A new method based on computer vision for non-intrusive orange peel sorting

Sajad Sabzi, Yousef Abbaspour-Gilandeh
Department of Biosystems Engineering
College of Agriculture, University of Mohaghegh Ardabili, Ardabil, Iran
e-mail: sajadsabzi2@gmail.com

Juan Ignacio Arribas
Department of Electrical Engineering
University of Valladolid, 47011 Valladolid, Spain
e-mail: jarribas@tel.uva.es

***Abstract*— As it is well-known, orange peel is used for making jam and oil. For this purpose, orange samples with high peel thickness are best. In order to predict peel thickness in orange fruit, we present a system based in image features, comprising: area, eccentricity, perimeter, length/area, blue value, green value, red value, wide, contrast, texture, wide/area, wide/length, roughness, and length. A novel identification solution based on the hybrid of particle swarm optimization (PSO), genetic algorithm (GA) and artificial neural network (ANN) is proposed. In addition, principal component analysis (PCA) has been applied to reduce the number of dimensions, without much loss of information. Taguchi's robust optimization technique has been applied to determine the optimal setting for parameters of PSO, GA, and ANN. The optimal level of factors were: Number of Neuron in first layer=7, Number of Neuron in second layer=2, Maximum Iteration=400, Crossover probability=0.7, Mutation probability=0.1, and Swarm (Population) Size=200. Results for prediction of orange peel thickness based on levels that are achieved by Taguchi method were evaluated by five performance measures: the coefficient of determination (R^2), mean squared error (MSE), mean absolute error (MAE), sum square error (SSE), and root mean square error (RMSE), reaching values of 0.8571, 0.0123, 0.0924, 1.392, and 0.1109, respectively.**

Keywords- computer vision;particle swarm optimization; peel;stochastic analysis ; thickness;Thompson orange;volume

I. INTRODUCTION

Orange is a delicious fruit consumed worldwide. It originated in Southeast Asia widely grown in Brazil, United States of America, India, Mexico, China, Spain, Italy, Iran, Egypt, and Pakistan, among others. It is consumed directly fresh or indirectly as jam or oil. According to statistics recently published, Iran is sixth citrus producer in the word [1]. Orange peel is used to make jam and oil. Orange peel can be thick or thin. Thus, oranges with higher peel thickness potentially have more oil and jam inside. But how can one anticipate orange peel thickness non-intrusively? Peel fruit grading provides information about its peel thickness. The use of machine vision techniques for the examination of vegetables and fruits has increased during the last decade. Nowadays, many producers around the world apply grading machines capable of pre-grading fruits by mass, color and size [2]. Determining a relationship between orange peel thickness and morphological characteristics may therefore be useful and applicable in peel thickness grading. [3] studied on prediction of texture characteristics from extrusion food surface images using a computer vision system, artificial neural networks and texture characteristics of 17 samples, detecting 25 different parts within one sample using a computer vision system and texture profile analysis in extruded food.

According to the linear fitting model, the hardness and gumminess score can be reflected directly by the a^* and intensity based on correlation coefficient of 0.9558, 0.9741 and 0.9429, 0.9619, respectively. [4] predicted firmness and soluble solids content (SSC) of blueberries using hyperspectral reflectance imaging. A pushbroom hyperspectral imaging system was used to acquire hyperspectral reflectance images from 302 blueberries in two fruit orientations (i.e., stem and calyx ends) for spectral region of 500–1000 nm. [5] studied on prediction of dissolved oxygen content in river crab culture based on least squares support vector regression optimized by improved particle swarm optimization. Authors present a hybrid dissolved oxygen content prediction model based on the least squares support vector regression (LSSVR) model with optimal parameters selected by improved PSO algorithm. In view of the slow convergence of PSO, improved PSO with the dynamically adjusted inertia weight was based on the fitness function value to improve convergence. Afterwards, a global optimizer, was employed to optimize the hyper parameters needed in the LSSVR model. [6] investigated postharvest citrus mass and size estimation using a logistic classification model and a watershed algorithm. An image processing algorithm was developed to identify fruit from images of the postharvest citrus from a commercial citrus grove. For fruit detection, logistic regression model based pixel classification algorithms was developed. To avoid misclassification due to highly saturated area on fruit and non-fruit regions, a highly saturated area recovering (HSAR) algorithm was developed that obviated the use of a "filling holes" operation.

The aim of this study was to evaluate orange peel thickness using image processing with the help of a hybrid of ANN-PSO-GA method.

II. MATERIAL AND METHODS

A. Samples

In this study 100 samples of Thompson orange were randomly selected and transferred to the laboratory of mechanics of agricultural machinery department, Razi University, Kermanshah (longitude: 7.03°E; latitude: 4.22°N), Iran. Selecting samples criteria were healthy, free from any injuries and got from an orchard in the north part of the country. True orange peel thickness was measured by using a digital caliper with a sensitivity of 0.01 mm.

B. Hardward system for imaging

The machine vision system consists of a digital camera (BOSCH, Portugal), an image acquisition system, a frame grabber (Pinnacle, China) and a PC, equipped with MatLab (ver. R2014a). The system provides digital images with a resolution of 352×288 pixels. The image acquisition system consisted of three types of lamps: fluorescent, LED, and tungsten. Lamps were arranged in three rows for each lateral plane of chamber. Fluorescent lamps have white light while tungsten bulbs have red, yellow and pink lights. Each set of LEDs have five rows consisting of a total of 40 LED lamps. LED lights color, top to bottom, were: orange, blue, red, white and green, respectively. Lamps were distributed around samples to avoid any shadow situations. A dimmer was built for high and low light intensity use. The digital camera was located in the center of image system.

C. Segmentation

In order to segment fruit from the background, the image must be changed into the red channel in RGB color space. After several trial and error, a proper threshold value was selected. Based on this threshold if red channel value is higher than green and blue channel values, that pixel is considered as fruit. Despite using best conditions for imaging, some segmentation error in the images were present (noisy pixels). Canny and Laplacian filters were employed for noise pixel removal. By Canny filtering, edges are detected with local maximum gradient of image f(x,y). Gradient is computed using a derivative of Gaussian filter. Next, two thresholds are used to identify strong and weak edges. With Laplacian filtering, edges are detected after filtering image f(x,y) using a Gaussian filter [7]. Distance between fruit sample and digital camera was selected fixed at 15 cm. For image acquisition, lamps of white LED light with a light intensity of 70.44 lux were selected.

(a) (b) (c) (d)

Figure 1. Pre-processing operations states: (a) original image, (b) segmented image, (c) binary image, (d) improved binary image

D. Features extraction in orange

Table I defines computed features from fruit images.

TABLE I. FEATURE EXTRACTION DEFINITIONS

feature	explanation
Area	The number of pixels that exist in region
Perimeter	The number of pixels around the boundary of each region
Length	The number of pixels that exist in major axis of the ellipse that has the same normalized second central moment as a region
Width	The number of pixels that exist in minor axis of the ellipse that has the same normalized second central moment as a region
Eccentricity	The ratio of distance from ellipse center to length of main axis
Roughness	Indicate variations of intensity values or gray levels in region
Contrast	The local variations of gray values
Entropy	Measure the value of random nature of texture
Red value	The number of Red pixels that exist in region
Green value	The number of Green pixels that exist in region
Blue value	The number of Blue pixels that exist in region

E. Principal component analysis (PCA)

PCA is a mathematical tool from applied linear algebra, which provides a roadmap for the way a complex dataset is reduced to a lower dimension. The steps involved in PCA are discussed below:

Step 1: subtract mean from each of the dimensions.

Step 2: compute covariance matrix; covariance is a measure for how much each of the dimensions varies from the mean with respect to each other. For two dimensions, covariance is introduced next:

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)} \qquad (1)$$

where X_i and Y_i are members of two different dimensions, $\bar{X}$ and $\bar{Y}$ are mean of data in each dimension, and *n* is the number of data samples. Covariance matrix for 3 dimensions case is shown next:

$$C = \begin{bmatrix} cov(X,X) & cov(X,Y) & cov(X,Z) \\ cov(Y,X) & cov(Y,Y) & cov(Y,Z) \\ cov(Z,X) & cov(Z,Y) & cov(Z,Z) \end{bmatrix} \qquad (2)$$

This matrix has several well-known properties:

1- Diagonal: variances of the variables
2- cov(X,Y)=cov(Y,X); hence, the matrix is symmetrical about the diagonal (upper triangular).
3- m-dimensional data will result in an $m \times m$ covariance matrix.

Step3: compute the eigenvectors and eigenvalues of the covariance matrix.

Step4: reduce dimensionality and form a feature vector. Notice that the eigenvector with the highest eigenvalue is the principal component of the dataset. Once eigenvectors are found from the covariance matrix, the next step is to order them by eigenvalues from the highest to lowest, which presents the components in the order of significance.

Step5: drive the new data: Final Data= Row Feature Vector × Row Zero Mean Data where: Row Feature Vector is the matrix with the eigenvectors in the transposed columns so that the eigenvectors are now in the rows with the most significant eigenvector at the top. Row zero mean data is the mean – adjusted data transposed; i.e. the data items are in each column with each row holding a separate dimension, [8].

III. RESULTS AND DISCUSSION

A. Hybrid PSO-GA-ANN system parameters tuning

Taguchi method was used in order to statistically calibrate the parameters in algorithm. We implement these experiments under MatLab (ver. R2014a), and run on a PC with 2.27GHz Intel(R) Core(TM) i3 CPU and 4 GB of RAM memory. Considered orthogonal array with six factors and three levels in Taguchi method is L_{27}. The orthogonal array L_{27} and results of experiment as the mean square errors (MSE) are presented in Table II. The relative percentage deviation (RPD) is used as a common performance measure to parameter setting;

$$RPD = \frac{Algorithm_{sol} - Min_{sol}}{Min_{sol}} \times 100 \quad (3)$$

Where $Algorithm_{sol}$ is result for a given algorithm and Min_{sol} is minimum result for all experiments, [9]. Results for each level are shown in Fig. 8 and 9. Fig.9 shows the mean S/N ratio for each level of the factors. The optimal level of factors A, B, C, D, E, and F become 3, 1, 2, 1, 1, and 3 respectively, being Number of Neuron in first layer(A)=7, Number of Neuron in second layer(B)=2, Maximum Iteration(C)=400, Crossover probability(D)=0.7, Mutation probability(E)=0.1, and Swarm (Population) Size(F)=200.

B. Prediction of orange peel thickness

Table III shows result for prediction OF orange peel thickness based on levels that were achieved by Taguchi method, including sum square error (SSE), mean absolute error (MAE), coefficient of determination (R^2),root mean square error (RMSE), and mean squared error (MSE), employed as performance indicators to evaluate prediction capability of the model, [10]. Fig. 2 plots peel thickness estimation over test data given by the hybrid of PSO-GA-ANN. From Fig. 2 it is clear that the proposed model of the hybrid of PSO-GA-ANN has good learning capability for samples and simultaneously achieves excellent generalization performance since the PSO-GA is a good compromise for guaranteeing improvement in both stability and accuracy, being aq suitable and effective method for predicting orange peel thickness. Regression analysis was conducted on the estimated peel thickness values with respect to the measured (true) peel thickness over the entire experiment sets. The highest R between the measured and the estimated peel thickness was 0.9258. Fig. 3 depicts the result of the regression analysis. Two tips should be considered in order to increase accuracy, first: orange peel thickness measurements should be done with high accuracy. Second: in the selection of factors and their levels in algorithm, high precision a good care is needed.

TABLE II. ORTHOGONAL ARRAY L_{27} MSE AND RPD

Factors / Experimeters	**A**	**B**	**C**	**D**	**E**	**F**	**MSE**	**RPD**
1	1	1	1	1	1	1	0.037267	146.85
2	2	2	1	1	1	1	0.04051	168.33
3	3	3	1	1	1	1	0.037167	146.19
4	1	1	2	2	2	1	0.050727	236.01
5	2	2	2	2	2	1	0.015097	0
6	3	3	2	2	2	1	0.030404	101.39
7	1	1	3	3	3	1	0.043062	185.23
8	2	2	3	3	3	1	0.043622	188.94
9	3	3	3	3	3	1	0.052355	246.79
10	2	1	3	2	1	2	0.041832	177.08
11	3	2	3	2	1	2	0.032117	112.73
12	1	3	3	2	1	2	0.047423	214.12
13	2	1	1	3	2	2	0.048598	221.90
14	3	2	1	3	2	2	0.038179	152.89
15	1	3	1	3	2	2	0.056181	272.13
16	2	1	2	1	3	2	0.057866	283.29
17	3	2	2	1	3	2	0.053948	257.34
18	1	3	2	1	3	2	0.060486	300.64
19	3	1	2	3	1	3	0.034186	126.44
20	1	2	2	3	1	3	0.046347	206.99
21	2	3	2	3	1	3	0.024718	63.72
22	3	1	3	1	2	3	0.042946	184.46
23	1	2	3	1	2	3	0.048785	223.14
24	2	3	3	1	2	3	0.050995	237.78
25	3	1	1	2	3	3	0.026258	73.92
26	1	2	1	2	3	3	0.05917	291.93
27	2	3	1	2	3	3	0.033233	120.12

TABLE III. REGRESSION AND STATISTICAL ERROR MEAN VALUES

R^2	SSE	MAE	MSE	RMSE
0.8571	1.392	0.0924	0.0123	0.1109

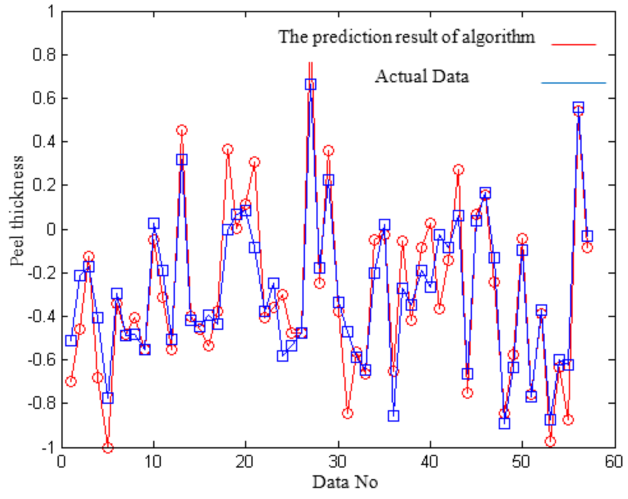

Figure 2. Peel thickness true versus predicted values based on the hybrid of ANN -PSO-GA

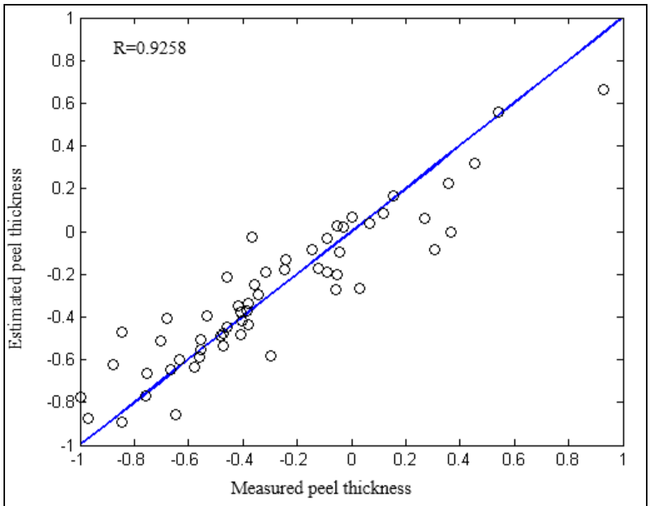

Figure 3. Regression analysis between measured (true) and estimated orange peel thickness (mm.)

IV. FINAL REMARKS: WAYS TO INCREASE ACCURACY

In this work, a hybrid method comprising PSO, GA and ANN, that is designed based on a stochastic search method, is used in non-intrusive estimation of Thompson orange peel thickness. In order to predict with high accuracy, some remarkable points are worth commenting on:

1. PCA reduces the number of dimensions, without much loss of information and this caused that computational load is reduced, and at the same time having increased accuracy and speed of calculation.
2. Clearly, hybrid of PSO-GA-ANN system, like most other searching algorithms, is mainly influenced by parameter values tuning. These parameters can be set manually or by using different setting approaches, such as full factorial experiment. This is a comprehensive approach but it would lose its efficiency by increasing the number of parameters. Thus, with Taguchi's proposed approach, it is expected that a large number of variables would be tuned through a small number of experiments.

The sorting system proposed in this study could be potentially used in the sorting industry, but to achieve a reliable (low error) sorting system, the following items should be taken into account:

1. Use of a camera with high resolution.
2. Locating the camera at a suitable distance from fruit, in such a way their real size of fruit is shown.
3. Appropriate lighting conditions, without any disturbance.
4. Suitable color as the background, omitting any light reflectance.

REFERENCES

[1] A. Adelkhani, B. Beheshti, S. Minaei, P. Javadikia, and M. Ghasemi-Varnamkhasti, "Taste characterization of orange using image processing combined with ANFIS," Measurement, vol. 46, pp. 3573–3580, 2013.

[2] J. Blasco, N. Aleixos, and E. Molto, "Machine Vision System for Automatic Quality Grading of Fruit," Biosystems Engineering, vol. 85 (4), pp. 415atio, 2003.

[3] F.H. Fan, Q. Ma, J. Ge, Q.Y. Peng, W.W. Riley, and S.Z. Tang, "Prediction of texture characteristics from extrusion food surface images using a computer vision system and artificial neural networks," Journal Food Engineering, vol. 118, pp. 426–43, 2013.

[4] G.A. Leiva-Valenzuela, R. Lu, and J.M. Aguilera, "Prediction of firmness and soluble solids content of blueberries using hyperspectral reflectance imaging," Journal Food Eng., vol. 115, pp. 91–98, 2013.

[5] S. Liu, L. Xu, D. Li, Q. Li, Y. Jiang, H. Tai, and L. Zeng, "Prediction of dissolved oxygen content in river crab culture based on least squares support vector regression optimized by improved particle swarm optimization," Computer and Electronics in Agriculture, vol. 95, pp. 82–91, 2013.

[6] J.S. Shin, W.S. Lee, and R. Ehsani, "Postharvest citrus mass and size estimation using a logistic classification model and a watershed algorithm," Biosystems Engineering, vol. 113, pp.42-53, 2012.

[7] R. C. Gonzalez, R.E. Woods, S.L, and Eddins, Digital Image Processing Using MatLab. Prentice Hall, 2004.

[8] S. K. Gauri, "Optimization of multi-response dynamic systems using principal component analysis PCA-based utility theory approach," International Journal of Industrial Engineering, vol. 5, pp. 101-114, 2014.

[9] E. Shokrollahpour, M. Zandieh, B. Dorri, "A novel imperialist competitive algorithm for bi-criteria scheduling of the assembly flowshop problem," Int. Journal Production Research, vol. 49, pp. 3087–3103, 2011.

[10] S. Sabzi, P. Javadikia, H. Rabani, and A. Adelkhani, "Mass modeling of Thompson orange with ANFIS and SPSS methods for using in machine vision," Measurement, vol. 46, pp. 3333–3341, 2013.