prediction (RMSEP). The main evaluation indices were r and RMSEP in our study. The bias was taken into consideration for distinguishing systematic error. These indices are defined as follows [4, 29]:

$$r = \frac{\sqrt{\sum_{i=1}^n (y \wedge_i - y_i)^2}}{\sqrt{\sum_{i=1}^n (y \wedge_i - y_m)^2}}, \quad (10)$$

$$\text{RMSEC} = \sqrt{\frac{1}{n_c} \sum_{i=1}^{n_c} (y \wedge_i - y_i)^2}, \quad (11)$$

$$\text{RMSEP} = \sqrt{\frac{1}{n_p} \sum_{i=1}^{n_p} (y \wedge_i - y_i)^2}, \quad (12)$$

$$\text{bias} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i), \quad (13)$$

where \hat{y}_i is the predicted value of the i -th observation, y_i is the measured value of the i -th observation, y_m is the mean value of the calibration or prediction set, and n , n_c , and n_p are the number of observations in the data set, calibration, and prediction set, respectively. Generally, a good model should have higher correlation coefficients and lower RMSEC, RMSEP, and bias values, but also a small difference between RMSEC and RMSEP.

3.9. Flowchart of Our Method. The flowchart of the proposed method is shown in Figure 4. The main steps include 3D reconstruction, feature extraction, feature fusion, multivariable calibration, and prediction.

4. Results and Discussion

4.1. Results of 3D Surface Reconstruction. Figure 5 shows the results of 3D surface reconstruction in three different forms, namely, 3D reconstruction, gray level height map images, and pseudocolor height map images. Figure 5(a) shows the 3D surface reconstruction with the actual diameter and height of the inspected apple samples. It is noted that the surface of 3D reconstruction is coarse due to the low cost of our system hardware and equation simplification, but the overall reconstruction results could be acceptable for volume and weight estimation. Figure 5(b) shows the 3D height map images in gray level (X and Y dimensions represent the spatial information, and Z dimension represents the intensity; in our research, different height was denoted as different gray level intensity from the value of 0 to 255). The projection area and 3D height information would be extracted from the gray level height map images. As humans are more sensitive to color images, pseudocolor images from the top view of apples are illustrated in Figure 5(c). Different height was denoted with a different color, and deeper color means the higher height. As shown in Figure 5(c), the pseudocolor height map images present a higher height in the central area and a lower height in the edge positions; stems and calyxes in apples also present a relatively lower height. It should be noted that only the upper surface could be reconstructed by our computer vision system and methods due to the camera could only capture the near-infrared linear-array structured light strip projecting on the upper half surface of the apple,

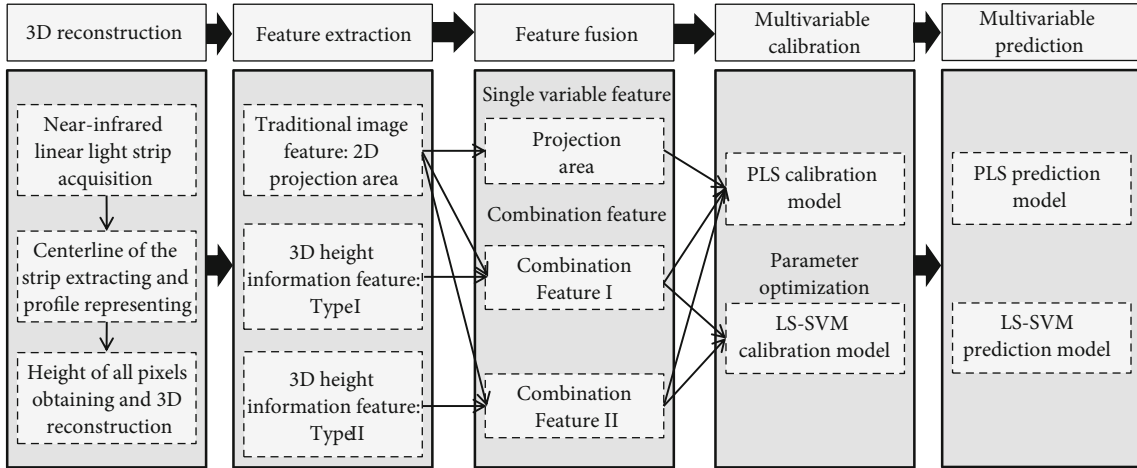


FIGURE 4: Flowchart of our methods for volume and weight estimation.

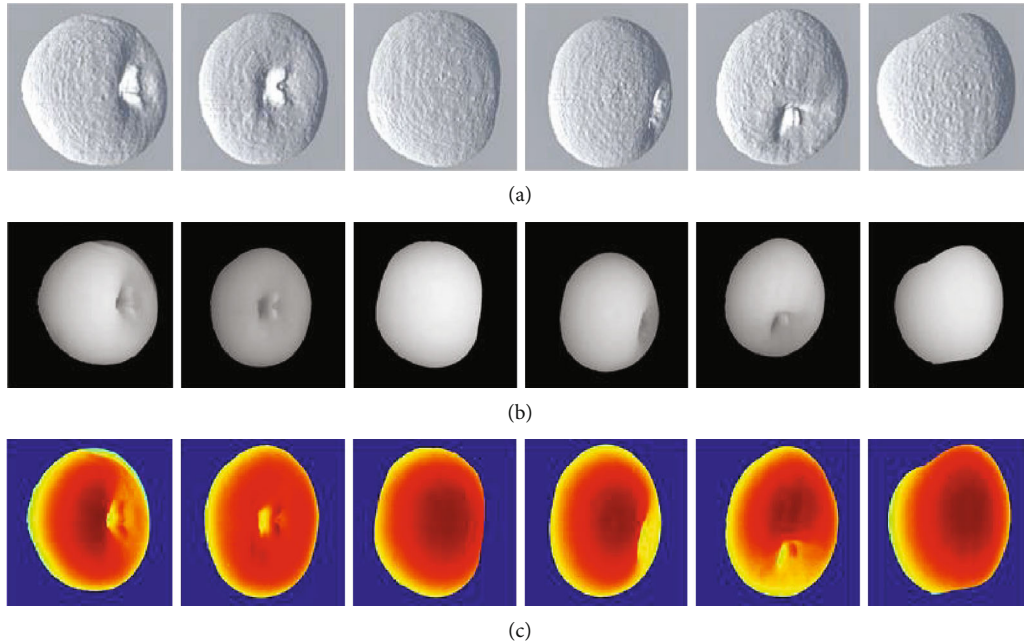


FIGURE 5: 3D surface reconstruction results: (a) 3D surface reconstruction with the actual diameter and height of the inspected apple samples, (b) 3D height map images in gray level, and (c) 3D height map images in pseudocolor.

and all the image processing and feature extraction were conducted in gray level height map images.

4.2. Estimation Results of PLS and LS-SVM for Volume. The traditional image feature (projection area) and the two types of combination features were used to establish the regression model for noncontact measuring by using two popular regression models PLS and LS-SVM, respectively. It is noted that both the models were established with the same samples, and normalization processing was also applied before calibration.

Before establishing the LS-SVM calibration model, three crucial problems need to be solved, namely, the determination of the optimal input feature subset, proper kernel function, and the best kernel parameters. Feature subset is

obtained in the feature extraction step, and two different combination features would be used as the input data set, respectively. Kernel function was chosen as RBF. So the remaining important problem is to decide the best kernel parameters, including regularization parameter gam (γ) and the RBF kernel function parameter $sig2$ (σ^2). These two parameters determine the learning ability, prediction ability, and generalization ability of LS-SVM [30]. Gam (γ) is used to maximize model performance (on training) and minimize the model complexity. Large gam (γ) implies little regularization and thus a more nonlinear model. $sig2$ (σ^2) influences the number of neighbors in the model. And large $sig2$ (σ^2) means more neighbors in the model which leads to a more nonlinear model. In this paper, a two-step Grid-Search technique using geometric steps with Leave-One-Out Cross-

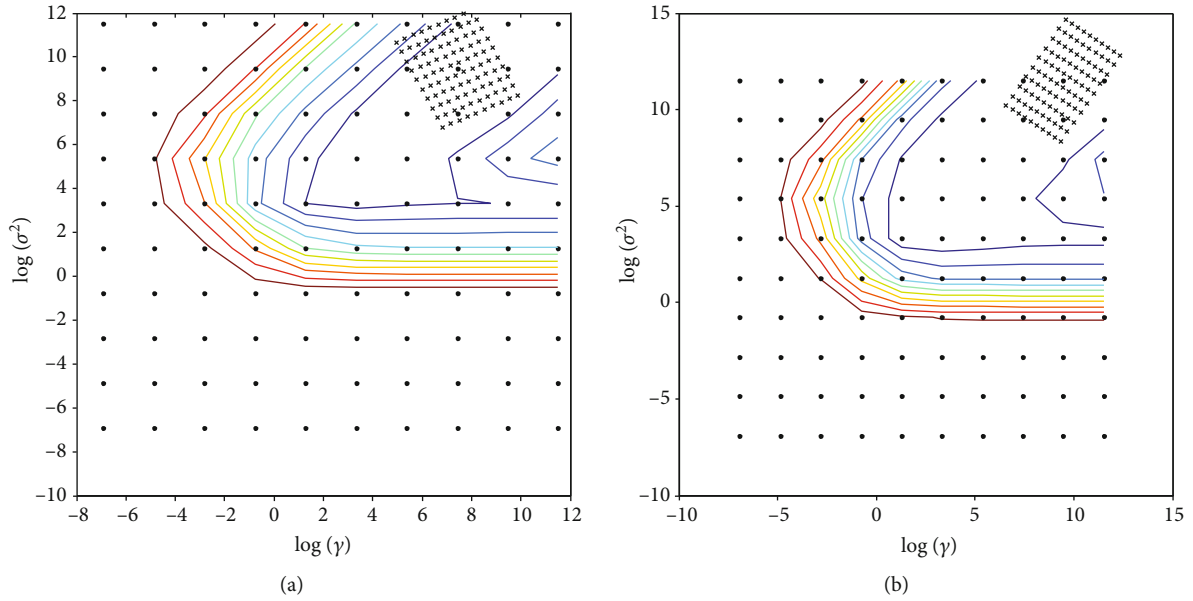


FIGURE 6: Optimization of gam (γ) and $sig2$ (σ^2) parameters for LS-SVM in volume estimation. (a) Optimizing processes for volume estimation by using Combination Feature I. (b) Optimizing processes for volume estimation by using Combination Feature II.

Validation was employed to obtain the optimal gam (γ) and $sig2$ (σ^2) within the region of (10^{-2} to 10^6) which were set based on experience. The first step of Grid-Search was for a crude search with a large step size and the second step for the specified search with a small step size. For each combination of gam (γ) and $sig2$ (σ^2) parameters, the root mean square error of cross-validation (RMSECV) was calculated and the optimum parameters were selected when a smaller RMSECV was produced. The optimizing processes for volume estimation by using Combination Feature I and Combination Feature II as input data are shown in Figure 6. The grids “.” in the first step are 10×10 , and the searching step in the first step is large. The optimal search area is determined by the error contour line. The grids “x” in the second step are 10×10 , and the searching step in the second step is smaller. The optimal search area is determined based on the first step. In the volume estimation by using LS-SVM, the initial value of γ and σ^2 were set to 0.01, and the optimal pair of (γ , σ^2) was found at the value of $\gamma = 7035.6$ and $\sigma^2 = 29983.9$ when using the Combination Feature I as the input data; the optimal pair of (γ , σ^2) was found at the value of $\gamma = 22712.6$ and $\sigma^2 = 89418.1$ when using the Combination Feature II as the input data. This indicates that the LS-SVM mode established with Combination Feature II is a more nonlinear model compared with that with Combination Feature I.

In order to find out the best features and models, five models were established based on the traditional 2D image feature and two different types of combination features. Figure 7 shows the noncontact measuring results versus actual volume measured by the water displacement method (WDM) charts for PLS and LS-SVM. Figure 7(a) shows the measurement results estimation by the PLS model using the traditional projection area are plotted against the actual volume measured by WDM. The solid line is the regression line corresponding to the ideal, unity correlation between the pre-

dicted and reference values. The correlation coefficients, RMSEP, for prediction sets were 0.8493 and 16.9978, respectively. Figure 7(a) indicates that the PLS model based on the projection area has some prediction power; as a popular traditional 2D image feature, the projection area could be used for volume measurement in the situation of demand on relatively low precision. Figures 7(b) and 7(c) show the measurement results estimation by the PLS model using the two different types of combination features are plotted against the actual volume measured by WDM. The correlation coefficients, RMSEP, for prediction sets of the PLS model with Combination Feature I were 0.6676 and 24.5473, and with Combination Feature II were 0.8618 and 16.3042. The unity correlation between the predicted and reference values in Figures 7(b) and 7(c) also indicated the PLS model with Combination Feature II outperformed the PLS model with Combination Feature I. Figures 7(d) and 7(e) show the measurement results estimation by LS-SVM model using the two different types of combination features are plotted against the actual volume measured by WDM. The correlation coefficients, RMSEP, for the prediction sets of the LS-SVM model with Combination Feature I were 0.8927 and 10.7424, and with Combination Feature II were 0.9032 and 10.1155. As the two different unity correlations between the predicted and reference values illustrated in Figures 7(d) and 7(e), the LS-SVM model with Combination Feature II could get better prediction results than that of the LS-SVM model with Combination Feature I.

Compared with the PLS model with the traditional 2D image feature, the PLS model with Combination Feature II could get better prediction performance. The better prediction performance mainly benefits from the height information from the 3D surface reconstruction. Compared with the PLS models, both the LS-SVM models could get a better prediction performance; the reason might be that the LS-

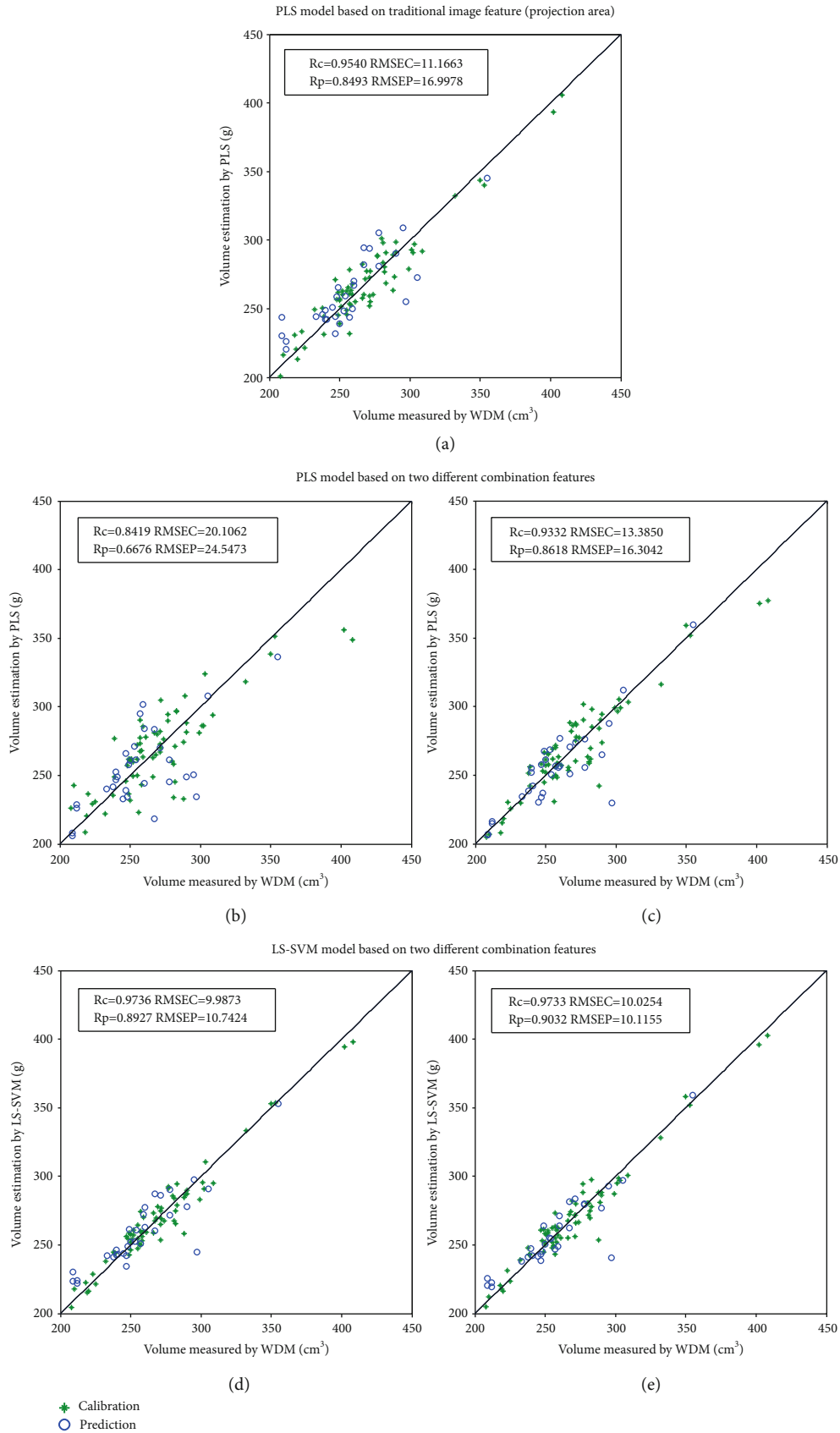


FIGURE 7: Measured versus estimated values for volume prediction by the PLS and LS-SVM models.

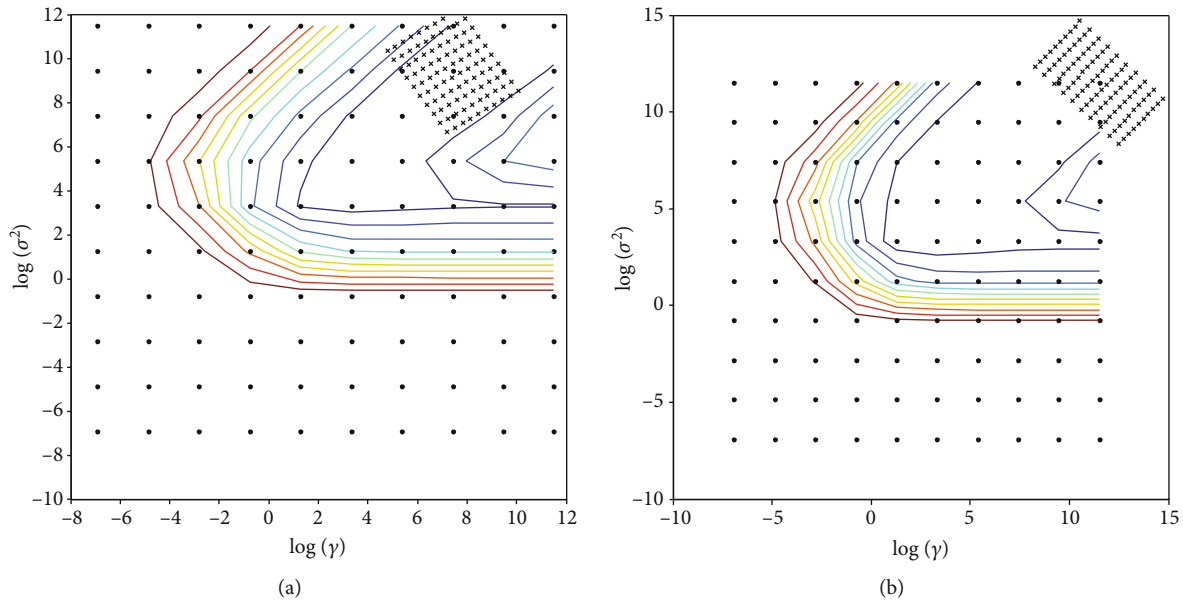


FIGURE 8: Optimization of gam (γ) and $sig2$ (σ^2) parameters for LS-SVM in weight estimation. (a) Optimizing processes for weight estimation by using Combination Feature I. (b) Optimizing processes for weight estimation by using Combination Feature II.

SVM models took the nonlinear information of the combination features into consideration, and the nonlinear information had improved the prediction precision. Compared with other models and features, the LS-SVM model with Combination Feature II has the strongest ability for volume non-contact estimation.

4.3. Estimation Results of PLS and LS-SVM for Weight. In order to find out the best model and features for apple weight estimation, a similar process with volume estimation was conducted. In the weight estimation by using LS-SVM, the initial value of γ and σ^2 were also set to 0.01, and the optimal pair of (γ, σ^2) was found at the value of $\gamma = 67935.7$ and $\sigma^2 = 373443.9$ when using Combination Feature I as the input data; the optimal pair of (γ, σ^2) was found at the value of $\gamma = 1739556.5$ and $\sigma^2 = 257323.3$ when using the Combination Feature II as the input data. The optimizing processes for weight estimation by using the Combination Feature I and Combination Feature II as input data are shown in Figures 8(a) and 8(b), respectively.

In order to find out the best features and models for weight estimation, five models were also established based on the traditional 2D image feature and two different types of combination features. Figure 9 shows the noncontact measuring results versus the actual weight measured by digital balance charts for PLS and LS-SVM. The solid line is the regression line corresponding to the ideal, unity correlation between the predicted and reference values. The correlation coefficient, RMSEP, for prediction sets were 0.8221 and 15.0121, respectively. Figure 9(a) indicates that the PLS model based on the traditional 2D image feature has some prediction power in weight estimation. Figures 9(b) and 9(c) show the measurement results estimation by the PLS model using the two different types of combination features are plotted against the actual weight measured by digital bal-

ance. The correlation coefficient, RMSEP, for prediction sets of the PLS model with Combination Feature I were 0.6394 and 21.1072, and with Combination Feature II were 0.8315 and 14.7238. Compared to the PLS model with the traditional 2D feature, the PLS models with combination features do not show any obvious advantages in weight estimation. Figures 9(d) and 9(e) show the measurement results estimation by the LS-SVM model using the two different types of combination features are plotted against the actual weight measured by digital balance. The correlation coefficient, RMSEP, for prediction sets of the LS-SVM model with Combination Feature I were 0.8602 and 9.9556, and with Combination Feature II were 0.8234 and 11.4991. Compared to the PLS models, the LS-SVM models get more satisfied measuring results in weight estimation. Both prediction precision of the LS-SVM models could be accepted. However, considering the LS-SVM model with Combination Feature II has the strongest ability for volume noncontact estimation, and extracting both of the two types of height information is time-consuming; in real-world applications, the LS-SVM model with Combination Feature II is the preferred model for volume and weight noncontact measuring applications.

It should also be noted that the weight estimation results of PLS and LS-SVM are similar with volume estimation by using the PLS and LS-SVM models; the reason might be that the apple density could be assumed constant, and results also indicated that there was a significant positive linear correlation between the volume and weight.

5. Conclusions

A computer vision system for apple volume and weight non-contact measurement was designed and developed. Estimation of the volume and weight of apples by using 3D reconstruction, image processing, and regression methods

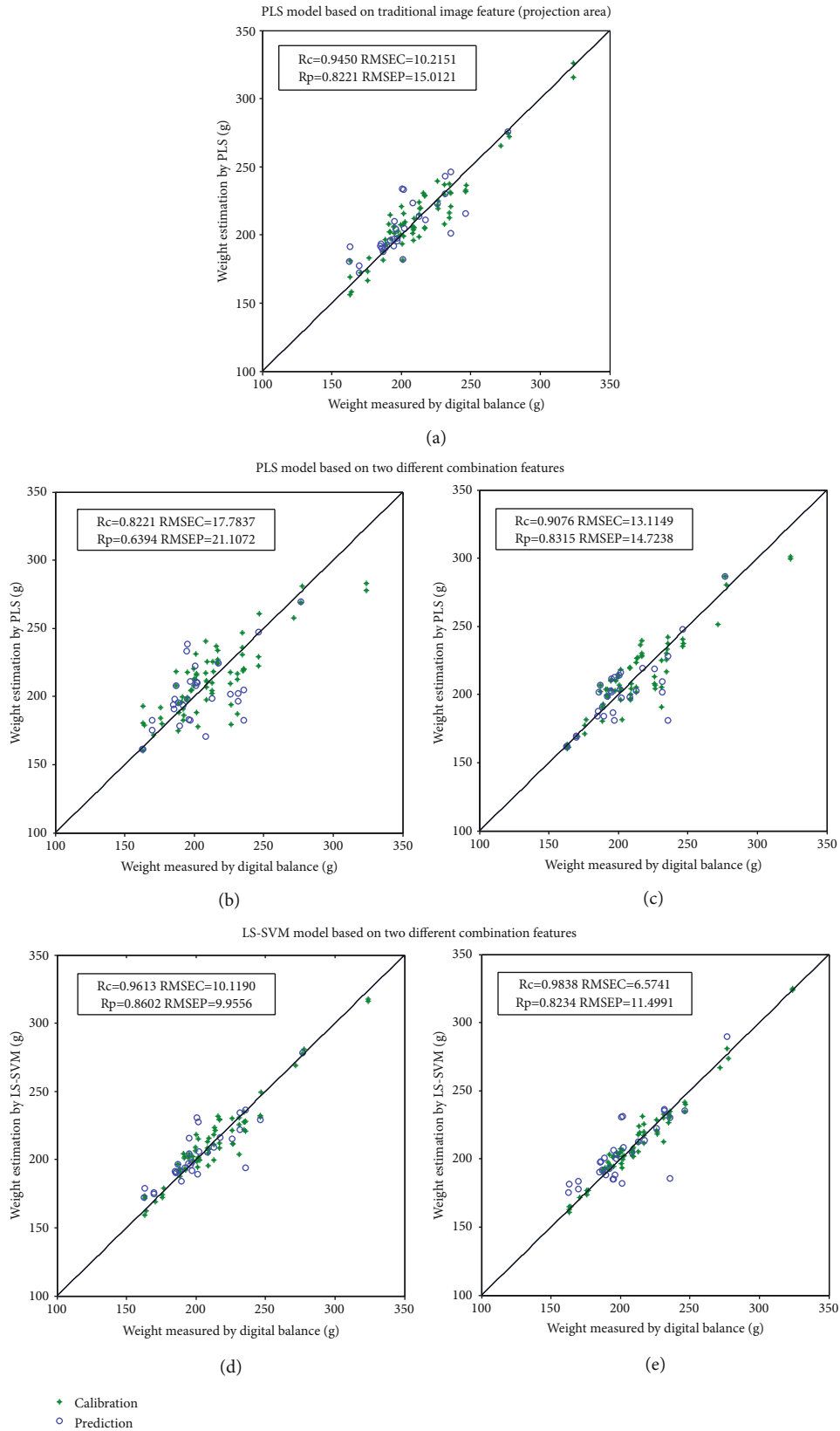


FIGURE 9: Measured versus estimated values for weight prediction by the PLS and LS-SVM models. (a) The measurement results estimation by the PLS model using the traditional projection area. (b) The measurement results estimation by the PLS model using Combination Feature I. (c) The measurement results estimation by the PLS model using Combination Feature II. (d) The measurement results estimation by LS-SVM model using Combination Feature I. (e) The measurement results estimation by LS-SVM model using Combination Feature II.

were also studied with the proposed platform of computer vision and near-infrared linear-array structured lighting system.

The 3D upper surface of apples was reconstructed by using the proposed system and triangulation method as moving the fruit on a conveyor belt. Both the traditional 2D image feature (projection area) and 3D height information (two types of fifty mean height values) were extracted by using image processing methods. Two popular regression modeling methods, PLS and LS-SVM, were developed with traditional and two types of combination features. RBF and a two-step Grid-Search technique were applied in LS-SVM models. The results indicated that the PLS model with Combination Feature II could get better prediction accuracy in volume and weight estimation than the PLS model with the traditional 2D image feature, and both LS-SVM models with two types of combination features performed better than all the PLS models. Results indicated that the 3D height information could greatly improve the prediction performance. For volume estimation, the LS-SVM model with Combination Feature II could get the best performance, and for weight estimation, the LS-SVM model with Combination Feature I could get the best performance. Considering that the LS-SVM model with Combination Feature II has the strongest ability for volume noncontact estimation, and extracting both of the two types of height information is time-consuming, in the real-world applications, the LS-SVM model with Combination Feature II could be chosen as the preferred model for volume and weight noncontact measuring applications.

The system and method developed in this study provide an alternative to the traditional methods for noncontract measurement of the volume and weight of agricultural products. The proposed system is easily constructed by using low-cost cameras without any complex calibration. The present work might be easily extended in 3D reconstruction, stem/calyx recognition, 3D shape detection, and whole surface inspection of axisymmetric agricultural products. And our future work will be focused on these extended areas by using the proposed near-infrared linear-array structured lighting vision system.

Data Availability

The data will be available on request.

Conflicts of Interest

The authors declare no competing financial interests.

Authors' Contributions

Baohua Zhang designed the study, developed the near-infrared linear-array structured light vision system, acquired the images, and analyzed the data. Jichao Huang, Baoxing Gu, and Jun Zhou wrote and edited the paper.

Acknowledgments

This work was supported by the Natural Science Foundation of Jiangsu Province (Grant no. BK20180515) and the Fundamental Research Funds for the Central Universities (project No. KYZ201325). The authors wish to thank Dr. Wenqian Huang and Dr. Jiangbo Li from Beijing Research Center of Intelligent Equipment for Agriculture for their help and guidance.

References

- [1] A. Siedliska, P. Baranowski, and W. Mazurek, "Classification models of bruise and cultivar detection on the basis of hyperspectral imaging data," *Computers and Electronics in Agriculture*, vol. 106, pp. 66–74, 2014.
- [2] A. Zdunek, J. Cybulska, D. Konopacka, and K. Rutkowski, "Interlaboratory analysis of firmness and sensory texture of stored apples," *International Agrophysics*, vol. 25, pp. 67–75, 2011.
- [3] D. Lorente, N. Aleixos, J. Gómez-Sanchis, S. Cubero, O. L. García-Navarrete, and J. Blasco, "Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment," *Food and Bioprocess Technology*, vol. 5, no. 4, pp. 1121–1142, 2012.
- [4] J. Li, W. Huang, C. Zhao, and B. Zhang, "A comparative study for the quantitative determination of soluble solids content, pH and firmness of pears by Vis/NIR spectroscopy," *Journal of Food Engineering*, vol. 116, no. 2, pp. 324–332, 2013.
- [5] J. Blasco, N. Aleixos, and E. Molto, "Computer vision detection of peel defects in citrus by means of a region oriented segmentation algorithm," *Journal of Food Engineering*, vol. 81, no. 3, pp. 535–543, 2007.
- [6] B. Zhang, W. Huang, L. Gong et al., "Computer vision detection of defective apples using automatic lightness correction and weighted RVM classifier," *Journal of Food Engineering*, vol. 146, pp. 143–151, 2015.
- [7] X. Zou, J. Zhao, Y. Li, and M. Holmes, "In-line detection of apple defects using three color cameras system," *Computers and Electronics in Agriculture*, vol. 70, pp. 129–134, 2010.
- [8] M. Hu, Q. Dong, B. Liu, and P. Malakar, "The potential of double K-means clustering for banana image segmentation," *Journal of Food Process Engineering*, vol. 37, no. 1, pp. 10–18, 2014.
- [9] S. Cubero, M. Diago, J. Blasco, J. Tardáguila, B. Millán, and N. Aleixos, "A new method for pedicel/peduncle detection and size assessment of grapevine berries and other fruits by image analysis," *Biosystems Engineering*, vol. 117, pp. 62–72, 2014.
- [10] W. Wang and C. Li, "Size estimation of sweet onions using consumer-grade RGB-depth sensor," *Journal of Food Engineering*, vol. 142, pp. 153–162, 2014.
- [11] Y. Zhang and L. Wu, "Classification of fruits using computer vision and a multiclass support vector machine," *Sensors*, vol. 12, no. 9, pp. 12489–12505, 2012.
- [12] A. Vidal, P. Talens, J. M. Prats-Montalbán, S. Cubero, F. Albert, and J. Blasco, "In-line estimation of the standard colour index of citrus fruits using a computer vision system developed for a mobile platform," *Food and Bioprocess Technology*, vol. 6, no. 12, pp. 3412–3419, 2013.
- [13] C. Garrido-Novell, D. Pérez-Marin, J. M. Amigo, J. Fernández-Novales, J. E. Guerrero, and A. Garrido-Varo, "Grading and color evolution of apples using RGB and hyperspectral

- imaging vision cameras,” *Journal of Food Engineering*, vol. 113, no. 2, pp. 281–288, 2012.
- [14] B. Zhang, W. Huang, J. Li et al., “Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: a review,” *Food Research International*, vol. 62, pp. 326–343, 2014.
- [15] A. B. Koc, “Determination of watermelon volume using ellipsoid approximation and image processing,” *Postharvest Biology and Technology*, vol. 2007, no. 45, pp. 366–371, 2007.
- [16] C. Teoh and A. M. Syaifudin, “Image processing and analysis techniques for estimating weight of Chokanan mangoes,” *Journal of Tropical Agriculture and Food Science*, vol. 35, p. 183, 2007.
- [17] S. Cubero, N. Aleixos, E. Moltó, J. Gómez-Sanchis, and J. Blasco, “Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables,” *Food and Bioprocess Technology*, vol. 4, no. 4, pp. 487–504, 2011.
- [18] H. Ishikawa and D. Geiger, “Occlusions, discontinuities, and epipolar lines in stereo,” in *European conference on computer vision*, pp. 232–248, Springer, 1998.
- [19] T. Chalidabhongse, P. Yimyam, and P. Sirisomboon, “2D/3D vision-based mango’s feature extraction and sorting,” in *2006 9th International Conference on Control, Automation, Robotics and Vision*, pp. 1–6, IEEE, 2006.
- [20] M. Omid, M. Khojastehnazhand, and A. Tabatabaefar, “Estimating volume and mass of citrus fruits by image processing technique,” *Journal of food Engineering*, vol. 100, no. 2, pp. 315–321, 2010.
- [21] H. Golnabi and A. Asadpour, “Design and application of industrial machine vision systems,” *Robotics and Computer-Integrated Manufacturing*, vol. 23, no. 6, pp. 630–637, 2007.
- [22] J. Geng, “Structured-light 3D surface imaging: a tutorial,” *Advances in Optics and Photonics*, vol. 3, no. 2, pp. 128–160, 2011.
- [23] B. Dellen and I. A. Rojas Jofre, “Volume measurement with a consumer depth camera based on structured infrared light,” *Proceedings of the 16th Catalan Conference on Artificial Intelligence, poster session*, pp. 1–10, 2013.
- [24] F. Storbeck and B. Daan, “Fish species recognition using computer vision and a neural network,” *Fisheries Research*, vol. 51, no. 1, pp. 11–15, 2001.
- [25] N. N. Mohsenin, *Physical properties of plant and animal materials. Vol. 1. Structure, physical characteristics and mechanical properties*, p. 1, 1970.
- [26] D. Unay and B. Gosselin, “Stem and calyx recognition on ‘Jonagold’ apples by pattern recognition,” *Journal of food Engineering*, vol. 78, no. 2, pp. 597–605, 2007.
- [27] Y. Shao, Y. He, and J. Mao, “Nondestructive measurement of the internal quality of bayberry juice using Vis/NIR spectroscopy,” *Journal of Food Engineering*, vol. 79, no. 3, pp. 1015–1019, 2007.
- [28] D. Wu, Y. He, S. Feng, and D. Sun, “Study on infrared spectroscopy technique for fast measurement of protein content in milk powder based on LS-SVM,” *Journal of Food Engineering*, vol. 84, no. 1, pp. 124–131, 2008.
- [29] F. Liu, Y. He, and L. Wang, “Comparison of calibrations for the determination of soluble solids content and pH of rice vinegars using visible and short-wave near infrared spectroscopy,” *Analytica Chimica Acta*, vol. 610, no. 2, pp. 196–204, 2008.
- [30] J. A. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle, “Weighted least squares support vector machines: robustness and sparse approximation,” *Neurocomputing*, vol. 48, no. 1–4, pp. 85–105, 2002.