Computers and Electronics in Agriculture 127 (2016) 395–405

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Design of an automatic apple sorting system using machine vision

M.M. Sofu[a,*], O. Er[b], M.C. Kayacan[a], B. Cetişli[b]

[a] Engineering Faculty, Mechanical Engineering, Süleyman Demirel University, Isparta, Turkey
[b] Engineering Faculty, Computer Engineering, Süleyman Demirel University, Isparta, Turkey

ARTICLE INFO

Article history:
Received 24 November 2015
Received in revised form 1 April 2016
Accepted 29 June 2016
Available online 6 July 2016

Keywords:
Machine vision
Online fruit classification
Apple

ABSTRACT

This study proposes an automatic apple sorting and quality inspection system, which is based on real-time processing. Golden and Starking Delicious, and Granny Smith apple cultivars are sorted into different classes by their colour, size and weight. It also detects apples affected by scab, stain and rot.

The proposed system consists of a roller, transporter and class conveyors combined with an enclosed cabin with machine vision, load cell and control panel units. The roller and transporter conveyors have two channels.

In order to analyze the visual properties of apples, two identical industrial colour cameras are set on the roller conveyor. Four images of any apple rolling on the conveyor can be captured and processed using image processing software in 0.52 s. As a result, the proposed machine can sorted averagely 15 apples in per second using two channels, in real time.

In the experimental studies, the system design was tested using three different conveyor band velocities and three apple cultivars to sort and inspect 183 samples with an average sorting accuracy rate of 73–96%.

1. Introduction

Nowadays, the fruits must be of a certain standard in order to be efficiently offered to better quality markets. Various automatic classification machines are used to ensure these standards are met. Especially apple fruit harvested with high tonnage, are needed for automatic classification machine with the aim of increasing the economic value. The apple fruit, in particular, has a very wide range of cultivars. Therefore, apples have different colours and dimensions, must be classified in order to be sold to the market as a better quality product. The quality of the apples is determined according to the colour, weight, dimension and their defects.

In any automatic apple sorting system mainly consists of machine vision, conveyor band, separator, and classifier. So, it has mechanical, electrical, electronics and software parts. The first processing about apple is done on machine vision section. Generally, the systems used for apple classification do not operate in real-time, take an image of a single apple under the camera (Sabliov et al., 2002; Nicolai et al., 2006). The apparatus used in real-time image processing studies obtain faster and more effective classification results (Fattal et al., 2008).

There are many studies about apple sorting using colour, size and stain features. The most important characteristic affecting the quality of the apple classification is stain and decay. Most studies performed are concerned with this area (Du and Sun, 2004). One of the reasons for the low percentage of accurate classifications of decayed and stained apples is that the natural geometry of the calyx and stem parts of the apples are perceived as being decayed or stained (Penman, 2001). Li et al. (2002) used neural networks (NNs) to solve this problem. However, their proposed system processed the apples with low tonnage capacity in real-time sorting because of their cumbersome structure. Pla et al. (2001) performed a classification in their study with a colour map created by using a colour classification in a classification mechanism working in real-time that is known as a Look up Table (LUT).

In many simple colour classification studies, as in the study carried out by Yam and Papadakis, 2004, the classification is made by using different colour spaces. In the prospective classification studies on the fruits, near infrared (NIR) or mid-infrared (MIR) hyperspectral imaging techniques and hardware were also used. This hardware and these techniques offer information about the aroma, sugar, juiciness ratio and internal structure disorders of the apples (Nicolai et al., 2006; Baranowski et al., 2013).

To distinguish the apples from the background is a very important problem due to non-uniform lighting of apple surfaces. Mizushima and Lu (2013) solved this problem using support vector

* Corresponding author.
E-mail addresses: mehmetsofu@sdu.edu.tr (M.M. Sofu), orhaner@sdu.edu.tr (O. Er), cengizkayacan@sdu.edu.tr (M.C. Kayacan), bayramcetisli@sdu.edu.tr (B. Cetişli).

http://dx.doi.org/10.1016/j.compag.2016.06.030
0168-1699/

machines and a proposed adaptive segmentation is employed using the Otsu algorithm. Zhang et al. (2015) used automatic light correction for detection of defects. The other most widely used classification criteria is the size. In theory, apples are classified as Class I, Class II and extra as determined in European standards (Anonymous, 1989). But these standards could be changed according to costumer demands.

There are different classification methods for apple sorting such as neural network (Unay and Gosselin, 2002; Li et al., 2002; Bhatt and Pant, 2015), support vector machines (Mizushima and Lu, 2013), decision tree (Kavdir and Guyer, 2008), rule based (Wen and Tao, 1999) and statistical based tree (Kavdir and Guyer, 2008) methods.

In this study, the automatic apple sorting system was designed and then implemented in hardware and software. Performance of the developed system was tested by using a total of 183 samples of Golden and Starking Delicious, and Granny Smith cultivar apples with different geometries and colour, and the results were published. Apples was classified by stain, colour, weight and size characteristics to the duration of the experiment.

2. Materials and methods

The apparatus used in general fruit classification is made by simple image processing equipment (Chen et al., 2002). In this study, the mechanical and conveyor parts were designed and realized by Gençgüçsan firm. Then we set up machine vision system on the main conveyor and we coded programs to control the systems and to process the images at CAD-CAM Research & Application Center of Suleyman Demirel University. The machine is over a formal ground, and in a closed area. Therefore, daylight and humidity are not affected to machine. Although, the enclosed cabin is decreased the environmental negativities.

The developed system was tested by using a total of 183 samples of Golden and Starking Delicious, and Granny Smith cultivar apples. The apples are processed in this apparatus by two cameras across two channels. The captured images were processed and the apples distinguished from the background using the K-means algorithm. The stain, defects, scab, stem and calyx were detected from the binary apple image. The apples could be classified as small, normal and large according to size and weight, or classified as light and dark according to colour. We also sorted apples as defective and non-defective. The C4.5 algorithm was preferred as the common classifier due to its simplicity and rule-based structure. What's new in this study, the two classification requirement at the same time as the software is able to do online. Stain classification feature can work together with the colour and size of the property.

2.1. Hardware

The proposed automatic apple sorting system is shown in Fig. 1. There are two different conveyors. The first is used for recognition of the apples. For this, a machine vision system was set up in an enclosed cabin. Two Charge-Coupled Device (CCD) cameras and lighting system were also positioned in the cabin.

The first conveyor rotates and shifts the apples along the two channels as in Fig. 2. In this way, all sides of the apple can be monitored.

The classified apples are transferred to the transporter conveyor using a brush. In the transporter conveyor, the apples are located in special bowls in order to maintain the location of the classified apples. At the beginning of the transporter conveyor, two load cells are placed under the apple bowls to measure the apple weights. When the apple bowls arrive at the classification slides, the bowls are open and the apples are slowly dropped into these slides. The classification slides, which are five conveyors, are also driven by motors. The motors, bowl opening and closing magnets, counting, triggering and load cell sensors are controlled by a Programmable Logic Controller (PLC). The PLC also communicates with a personal computer (PC).

The PC is also used for image capturing, acquisition and software processing. Software flow diagram is shown in Fig. 3.

The classification data of the apples is sent to the PLC via a parallel port. The properties of the machine vision and control elements are shown in Table 1.

After the machine vision system reaches the decision about the apple classification, the apples are transferred to the transporter conveyor. At this time, the PC sends the classification data of the apples to the PLC via a parallel port without losing the apple location. The PLC also uses the load cell information, and then opens the bowls according to the class of the apple as in Fig. 4.

The bowl system allows the apples to be sorted into the classes by opening them using the trigger information sent by the PLC to 48v electromagnets.

2.2. Image acquisition

The main problem in apple sorting is to achieve the synchronization of the camera and conveyor. The other problem is to capture four images of each apple in the enclosed cabin. Therefore, an encoder is used to trigger the camera. The camera and PC easily communicate via a USB 2.0 port. A MATLAB Image Acquisition Toolbox is used to acquire the video signal of the camera. In this way, use of a frame grabber is unnecessary. The captured images are shown in Fig. 5.

Deciding the apple classification is made after the processing of the fourth image. In this time, eight apples can be processed within the two channels. The processed images and apple features are buffered in memory.

The position of apples under the camera with different processing times can be demonstrated in a matrix form $\mathbf{R}$ as,

$$\mathbf{R} = \begin{bmatrix} \mathbf{D}_{t+3} & \mathbf{C}_{t+2} & \mathbf{B}_{t+1} & \mathbf{A}_t \\ \mathbf{E}_{t+3} & \mathbf{D}_{t+2} & \mathbf{C}_{t+1} & \mathbf{B}_t \\ \mathbf{F}_{t+3} & \mathbf{E}_{t+2} & \mathbf{D}_{t+1} & \mathbf{C}_t \\ \mathbf{G}_{t+3} & \mathbf{F}_{t+2} & \mathbf{E}_{t+1} & \mathbf{D}_t \end{bmatrix}, \quad (1)$$

where $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}, \mathbf{F}, \mathbf{G}$ denote the different apples. The subscript t denotes the processing time. In this way, all apples have four successive pictures.

Image extraction is completed in three steps. The camera shoots sixteen photographs in a second on the real-time working conveyor band.

The first step is triggering. A triggering solution is created to always keep the apples in the same position and to use the memory efficiently. Triggering catches the proper photograph from the sixteen pictures taken per second using a token with any conveyor band velocity. The chain connection located on the band is painted with a white colour in front of the camera for this reason.

The second step is the apple extraction from the captured image. In this way, all the surface of the apple is processed.

The third step is labelling each apple in the image. In addition, it is controlled to allow for an apple image to be present or not. The labelled apple with t index is processed by the software and its information, including the four surfaces, is stored in a matrix. The colour, detecting the maximum size, and the number and size of defective regions from the images are the main information used to classify the apples. At the end, the load cell and image information are combined to give a decision about specific apples.

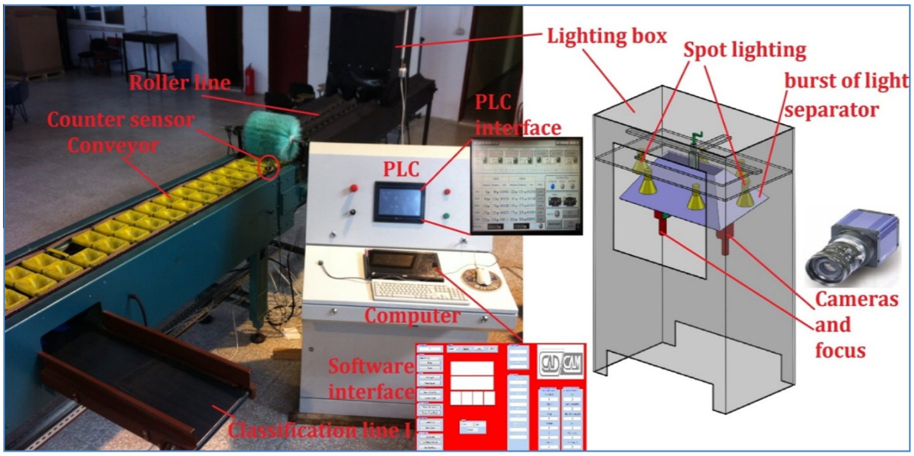

Fig. 1. The automatic apple sorting system.

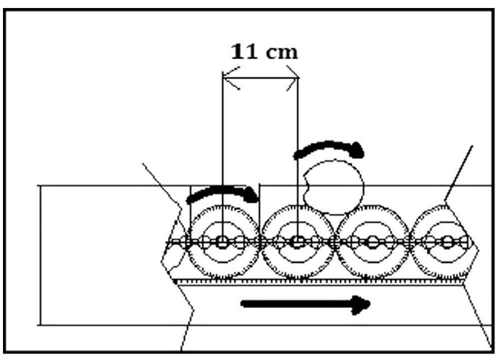

Fig. 2. The progress and return of apples.

2.3. Apple features

Certain features are very important for sorting apples successfully. In industrial applications, the colour, size and weight of apples are generally the features used. The colour and size features can be obtained with image processing.

2.3.1. Colour

Most of the time, the producers never mix the different apple cultivars when the apples are harvested. However, the light and dark colour detection could be necessary for any cultivar of apples.

The RGB colour model is used in the colour classification. Leemans et al. (1997) describe the surface colour classification of the apples. The numerical values of the colour tones of the surface colours of the apples are used for classifying the colour of different apples.

The studies performed on the surface colour shades of apples show that the classification can be made by looking at the red colour density in the red apples of Jonagold and Starking cultivars. Furthermore, the red colour density helps different yellow apples to be classified as light and dark. The density difference of the red colours is used in Golden cultivar. The difference between the blue colour densities helps to classify the Granny Smith cultivar.

Özongun et al. (2014) reported in their study that differences are observed in the perceived colour shades of the same cultivar of apples depending on environmental factors such as altitude and sunlight. Altitude is prevented from being a problem in this study because the camera is fixed but there is a slightly different situation with the light. Sunlight can enter the enclosed cabin of apples. A tarpaulin is attached to where the apples enter to reduce the light input but some light does enter. It was tried to be prevented by lack of hardware via software.

In this study, the differences between red, green and blue colours are used. In this way, the recognition rates were increased.

2.3.2. Size

The size of apples is determined from the binary image as seen in Fig. 6. The maximum value b_{max}, which is calculated from the height b_1 and width b_2 of four images, is accepted as the diameter of apple. The pixel value of lengths can be converted into the centimetre value B_{real}. The apples can then be sorted into three or more classes using their sizes. In this study, the apple sizes are evaluated as small, medium and big.

2.3.3. Defective regions

There are some methods of defining any stain on the apple. Arlimatti (2012) used a simple window-based method in his study. With this method, statistical features like averages are used – by dividing the apples into 5×5 pixels windows and standard deviation in the calculation of the remaining window values. Thus, the apples can be deemed defective or not. In each window,

$$\text{Average}: \quad m = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2}$$

$$\text{Standard deviation}: \quad \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - m)^2} \tag{3}$$

Correlations are used as shown in Eqs. (2) and (3) between the pixels. By calculating the average and standard deviation values, a list is created by sorting from small to big. The first few values in this list are accepted as the threshold values of each class. If bigger values than the threshold values are obtained, the apple is considered to be stained.

Colour histograms of all of the digital images are formed creating a brightness histogram or the RGB components for polygonal areas of interest or for a specific line (ROI: Region of Interest or AOI: Area of Interest). The average values and the standard deviations of these regions can be obtained easily with the related software. Thresholding is performed for the pixels having values between 0 and 255 in the image. Thus, only the ones staying between certain values or outside can be chosen (Baxes, 1994).

As a result of the polygonal area method (ROI) used in the experimental studies and the analysis performed during the study,

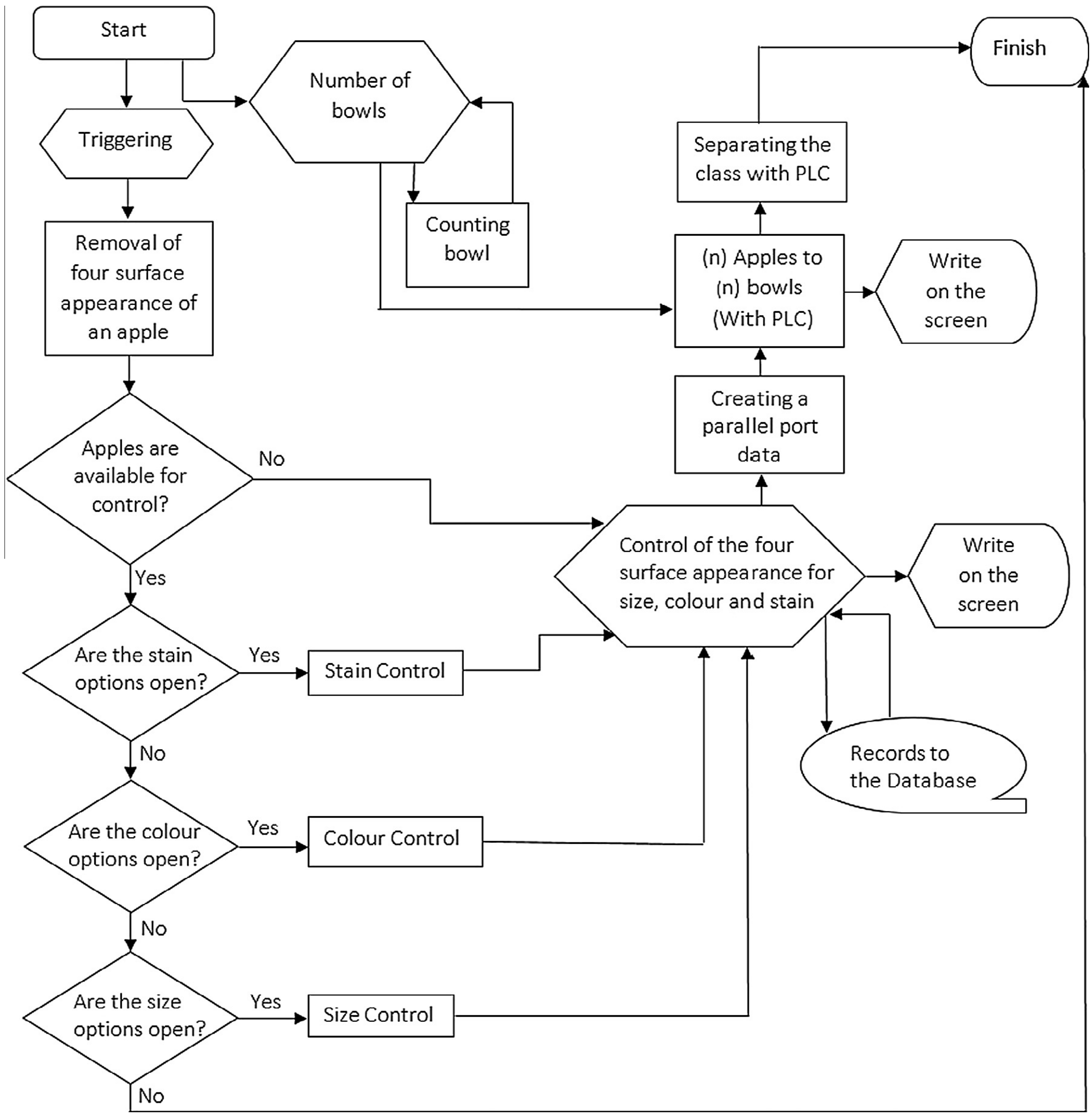

Fig. 3. Software flow diagram.

Table 1
The properties of the hardware used.

Element	Property
CCD colour camera	Firm: matrix vision, model: mv blue fox-124c, resolution: 1600 × 1200, frame rate: 16, connection: USB 2.0
Lens	Firm: fujinon, model: HF16HA-1B, mount: C, focal length: 16 mm
Lighting	Firm: OSRAM, type: halogen spot, 220 V, 50 W, 2700 K, white colour
Cabin	Sheet metal, 50 × 25 × 100 cm^3
PC	Firm: asus, model: P8hlx61, RAM: 8 GB, MP: Intel I7 3, 4 GHz
PLC	Firm: weintek, model: MT61001 v2ev
Load cell	Firm: VPG transducers, model: 1042, 5 kg capacity

the petal and the stems of the apples are determined to be generally of a circular form (Baxes, 1994). If there are areas having geometries other than a circular form obtained from the images, they are sensed as stains. The areas in the images are compared to each other using eccentricity values. The eccentricity is the ratio of the between width and height of the area. An area whose eccentricity is 0 is actually a circle, while an area whose eccentricity is 1 is a line segment. If this ratio is bigger than 0.6, the apples are determined as stained.

Unay and Gosselin (2002), performed studies on the images of stained Golden apples having eighteen different geometries. By dividing the areas formed by the stains into the total surface area of the apples, he found that the stains are 18.6% on average of the stained apples. He reported that the petal and the stems sections are equivalent to a 2.81% area.

Nicolai et al. (2006) performed a calculation by numbering the stains in the case where more than one stains is located on the apple images in his study. Each stain number contains a dimensional identity and a different number is assigned to each stain.

The first deterministic factor is the existence of a geometric difference among the calyx, stem and stains on the apple. In Fig. 7(a), a defective apple is shown. There are three possible defect regions. The image pixels are clustered into two groups with K-means algorithm (Jang et al., 1997). One of the groups denotes the background. The second group is the apple. In this way, a threshold value is determined to distinguish the apple from the

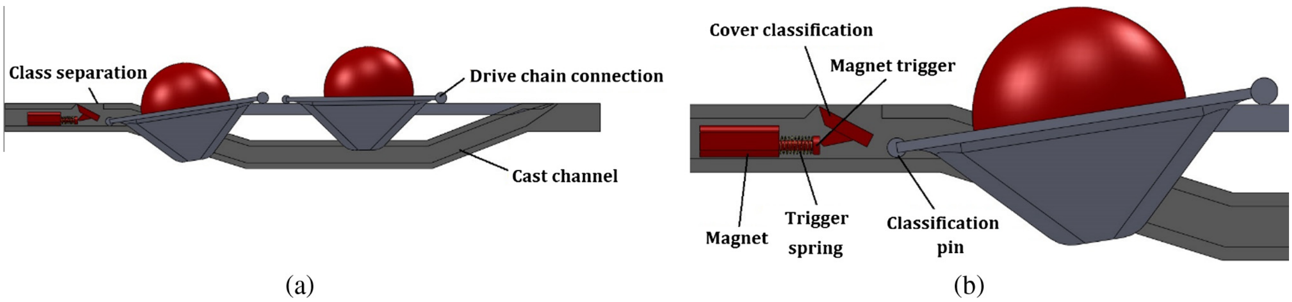

Fig. 4. Bowl system (a) arrangement allows the separation of classes of apples. (b) Triggering system of the bowl.

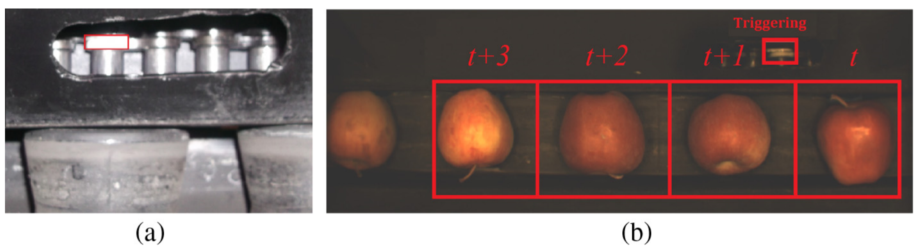

Fig. 5. Triggering system: (a) white coloured chain. (b) Captured image with visual triggering.

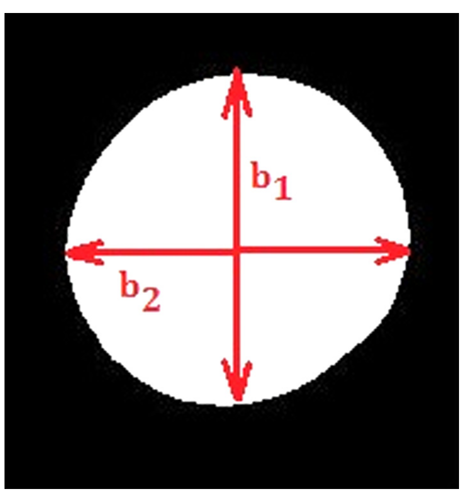

Fig. 6. The length and width of an apple.

background. Binary apple images are clustered again to detect the scab, stain, stem, and calyx regions. Fig. 7(b) denotes the binary apple image with possible defect regions. Subsequently, the geometry of each region is analyzed.

In Fig. 7(b), r_1 is the apple scab region and it has 0.7 eccentricity value. The r_2 region is the stem of the apple and it has 0.47 eccentricity value. The r_3 area is the bruised region of the apple and it has 0.88 eccentricity value. According to their eccentricity values and mean values of colour, they can be easily distinguished from the others. Results of studies of the apple as shown in Fig. 7(b), stains, scabs and rots could be distinguished from the calyx and stems.

2.3.4. Weight

The weight of apples is a very important feature used to classify apples. The weight demonstrates the apple's ripeness. In most sorting machines, the weight of apples is measured with load cell sensors. We also measured the apple weights using two load cell sensors in a real-time application. However, in this study, we also estimated the weight of apples from the image area using Least Squares Estimation (LSE).

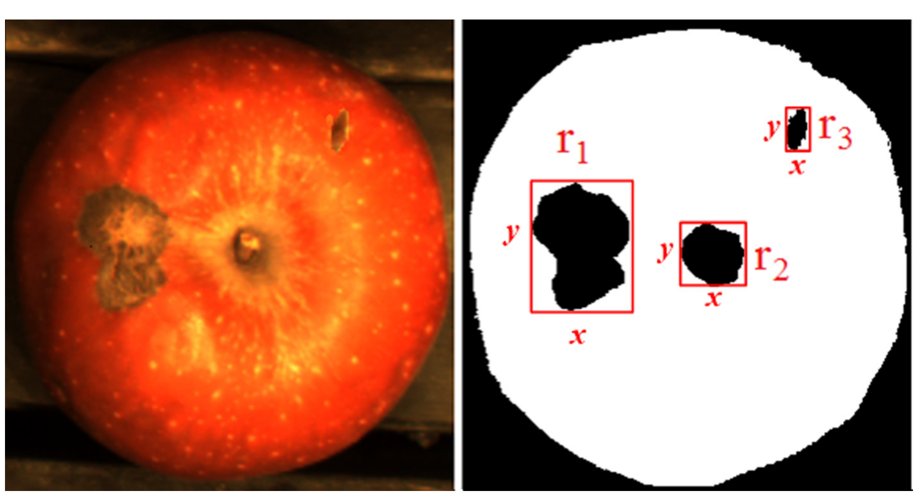

Fig. 7. Apple defects (a) original apple with the calyx and stain. (b) Binary image and detection of stains.

2.4. Least squares estimation

Curve fitting methods were used to perform weight estimation, such as interpolation and regression. There are many different types of curve fitting methods such as linear, spline and polynomial. Polynomial regression was used in this study.

The collected data is used to fit a polynomial curve function. In this study, the most commonly used method of least squares error estimation of the regression type was applied to fit polynomial parameters. In this method, the polynomial gives the relation between input and output variables. This relation can also be defined a linear or nonlinear equation system. (Cetişli and Kalkan, 2011) Eq. (4) denotes that the relationship between the input data and their target values.

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{M1} & x_{M2} & \cdots & x_{MN} \end{bmatrix} \times \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_N \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}. \tag{4}$$

This equation could be briefly rewritten as

$$\mathbf{X\theta} = \mathbf{y} \tag{5}$$

Eq. (4) can be solved in different ways such as the Gauss-elimination method or dealing with the inverse matrix. When using an inverse matrix, each side of the equation is multiplied by $\mathbf{X}^{-1}$ (inverse of $\mathbf{X}$).

$$\mathbf{X}^{-1}\mathbf{X\theta} = \mathbf{X}^{-1}\mathbf{y} \tag{6}$$

Please note that $\mathbf{X}$ must be a square and singular matrix in order to do this matrix multiplication operation. Our goal here is to find the coefficient and the last instance of the equation is:

$$\mathbf{\theta} = \mathbf{X}^{-1}\mathbf{y} \tag{7}$$

If $M \geqslant N$, $\mathbf{X}$ is a singular matrix. This means that $\mathbf{X}$ has no inverse matrix at the same time as its determinant is equal to zero. That's why Eq. (7) and its description are not adaptable or applicable. Hence, $\mathbf{X}$ needs to be in a state where it is a non-singular matrix case. $\mathbf{X}^T\mathbf{X}$ is the only state that allows it.

$$\widehat{\mathbf{\theta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{8}$$

The $\widehat{\mathbf{\theta}}$ shows the unknown coefficient in Eq. (8).

$$\mathbf{X}\widehat{\mathbf{\theta}} + \mathbf{e} = \mathbf{y} \tag{9}$$

In determining the coefficients used in the estimates, it is obvious there will be a certain error. Therefore, Eq. (4) is expanded as in Eq. (8). So, $\mathbf{e}$ and $\mathbf{X}\widehat{\mathbf{\theta}}$ indicates the error and predicted results, respectively. The solution of the system is achieved with LSE (Cetişli and Kalkan, 2011).

$$E(\mathbf{\theta}) = \sum_{k=1}^{M} (\mathbf{y}_k - \mathbf{X}_k\widehat{\mathbf{\theta}})^2 \tag{10}$$

The LSE method reduces the total square error as shown as Eq. (10). Here $\mathbf{X}_k$ shows the k-th sample of data. A linear solution is sufficient in many problems but some problems are not linear. Higher-degree polynomials may be used at this point. The solution method also applies to non-linear systems just as for the linear system.

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^N \\ 1 & x_2 & x_2^2 & \cdots & x_2^N \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_M & x_M^2 & \cdots & x_M^N \end{bmatrix} \times \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_N \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \tag{11}$$

$$\mathbf{y} = f(\mathbf{x}) = \theta_0 + \theta_1\mathbf{x} + \theta_2\mathbf{x}^2 + \ldots + \theta_N\mathbf{x}^N \tag{12}$$

A non-linear system equation and function are shown in Eqs. (11) and (12), respectively. The $f(\mathbf{x})$ is a higher-order polynomial.

In this study, firstly, the weight of the apples is measured at the beginning of the transporter conveyor under the apple bowls. Secondly, these measurements are compared with the estimated weights using LSE. Estimation was carried out using the area and minimum-maximum height of apple information.

2.5. C4.5 classification

Classifier selection is very important for automatic apple sorting. We preferred a simple classifier, which is a rule-based structure. A C4.5 decision tree is one of the most widely used inductive inference tools (Quinlan, 1993). The tree is generally constructed in a top-down manner.

The construction begins at the root node where each attribute is evaluated using a statistical test to determine how well it can classify the training samples. The best attribute is chosen as the test at the root node of the tree as (Setsirichok et al., 2012). A descendant of the root node is then created for either each possible value of this attribute – if it is a discrete-valued attribute, or each possible discrete interval of this attribute – if it is a continuous-valued attribute.

Next, the training samples are sorted to the appropriate descendant node. The process is repeated using the training samples associated with each descendant node to select the best attribute for testing at that point in the tree. This forms a greedy search for a decision tree, in which the algorithm never backtracks to reconsider earlier node choices. Although it is possible to add a new node to the tree until all samples that are assigned to one node belong to the same class, the tree is not allowed to grow to its maximum depth. A node is introduced to the tree only when there are a sufficient number of samples left from sorting. After the complete tree is constructed, a tree pruning is usually carried out to avoid data over-fitting.

A statistical test used in C4.5 for assigning an attribute to each node in the tree also employs an entropy-based measure. The assigned attribute is the one with the highest information gain ratio among attributes available at that tree construction point. The information gain ratio $GainRatio(F, \mathbf{X})$ of a feature F relative to the sample set $\mathbf{X}$ is defined as (Setsirichok et al., 2012)

$$GainRatio(F, \mathbf{X}) = \frac{Gain(F, \mathbf{X})}{Split\_Information(F, \mathbf{X})}, \tag{13}$$

where

$$Gain(F, \mathbf{X}) = Entropy(\mathbf{X}) - \sum_{s \in F} \frac{|\mathbf{X}_s|}{|\mathbf{X}|} Entropy(\mathbf{X}_s) \tag{14}$$

$$Split\_Information(F, \mathbf{X}) = -\sum_{s \in F} \frac{|\mathbf{X}_s|}{|\mathbf{X}|} log_2\left(\frac{|\mathbf{X}_s|}{|\mathbf{X}|}\right) \tag{15}$$

$\mathbf{X}_s$ is the subset of $\mathbf{X}$ for which the attribute F has the value s. Obviously, the information gain ratio can be calculated straightaway for discrete-valued attributes. In contrast, continuous-valued attributes need to be changed to discrete values prior to the information-gain ratio calculation.

3. Experimental studies

After the installing of the sorting machine, the PLC program – for controlling motors and actuators using sensors, and the machine vision program and graphical user interface (GUI) using MATLAB – was coded. The PLC control screen and sorting GUI are shown in Fig. 8 and for exempla Golden Delicious (yellow) are shown in Fig. 9, respectively.

Experiments were carried out on total of 183 apples from three different apple cultivars. The used apples are shown in Fig. 10.

Before the trials, the apples were classified by size, weight and colour and numbered in the petal sections by an apple-sorting expert. Also, the apples were sorted as either normal or abnormal according to defects. The class information is given in Table 2 for three cultivar apples.

Although there are many classification methods for apple sorting, the C4.5 algorithm is preferred because of its simplicity, rule-based structure and non-training properties. A decision tree was created according to the colour of apples and shown in Fig. 11.

In this study, the colours of any apple can be mixed with apples of other classes. For that reason, we used the sum of pixels that are the difference of colour spaces, such as $G-R, R-B, R-G$. These numbers should be positive. In Fig. 11, *SCA*, *GA* and *RA* denote Sorted Current Apple, Green Apple and Red Apple, respectively. The threshold levels of apple classes are given in Table 3.

The threshold values in Table 3 were obtained within 242 × 320 pixels of the apple image size. All of the apple cultivars were also evaluated together according to the threshold values above. The classification results for 0.1 m/s conveyor band speed are given in Table 4.

According to Table 4, while the Starking Delicious is easily separated from others, the Granny Smith and Golden Delicious cultivars can be confused due to colour similarity.

We also defined rules for sorting the apples according to their size.

$$\begin{aligned}&\text{If the size of apple } B_{real} \leqslant 6.5\,\text{cm then the apple is small.}\\&\text{If the size of apple } 6.5\,\text{cm} < B_{real} \leqslant 8\,\text{cm then the apple is medium.}\\&\text{If the size of apple } 8\,\text{cm} < B_{real} \text{ then the apple is big.}\end{aligned} \tag{16}$$

In order to detect defective apples we defined a rule as:

$$\text{If } n > 1 \text{ and } \alpha_i > 2.5\,\text{cm}^2 \text{ and } \beta_i > 0.6 \text{ then the apple is defective} \tag{17}$$

where n denotes the number of possible defect regions and α_i, β_i the area and eccentricity of evaluated ith region, respectively. The eccentricity, which is known as elongation, is a second-order moment information about the current region. It denotes the circularity of the region (Liu et al., 2015).

The apple weight is estimated from its pixel area using the LSE method. There are different types of formula that demonstrate the relation between area (a) and weight (w). The best formula is:

$$w = f(a) = 0.0141a - 1.93 \times 10^{-7}a^2 + 2.083 \times 10^{-11}a^3 \tag{18}$$

Eq. (18) shows that there is a linear relation between the weight and area. The weight estimation results for the experimental apple data are depicted in Fig. 12 and are given in Table 5.

RMSE (Root Mean Square Error) is used for evaluating the test result. RMSE is a frequently used measure of the differences between actual and estimated values (Hyndman and Koehler, 2006). If the RMSE value is much closer to zero, the estimation method success is good, otherwise, its success is poor. The RMSE is calculated and formulated with the root of sum of mean square error.

$$RMSE = \sqrt{\frac{1}{M}\sum_{i}^{M}(actual_i - estimated_i)^2} \tag{19}$$

Here M shows the number of samples, $actual_i$ and $estimated_i$ shown with actual and estimated value of ith sample, respectively.

According to Table 5, while the weight estimation error for Golden cultivar is very small, the weight estimation error for Granny Smith cultivar is more. As a result, the weight could be estimated instead of measuring. In this way, the installation cost

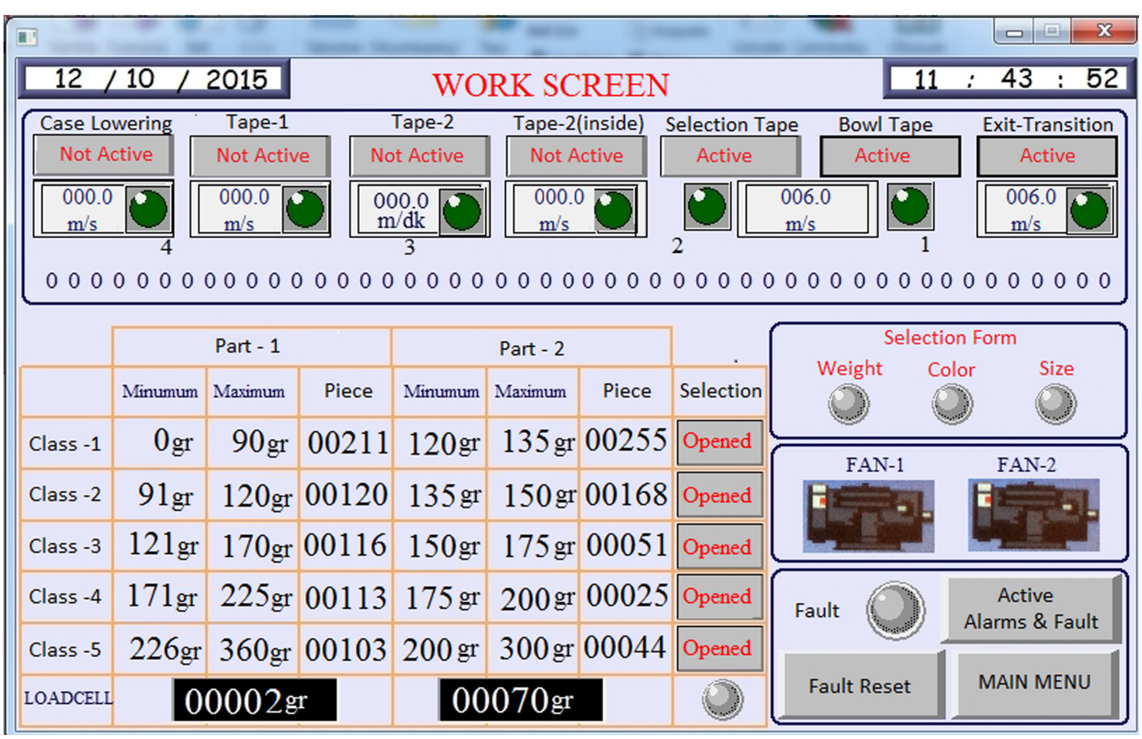

Fig. 8. PLC control screen.

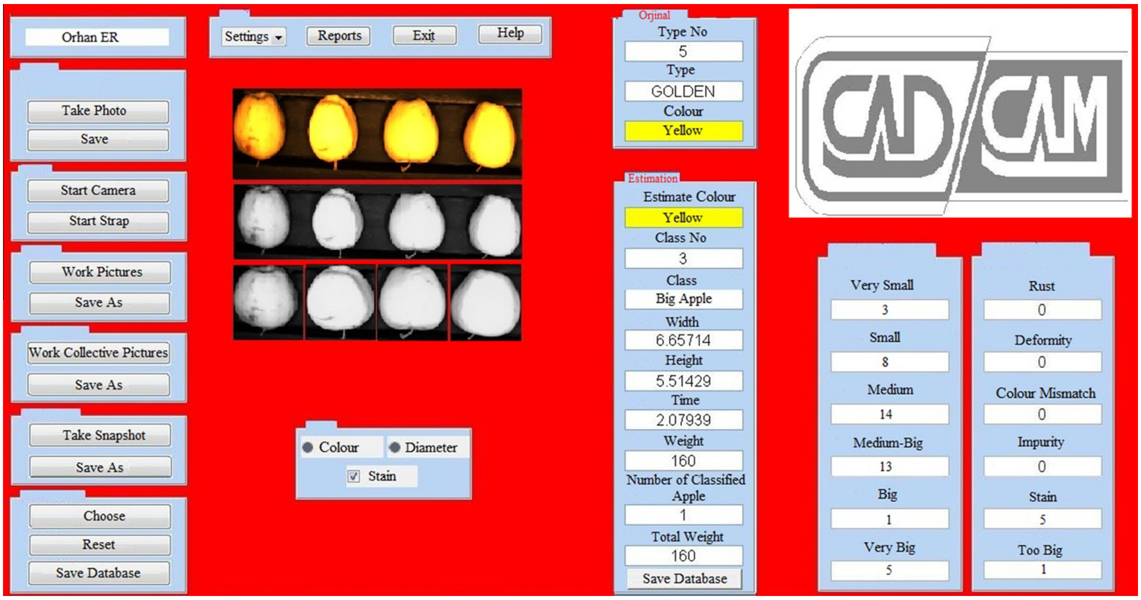

Fig. 9. Apple sorting GUI.

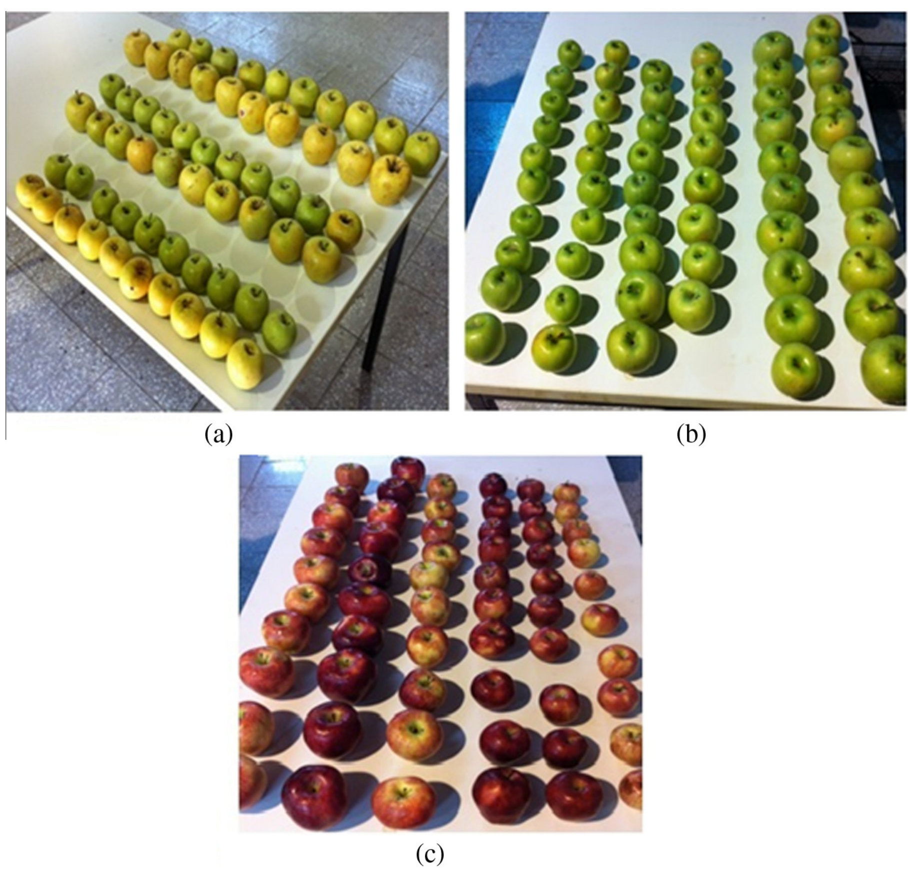

Fig. 10. The used apple cultivars in the experimental studies: (a) golden delicious, (b) granny smith and (c) starking delicious.

of a sorting machine will be decreased and the sorting speed will be increased.

The experiments are repeated three times for standard deviation and variation analysis using three different velocities: $v_1 = 0.05$ m/s, $v_2 = 0.1$ m/s, $v_3 = 0.2$ m/s to observe the performance change. The tests were completed in four days since the apples deformed quickly.

There are five class slides at the end of the machine. The light- and dark-coloured apples are sorted into the first two slides. The small, medium and big class apples are also sorted into the last

Table 2
Information of experimental apples.

Cultivar of apples	Total (number)	Colour		Size			Condition	
		Light	Dark	Small	Middle	Big	Abnormal	Normal
Starking (red)	61	31	30	20	21	20	10	51
Golden (yellow)	61	30	31	20	21	20	10	51
Granny (green)	61	30	31	20	21	20	10	51

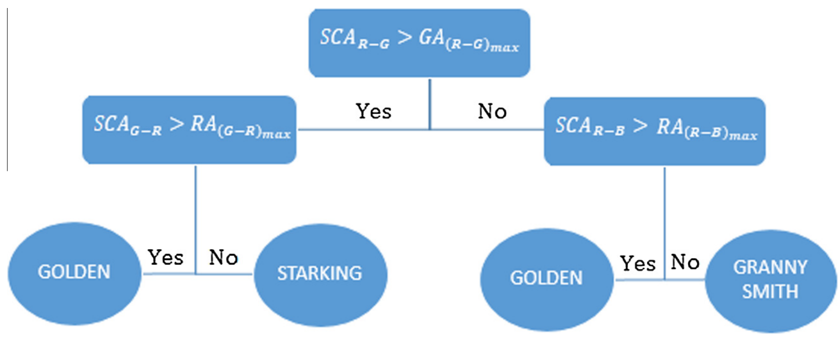

Fig. 11. C4.5 decision tree of apples according to the colour features.

Table 3
Threshold values of experimental apple data.

Apple cultivar	R–G		G–R		R–B	
	Min	Max	Min	Max	Min	Max
Starking	1417	2276	0	99	0	8
Granny	0	6	370	1488	0	29
Golden	0	328	60	195	876	1437

Table 4
Sorting results of experimental apple data with only their colour feature.

Apple cultivar	Sorting rates (%)
Starking	100.00
Granny	93.44
Golden	95.08

Table 5
Weight estimation results of three cultivar apples.

Apple cultivar	RMSE
Starking	6.43
Golden	0.80
Granny Smith	9.09
All data	5.44

Table 6
Average sorting results with four features for different speeds.

Speed (m/s)	Sorting rate (%)	Amount of apples (apples/hour)
0.05	89.00	18,200
0.1	82.41	36,250
0.2	79.00	54,000

three slides. If any apple has defective regions, this apple is discharged at the end of the band without classification. The defect, size, colour and weight classification can be done simultaneously during the selection.

As a result of the experimental studies, the classification results are obtained for three different velocities, which are shown in Table 6. The sorting rate denotes the average values of all apple cultivars and classification types such as colour, size, weight and defects.

As shown by Table 6, while the machine velocity increases, the sorting rate decreases. The proposed software can classify the apple with its four face images in at least 0.52 s. The maximum band speed of sorting machine is 0.5 m/s. When the band speed is increased, the efficiency is reduced. Therefore, the highest band speed is set to 0.2 m/s value.

The classification curve graphs of three different cultivar apples depending on the velocity are given in Fig. 13.

4. Results and discussions

Since the geometrical regularity and colour shades are homogeneous for the Golden cultivar apples, classifying this cultivar of

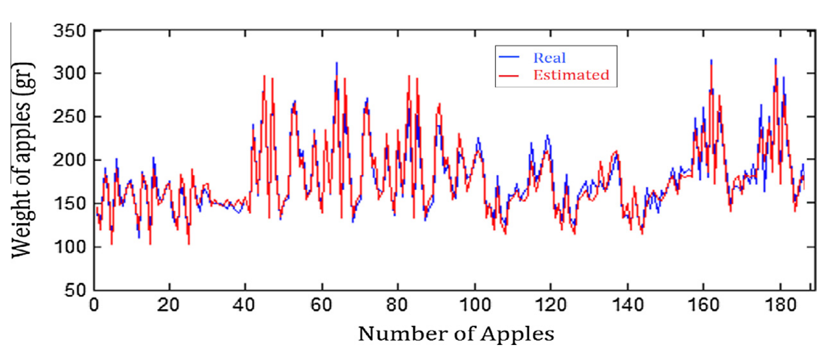

Fig. 12. Weight estimation of experimental apple data using Eq. (18).

apples is the easiest. It is seen from previous studies in literature that they are generally carried out using the Golden Delicious or a relative apple cultivars. Kavdir and Guyer (2008) sorted the Golden Delicious apples with their nine features such as colour, shape defect, circumference, firmness, weight, blush percentage, russeting, size of bruises, and size of natural defects. They used K-nearest neighbor and multilayer perceptron classifiers and obtained 90% classification accuracy rates. This result is similar with our results.

Having the close values for the results in each velocity leads us to conclude that the size classification of this apple with proper geometry is performed easily (Fig. 13). The case when the classification performances of the light and dark colour apples are close to each other means that the light and dark apples can be distinguished easily. It can be seen from Fig. 13 that the efficiency is quite high in the determination of stains on the apple. High stain classification efficiency is achieved by clearly seeing the stains on the apple since the colour shade of this apple is homogeneous and in light colour shades.

Starking cultivar apples can be considered as the most difficult group to be classified by using the image processing techniques. The geometry is bad and the colour shade carries quite variable colour density across the surface. For these reasons, Starking cultivar apples are not studied as much. The size classification for the Starking cultivar apples has the smoothest variation. As a result of finding the biggest diameters by using the software size algorithm, size classification and efficiency variation are smooth. The colour density for the Starking cultivar apples causes the surfaces exposed to the sun to form a dark surface colour. The resulting colour difference increases the colour-shade value interval. Although the average of the apple image from 360 degrees is taken, spatial value change causes the classification efficiency to be destabilized. Since the stain and deformation identification becomes harder for the dark colour shaded surfaces, the stain-sorting rate varies between 65% and 80% as seen in Fig. 13(c).

Granny Smith cultivar apples are more rigid and durable compared to the other cultivars. Because of this feature, the natural stain formation and the stain-colour shade differences occur less on its surfaces. The stain-classification performance fluctuations in the graph shown in Fig. 13(b) occurred for that reason.

Leemans et al. (2002) reported that the global classification ratio is about 70–78%. The efficiency value of the selection machine tested is found to be close to that ratio. As shown by the overall system efficiency values, the efficiency varies depending on the velocity. Band velocity affects the accurate classification capability of the software at a very low rate. The other findings decreasing the efficiency are caused by the natural geometric distortion of the apples and excessive colour shade differences on an apple.

The main problem with our design is that there is not a single conveyor band. When the apple is recognized at the end of the roller conveyor, the brush system pushes the apple to the transporter conveyor. Sometimes, the location of apple could have changed and this apple is sent to the wrong slider conveyor. In the next design, a single conveyor band should be used for machine vision and transport.

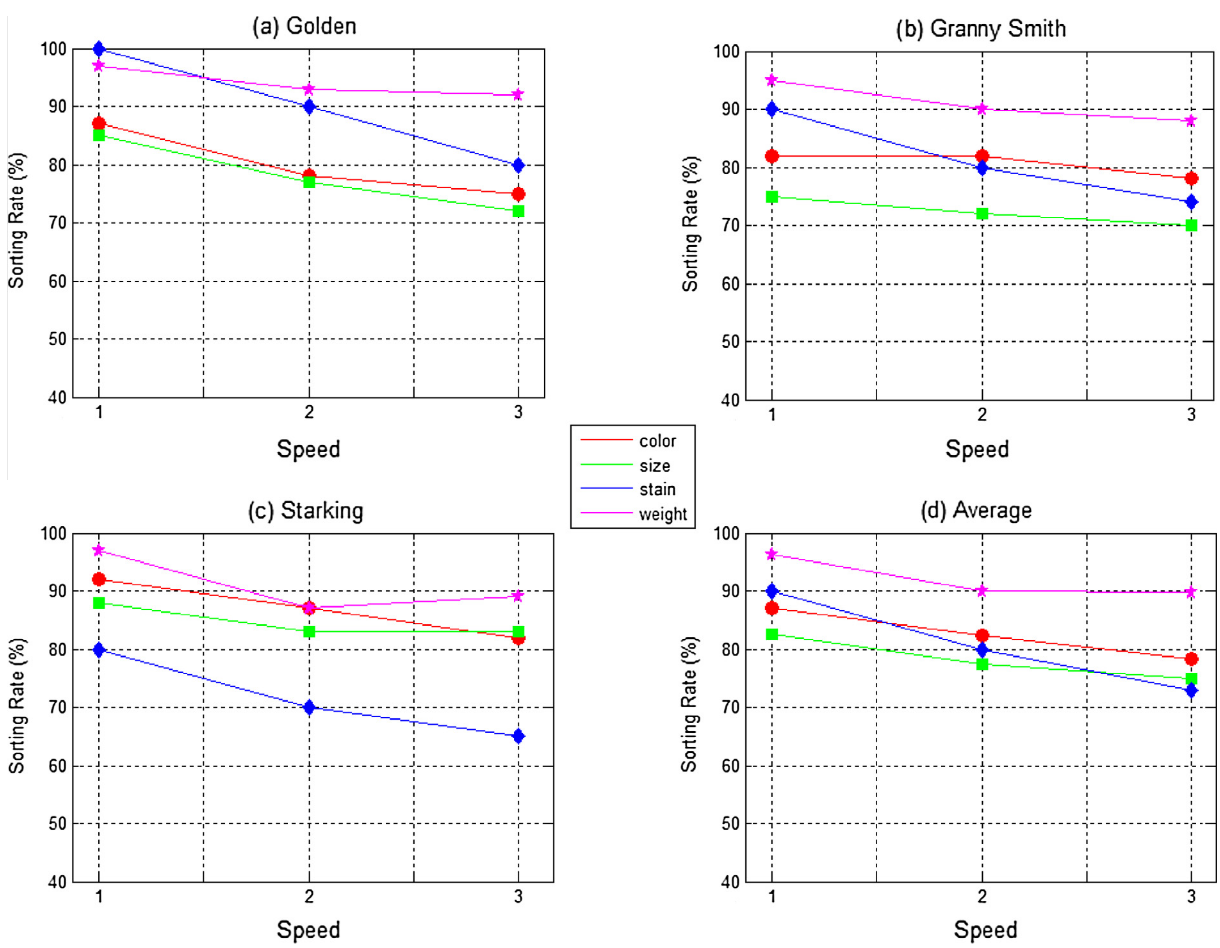

Fig. 13. Sorting rates of three apple cultivars and their average values.

When the proposed machine is compared with other studies according the capacities, there are a little studies. Golpira and Golpîra (2012) proposed an apple sorting machine and their grading capacity is 130 kg/h. If we assume that an apple has 200 g averagely, the capacity of our machine is 10,800 kg/h. This result is better than Golpira and Golpîra. But, this result is half of Greefa and Aweta apple sorting machines.

5. Conclusions

In this study, an automatic apple sorting machine was designed and realized. This machine consists of mechanical parts, such as a roller, transporter, apple class sliders and brush; electrical and electronic parts, such as motors, sensors, lamps, computer and PLC, and software, such as PLC, image acquisition and processing, sorting, estimation and curve fitting programs. The aim of this study was to design a hybrid machine that meets the market needs: a machine to sort at least 200.000 apples/day with a better than 75% sorting accuracy rate. The proposed machine achieved these aims with 432.000 apples/day and 79% sorting accuracy scores, when a day is accepted as 8 working hours.

The roller conveyor was used to take at least four images of each apple. The enclosed cabin was designed to house the camera and lighting system. In this way, homogeneous illumination was provided for successful apple sorting. The machine vision system obtains the visual features of the apples. The brush system moved the apples from roller conveyor to transporter conveyor. The load cells measured the weight of apples. The bowl and its triggering systems were used for successful sorting. The PLC controlled all the actuators, conveyors and bowls.

The captured and processed images displayed the colour, size, area, and defective regions of the apples. The C4.5 classifier, which is a rule-based method, was used to sort apples according to their features. In experimental studies, although Golden Delicious, Granny Smith and Starking Delicious were evaluated, other cultivars of apple can also be evaluated in this machine.

We also studied estimation of apple weight using a second order polynomial using only the area of apple. We estimated the apple weights to within 5–6% gram error rates. This result is sufficient, but the ratio between the weight and area changes according to the apple cultivar, apple life and apple-keeping conditions. For those reasons, using the load cell instead of estimation is more suitable to determine the apple weight.

Experimental studies showed that while the defective region detection is very easy for Golden Delicious and Granny Smith apple cultivars, it is very hard for Starking Delicious, due to dark and light colour surfaces. The weight is a very important feature for all apple cultivars. For that reason, the traditional sorting systems usually sort apples using only the weight feature.

The size determination is very hard. While the smaller apples can easily whirl round, the bigger apples cannot whirl round because of the roller structure. As a result, to catch four perfect images that define all sides of the apple is very difficult.

The proposed machine can be used for different fruits and vegetables, such as oranges, potatoes and so on.

Acknowledgements

This research is funded by the Ministry of Science, Industry and Theology of Turkey. Project no 00076.STZ2007-1 (SANTEZ) and we would like to acknowledge technical support from Süleyman Demirel University CAD/CAM Research and Application Center.

References

Anonymous, 1989. Laying down quality standards for carrots, citrus fruit and desert apples and pears and amending commission regulation. Off. J. L 097 (58), 0019–0039.
Arlimatti, S.R., 2012. Window based method for automatic classification of apple fruit. Int. J. Eng. Res. Appl. 2 (4), 1010–1013.
Baranowski, P., Mazurek, W., Pastuszka-Woźniak, J., 2013. Supervised classification of bruised apples with respect to the time after bruising on the basis of hyperspectral imaging data. Postharvest Biol. Technol. 86, 249–258.
Baxes, G.A., 1994. Digital Image Processing, Principles and Applications. John Wiley & Sons Inc., USA, p. 445.
Bhatt, A.K., Pant, D., 2015. Automatic apple grading model development based on back propagation neural network and machine vision, and its performance evaluation. AI & Soc. 30 (1), 45–56.
Cetişli, B., Kalkan, H., 2011. Polynomial curve fitting with varying real powers. J. Electron. Electr. Eng. 6 (112), 117–122.
Chen, Yud-Ren, Chao, Kuanglin, Kim, Moon S, 2002. Machine vision technology for agricultural applications. Comput. Electron. Agric. 36, 173–191.
Du, C.-J., Sun, D.-W., 2004. Recent developments in the applications of image processing techniques for food quality evolution. Trends Food Sci. Technol. 15, 230–249.
Fattal, M., Turk, A.E., Dahrouj, M., Badereddin, A., 2008. Automatic apple sorting and grading using machine vision. In: 7th FEA Student Conference 21–22 May.
Golpira, H., Golpîra, H., 2012. Improvement of an apple sorting machine using PSNR criterion. In: IEEE Int. Conf. on Advanced Mechatronic Systems (IEEE ICAMechS), Japan, Tokyo, pp. 729–732.
Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22 (4), 679–688.
Jang, J.S.R., Sun, C.T., Mizutani, E., 1997. Neuro-Fuzzy and Soft Computing. Prentice Hall, Upper Saddle River.
Kavdir, I., Guyer, D.E., 2008. Evaluation of different pattern recognition techniques for apple sorting. Biosyst. Eng. 99 (2), 211–219.
Leemans, V., Magein, H., Destain, M.-F., 1997. The quality of Golden Delicious apples by color computer vision. Preprints from the Mathematical and Control Applications in Agriculture and Horticulture, Int. Fed. of Automatic Control (IFAC) Workshop, 28 September–2 October 1997, Hannover, Germany, pp. 175–179.
Leemans, V., Magein, H., Destain, M.-F., 2002. On-line fruit grading according to their External quality using machine vision. Automat. Emerg. Technol., Biosyst. Eng. 83 (4), 397–404.
Li, Q., Maohua, W., Weikang, G., 2002. Computer vision based system for apple surface defect detection. Comput. Electron. Agric. 36, 215–223.
Liu, D., Ning, X., Li, Z., Yang, D., Li, H., Gao, L., 2015. Discriminating and elimination of damaged soybean seeds based on image characteristics. J. Stored Prod. Res. 60, 67–74.
Mizushima, A., Lu, R., 2013. An image segmentation method for apple sorting and grading using support vector machine and Otsu's method. Comput. Electron. Agric. 94, 29–37.
Nicolai, B.M., Lötze, E., Peirs, A., Scheerlinck, N., Theron, K.I., 2006. Non-destructive measurement of bitter pit in apple fruit using NIR hyperspectral imaging. Postharvest Biol. Technol. 40, 1–6.
Özongun, Ş., Dolunay, E.M., Öztürk, G., Pektaş, M., 2014. Performance of some apple varieties in Eğirdir (Isparta) conditions. Fruit Sci. 1 (2), 21–29 (in Turkish with English abstract).
Penman, D.W., 2001. Determination of stem and calyx location on apples using automatic visual inspection. Comput. Electron. Agric. 33, 7–18.
Pla, F., José Sanchiz, M., Sánchez, J.S., Ugolini, N., Diaz, M., 2001. A machine vision system for on-line fruit colour classification. In: Pattern Recognition and Image Analysis: Proc. of the IX Spanish Symp. on Pattern Rec. and Image Anal., Castellón, Spain, 16–18 May 2001, pp. 323–330.
Quinlan, J.R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann Publisher Inc., San Francisco, USA, p. 299.
Sabliov, C.M., Boldor, D., Keener, K.M., Farkas, B.E., 2002. Image processing method to determine surface area and volume of axi-symmetric agricultural products. Int. J. Food Prop. 5 (3), 641–653.
Setsirichok, D. et al., 2012. Classification of complete blood count and hemoglobin typing data by a C4.5 decision tree, a naïve Bayes classifier and a multilayer perceptron for thalassemia screening. Biomed. Signal Process. Control 7 (2), 202–212.
Unay, D., Gosselin, B., 2002. Apple defect detection and quality classification with MLP-neural networks. Processing of the Risk Workshop on Circuits. Available from: <http://www.tcts.fpms.ac.be/publications/papers/2002/prorisc02_dubg.pdf>.
Wen, Z., Tao, Y., 1999. Building a rule-based machine-vision system for defect inspection on apple sorting and packing lines. Expert Syst. Appl. 16 (3), 307–313.
Yam, K.L., Papadakis, S.E., 2004. A simple digital imaging method for measuring and analyzing color of food surfaces. J. Food Eng. 61, 137–142.
Zhang, B., Huang, W., Gong, L., et al., 2015. Computer vision detection of defective apples using automatic lightness correction and weighted RVM classifier. J. Food Eng. 146, 143–151.