Non-intrusive image processing Thompson orange grading methods

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*—A key issue in fruit export is classification and sorting for marketing. In this work image processing techniques are used to grad Thompson orange fruit. For this purpose, fourteen parameters were extracted, comprising area, eccentricity, perimeter, length/area, blue value, green value, red value, width, contrast, texture, width/area, width/length, roughness, and length. Adaptive neuro fuzzy inference system (ANFIS), linear, and nonlinear regression methods were used. Based on results, mean square error (MSE), sum squared error (SSE) and coefficient of determination (R^2) were 3.47e-08, 3.47e-07, 0.988 (ANFIS), 51.33, 4927.59, 0.866 (linear reg.) and 64.85, 6092.5, 0.832 (non-linear reg.), respectively. ANFIS model was shown as the best fit model based on previously listed performance evaluation criteria.**

Keywords-ANFIS;machine vision;grading;image processing;linear regression; non-linear regression;orange; sorting; volume

I. INTRODUCTION

Fruits are attractive and nutritional foods, because of their shape, unique taste and smell, color, and richness in minerals, vitamins and other beneficial components [1].

Orange is certainly one of the most popular fruit in several countries. Orange fruit, which is really meaty, is harvested ripe and consumed as oil, jam or fresh. It has several nutrients such as flavonoids, carotenoids, sugar and C vitamin. Based on published statistics, Iran is sixth citrus producer around the world [2].

There are several methods used in grading, but nowadays, non-destructive methods have more importance. One of the non-destructive grading methods is computer or machine vision. This method consists of both hardware and software parts. There are some adjustable objects in hardware part, such as camera, light source, type of light, the light source intensity and height of camera. To reduce error, these parameters must be set up correctly. Software part includes programming in order to properly extract and model product characteristics for use in grading. [3] used machine vision to estimate quality of seeds of tomato, cucumber, aubergine and pepper based on leaf area; a decision method and a methodology were improved to watershed segmentation for overlapping leaf images. The relative identification accuracy of seedling quality was 98.6%, 96.4%, 98.6% and 95.2% for tomato, cucumber, aubergine and pepper, respectively. [4] proposed an image segmentation technique for apple grading and sorting using Otsu's method and SVM. This technique automatically modified the classification hyperplane that was computed by using linear SVM requiring minimum time and training. The segmentation error fluctuated from 3% to 25% for fixed SVM, while adjustable SVM reached reliable and exact results for each training set, with a segmentation error of less than 2%. [5] investigated an expert system based on computer vision to estimate the content of impurities in olive oil samples. In this work a system based on computer vision and pattern recognition was developed, in order to classify the content of impurities of the olive oil samples in three sets, indicative of the goodness of the separation process of olive oil after its extraction from the paste. Beginning from the histograms of the channels of the Red–Green–Blue (RGB), CIELAB and Hue–Saturation–Value (HSV) color spaces, an initial input parameter vector was constructed via feature extraction previous to classification. Various linear and non-linear extraction techniques were computed, and classifiers used included SVMs and artificial neural networks (ANNs). Best classification rate was 87.66%, obtained using kernel principal components analysis and a third grade polynomial kernel SVM. The best result using ANNs was 82.38%, yielded by the use of principal component analysis with a multilayer perceptron. [6] investigated the application of hybrid image characteristics for rapid and non-invasive classification of raisin. In this study a new method for the non-invasive classification of raisins is presented based on hybrid image characteristics: color, morphological and texture. A whole set of 74 characteristics (30 color, 8 morphological and 36 textural) were extracted from RBG images. Seven types of models were recognized based on various characteristic sets. There were three types of models recognized based on single characteristic set, three types of models recognized based on the combination of two characteristics set, and one model recognized based on the combination of all characteristic sets. Results showed that characteristic combination is effective to enhance the accuracy of raisin classification. It was concluded that the varieties of raisin could be accurately classified based on RGB image characteristics and the combination of color, morphological and texture features was an accurate way to enhance classification performance. Machine vision is one of the most effective tools for measuring external features such as color homogeneity, color intensity, bruises, shape, size, physical attributes of fruits and irregular-shaped objects and stem identification, [7]. [8] used image processing techniques to compute the volume and projected area of orange. They presumed each orange has been composed of several right elliptical cones. In [9] an image processing method is used to estimate the surface area and the volume of limes, lemons and peaches. [10] uses a machine vision approach for estimating the volume of irregular-shaped agricultural products using radial projections. In line with previous review of the state of the art, we focus now on orange grading.

In this study used a computer program, written in MatLab (ver. R2014a) for processing and analyzing the captured digital images from oranges. Fourteen parameters were extracted from Thompson orange such as: *area, eccentricity, perimeter, length/area, B_ave, G_ave, R_ave, wide, contrast, texture, wide/area, wide/length, roughness,* and *length*.

II. Preliminaries

A. Input data samples

100 samples 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 included healthy and free from any injuries oranges, got from orchard in North of the country. Fruit mass was measured by using a digital balance with an accuracy of 0.01 gr. There are different methods to determine volume including geometric mean diameter, water displacement method (WDM), and gas displacement method, [7]. Here, volume was determined WDM method. A sinker was used for the immersion since Thompson orange was lighter than water. Volume was calculated using (1), [11]:

$$V = W / \rho \qquad (1)$$

where, V is the volume (cm.^{3}), W is mass of displaced water (kg.), ρ is water density (kg. cm.$^{-3}$).

B. Framework description

This study was done off-line in lab. There are several stages in this method:

1- Design and development of imaging chamber.
2- Determination of the best condition for imaging.
3- Development of a program in MatLab to extract several features from orange.
4- Development of volume prediction model of orange based on ANFIS, linear and nonlinear regression models.

C. Image acquisition system

The proposed system consists of an image acquisition system, digital camera (BOSCH, Portugal) a frame grabber (China) and a PC equipped with MatLab. The camera provided images with a resolution of 352×288 pixels, appropriate for potential use in an on-line real industry conditions application. There are three lamp types in this chamber: LED, fluorescent 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 LED has five rows consisting of 40 LED lamps each. The color of LED light were orange, blue, red, white and green, from top to bottom. Lamps were homogenously distributed around samples to avoid any shadow.

D. Determination of the best condition for imaging

The camera captures image from fruit samples, and transfers to the PC through the video capture card. The image is digitized and stored into four user-defined buffers in RGB color coordinates and gray scale, for further analysis.

Five colors for background were proved: orange, white, black, brown and cream. After imaging in different states, minor and major diameters and surface area of samples were measured using caliper and planimeter respectively. Using algorithms written in MatLab, we computed minor and major diameters, surface roughness, surface area, ratio of major diameter to surface area, ratio of minor diameter to surface area, center of mass and average orange RGB, as well as segment orange (foreground) from the background.

Foreground pixels remaining in the background after application of filter are known as noise pixels. Total error between the actual values and algorithm output were calculated using (2) and (3):

$$E^2 = (A-a)^2 + (B-b)^2 + (C-c)^2 + N^2 \qquad (2)$$

$$e = \sqrt{E} \qquad (3)$$

where A is the ratio of major diameter to surface area (estimated), a ratio of major diameter to surface area (true value), B ratio of minor diameter to surface area (estimated), b ratio of minor diameter to major diameter (true), C ratio of minor diameter to major diameter (estimated), c ratio of minor diameter to major diameter (true), and N number of noise pixels in the background, [2].

Best distance between sample and digital camera was set to 10 cm. by trial and error, and fluorescent lamp with light intensity of 80.65 lux was selected (Fig. 1). A black cardboard was used as background surface to make segmentation task easier.

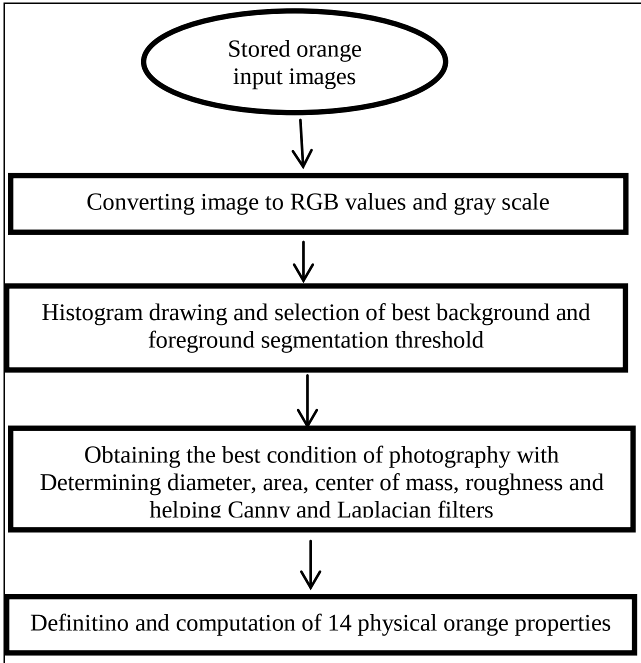

Figure 1. Methodology adapted in this research

To separate fruit from background pixels, image segmentation was used (Fig. 2). Peaks in Fig. 2 histogram correspond to background and fruit (foreground), left to right. If these two peaks are set as much away as possible from each other, the segmentation results easier. Thus, red component image has more favorable conditions for fruit to background segmentation. Canny and Laplacian filters were used for noise removal. By a Canny filter edges are identified with local maximum gradient of image data f(x,y). Gradient is calculated using a derivative of a Gaussian filter. Two different thresholds were used to identify strong and weak edges. In Laplacian filter, the edges are identified after filtering image f(x,y) using a Gaussian filter [12].

E. *Orange feature extraction*

Table 1 summarizes how orange image features are defined.

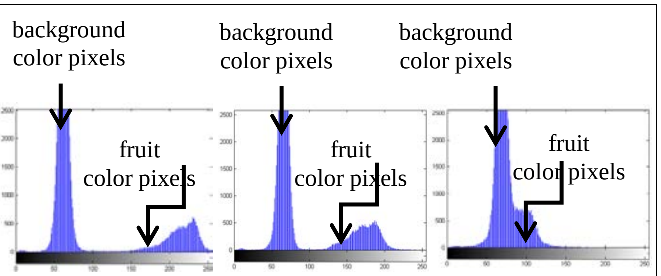

Figure 2. Image segmentation in the RGB color space: R, G, and B channels

TABLE I. FEATURE EXTRACTION DEFINITIONS

feature	explanatory definition
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 moments as the region
Width	The number of pixels that exist in minor axis of the ellipse that has the same normalized second central moments as the 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

III. DATA PROCESSING

A. *Pearson's correlation*

Pearson's correlation is a method that determines correlation between two variables. Pearson's correlation range is [-1,1] where -1, 0 and 1 mean negative, no and positive correlation. Pearson's correlation is introduced next in (4):

$$P = \frac{cov(\varepsilon, \rho)}{\sqrt{var(\varepsilon) var(\rho)}} \quad (4)$$

where $var(\varepsilon)$ and $cov(\varepsilon,\rho)$ represent the variance of feature and the covariance between feature ε and ρ, respectively, [13]. Table 2 shows Pearson's correlation for each extracted feature.

B. *Thompson orange volume modelling*

ANFIS, linear and nonlinear regression models were applied to determine the volume of orange by an image processing algorithm that written in MatLab. Several models for orange volume based on ANFIS, linear and nonlinear regression, are presented in Tables 3, 4 and 5, respectively. As seen in Tables, ANFIS method is far better than linear and nonlinear regression statistical methods. Figs. 3 and 4 depict the result of regression analysis for ANFIS and linear regression, respectively.

TABLE II. PEARSON'S CORRELATION COEEFFICIENTES COMPUTED FOR THE VARIOUS FEATURES CONSIDERED

Features	Mean	Std.	Pearson's correlation	Sig. (2 tailed)
Red value	156.15	30.92	0.021	0.833
Width	174.59	10.87	0.777	0.000**
Green value	118.71	13.46	-0.021	0.835
Contrast	0.026	0.0016	-0.775	0.000**
Length	194.02	11.25	0.726	0.000**
Perimeter	620.68	35.44	0.811	0.000**
Area	26644.9	2901.1	0.837	0.000**
entropy	16053.3	403.63	0.053	0.602
Roughness	0.828	0.0414	0.834	0.000**
Eccentricity	0.712	0.235	-0.832	0.000**
Wide/Length	0.9	0.046	0.116	0.252
Length/Area	0.0073	0.00044	-0.074	0.165
Wide/Area	0.0066	0.00038	-0.024	0.232
Blue value	81.43	5.67	0.023	0.817

**correlation is significant at the 0.01 level (2-tailed)

TABLE III. SUMMARY OF PROPERTIES FROM VARIOUS ANFIS MODELS FOR THOMSON ORANGE INPUTS

No	Mf Input	Mf Number	Epoch Number	Mf Output	Input1	Input2	Input3	R^2	SSE	MSE
1	*pimf*	3 3 3	3	constant	contrast	area	width	0.988	3.74E-07	3.74E-08
2	*gbellmf*	3 3 3	20	linear	length	area	width	0.77	2873.21	95.77
3	*pimf*	2 2 2	50	linear	length	area	-	0.816	2714.3	90.48
4	*trapmf*	5 5 5	120	constant	length	contrast	area	0.84	1726.85	57.56
5	*gbellmf*	2 2 2	10	constant	roughness	perimeter	area	0.875	1538.37	51.28
6	*psigmf*	3 3 3	30	constant	roughness	perimeter	contrast	0.888	1302.216	44.90
7	*gsigmf*	5 5 5	80	linear	roughness	perimeter	Eccentricity	0.756	2482.35	82.745

TABLE IV. SUMMARY OF PROPERTIES FROM VARIOUS LINEAR REGRESSION MODELS FOR THOMSON ORANGE INPUTS

No	Input1	Input2	Input3	Regression equation	R^2	SSE	MSE
1	contrast	area	width	V=0.433c+0.006a+8170.97w-341.36	0.868	4927.59	51.33
2	length	area	width	V=-2.38l-2.88a+0.021w+506.611	0.615	18184.22	189.42
3	length	area	-	V=70.371l+.0638a+91.938	0.678	13184.95	135.93
4	length	contrast	area	V=98.172l+3.356c+25.801a+65.714	0.723	11260.76	119.79
5	roughness	perimeter	area	V=145.043r-74.645p+5.697a+91.589	0.865	5012.919	52.218
6	roughness	perimeter	contrast	V=0.733r+5827.645p-108.017-393.85	0.875	4636.848	48.301
7	roughness	perimeter	eccentricity	V=-0.956r+0.391p+3.59e-128.86	0.71	13672.192	142.419

TABLE V. SUMMARY OF PROPERTIES FROM SOME NONLINEAR REGRESSION MODELS FOR THOMPSON ORANGE INPUTS

No	Input1	Input2	Input3	Regression equation	R^2	SSE	MSE
1	contrast	area	width	V=23.2ln(w)	0.153	31531.4	328.5
2	area	length	width	$V=-2\ln(a)+32.7\ln(l)-2\ln(w)+\ln(7.2\times10^{-9})$	0.178	60870.4	310.6
3	length	area	-	V=3ln(l)+2.36ln(a)+ln(3.2)	0.11	45214.3	423.2
4	length	contrast	area	V=2.1ln(a)+ln(5.23)	0.09	52326.3	523.2
5	roughness	perimeter	area	$V=251.7\ln(r)+24.97\ln(p)+7.2\times10^{-12}\ln(w)$	0.824	6450.6	68.62
6	roughness	perimeter	contrast	V=310.9ln(r)+110.5ln(p)+95.9ln(a)+ln(1.69)	0.832	6092.5	64.85
7	roughness	perimeter	eccentricity	V=254.6ln(r)+25.6ln(p)+1.5ln(c)+ln(0.105)	0.825	6390.2	67.98

IV. POTENTIAL APPLICATION TO REAL CASE STUDY

A. *Proposed system for on-line orange grading*

Machine vision provides the ability to grad automatically, without human intervention. Artificial intelligence techniques are among the best methods in data analysis. So, using one of artificial intelligence methods and image processing, it can be launched a fully automated sorting system with low error percentage. Benefits of this sorting system are: 1- increased sorting speed; 2- increased accuracy and costs cut due to lack of man work. The proposed sorting system prototype is depicted in Fig. 5. Thompson oranges with different volumes are placed on conveyor belt and the camera takes images from them. Captured Images are then transferred to the computer for analysis by an algorithm. After analysis, fruits are divided into three mass groups: *small, medium and large* (Table 6). Three lines are built to transfer oranges with different volumes (*small*, *medium*, and *large*) to corresponding packaging areas. This grading layout might be designed and fabricated for packaging of orange in the relevant food industries.

V. SUGGESTIONS FOR HIGH ACCURACY ESTIMATIONS

In this work, ANFIS method that was based on a stochastic search method and image processing were used for predicting the volume (mass) of orange. To predict with high accuracy, at least two points should be taken into account:

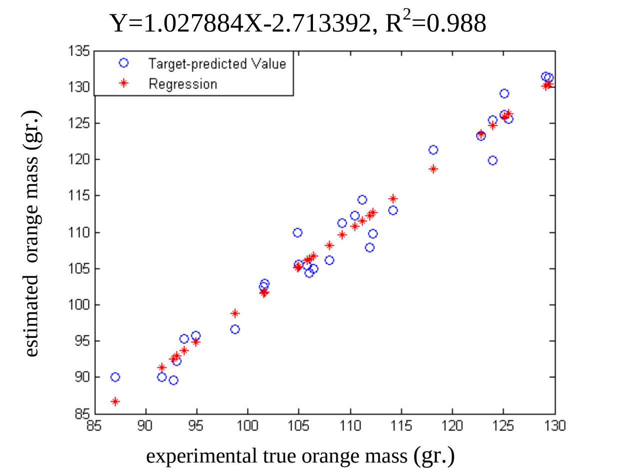

Figure 3. Experimental true versus ANFIS predicted orange mass (gr.)

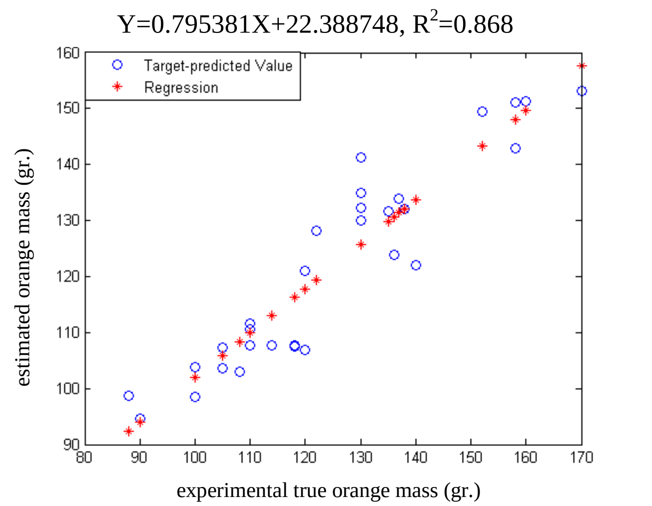

Figure 4. Experimental true versus linear reg. estimated orange mass (gr.).

1. Orange volume measurements should be done with high accuracy.
2. ANFIS method modelling must do five adjustments: membership input function, membership function number of inputs, membership function number of outputs, optimization method, and epoch number. Only whenever previous five adjustments are correct and optimized, grading mass model is accurate.

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

1. Use of a camera with a high resolution.
2. Install the camera at a suitable fix distance from the fruit in such a way that the real size of the fruit is shown.
3. Provide appropriate lighting conditions.
4. Provide a suitable background color and avoid any light reflectance.

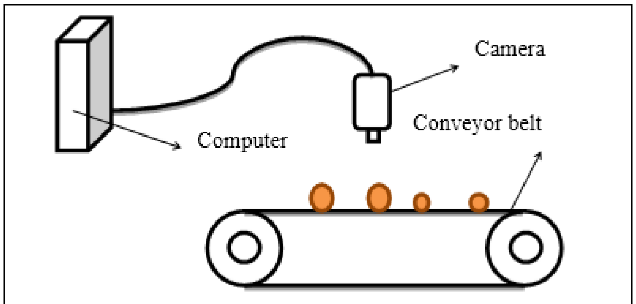

Figure 5. Layout of a computer vision orange sorting system prototype

TABLE VI. MEAN AND STANDARD DEVIATION OF MASS GROUPS (GRAMS)

Mean and std. orange mass (gr.)			
group	*small*	*medium*	*large*
Mean	101.37	126.98	152.43
standard deviation	6.99	9.15	8.42

REFERENCES

[1] A. Cassano, E. Drioli, G. Galaverna, R. Marchelli, G.D. Silvestro, and P. Cagnasso, “Clarification and concentration of citrus and carrot juices by integrated membrane processes,” J FOOD ENG vol. 57,pp. 153-163, 2003.

[2] 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.

[3] J.H. Tong, J.B. Li, and H.Y. Jiang, “Machine vision techniques for the evaluation of seedling quality based on leaf area,” biosystems engineering, vol. 115, pp. 369-379, 2013.

[4] A. Mizushima, and R. Lu, “An image segmentation method for apple sorting and grading using support vector machine and Otsu’s method,” Computers and Electronics in Agriculture, vol. 94, pp. 29-37, 2013.

[5] P.C. Marchal, D.M. Gila, J.G. García, and J.G. Ortega, “Expert system based on computer vision to estimate the content of impurities in olive oil samples,” Journal of Food Engineering, vol. 119, pp. 220–228, 2013.

[6] S. Wang, K. Liu, X. Yu, D. Wu, and Y. He, “Application of hybrid image features for fast and non-invasive classification of raisin,”, Journal of Food Engineering, vol. 109, pp. 531–537, 2013.

[7] M. Omid, M. Khojastehnazhand, and A. Tabatabaeefar, “Estimating volume and mass of citrus fruits by image processing technique,” Journal of Food Engineering vol. 100, pp. 315-321, 2010.

[8] M. Khojastehnazhand, M. Omid, and A. Tabatabaeefar, “Determination of orange volume and surface area using image processing technique,” Int. Agrophysics, vol. 23, pp. 237-224, 2009.

[9] T.Y. Wang, and S.K. Nguang, “Low Cost Sensor for Volume and Surface Area Computation of Axisymmetric Agricultural Products,” J. Food Eng vol. 79, pp. 870-877, 2007.

[10] J.D. Eifert, G.C. Sanglay, D.J. Lee, S.S. Sumner, and M.D. Pierson, “Prediction of raw produce surface area from weight measurement,” J. Food Eng vol. 74, pp. 552–556, 2006.

[11] N.N. Mohsenin, Physical properties of Plant & Animal Materials, Gordon & Breach Sci.publ, New York, 1986.

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

[13] I.H. Laradji, M. Alshayeb, and L. Ghouti, “Software defect prediction using ensemble learning on selected features,” Information and Software Technology , unpublished.