A Study on Optimum Feature Selection for Recognition of Starking Apple

Hayrettin Toylan
Pinarhisar Vocational High School Department of Electricity, Kirklareli University, Kirklareli, Turkey
Email: hayrettintaylan@hotmail.com

Hilmi Kuscu
Faculty of Architecture and Engineering, Department of Mechanical Engineering, Trakya University, Edirne, Turkey
Email: hilmi@trakya.edu.tr

***Abstract*—Pattern recognition problems require selection of feature subset in the image process. In order to increase the performance and to reduce the fails of classifier, it is needed to construct a sensitive feature selection. In this study, firstly apple images were taken from different angles of each apple via developed computer vision system. These images were then segmented and 25 features of the pattern were extracted by colors, texture and statistical. Feature selection methods which are SFS (Sequential Forward Selection), SBS (Sequential Backward Selection), SFFS (Sequential Floating Forward Selection), SBFS (Sequential Backward Floating Selection) were compared to specify the feature subset which has the maximum discriminatory. To perform this, a database used which has 726 patterns of 100 apples. According to experimental results, feature subset with 11 elements was created with SBFS method which is more successful than the other methods with % 83,1 accuracy rate. This method provides more successful labeling of patterns (defect, apple calyx, apple stem and healthy tissue).**

***Index Terms*—computer vision, feature selection, pattern recognition, apple**

I. Introduction

Most recently, determining or classifying the feature of an apple by using computer vision and image processing has attracted much attention [1]-[4].

One of the biggest problems of the apple classification is separating the apple calyx and apple stem from the defected patterns of apple. Classifying these patterns is important in two aspects. Firstly, radius of the apple is defined as the largest equatorial radius of the axis which is perpendicularly combines apple calyx and apple stem. Secondly, by separating apple's surface defects from calyx and stem, it helps to stop increasing the rotten parts of apple surface. Process starts with segmentation of apple surface and the feature of the segmentation is obtained. Groups of patterns are defined according to the feature.

The colour is an important qualification when evaluating the surface of apple. Most of the control systems use only colour for the segmentation [5]. Researchers had used different methods to define patterns which were obtained after the image segmentation. Leamans and Destain characterised blobs with 16 features as colour, texture, position and shape [1]. Kleynen and others identified apple stem and calyx by using similar model matching algorithm [6]. In their research, Unay and Gasselin used thresholding method for segmentation. They defined 11 features of pattern as statistical, textural, and shape. The most successful features were chosen by SFFS feature selection method. They defined patterns according to features which they specified [7]. Different applications of the image processing were used for feature selection method. Ekenel and Sankur used SFFS method for face recognition [8].

II. Materials and Methods

A. Image Acquisition

Image acquisition device was used for this research is a 1.3 mega pixel (H x V = 1280 x 1024 pixel) resolution 25 fps (frame per second) Complementary Metal-oxide Oxide Semiconductor (CMOS) camera with a manually adjustable 6mm focus length which was mounted on camera, C-mount lens. A lighting system, 600 mm x 670 mm x 300 mm white painted diffusely illuminated tunnel with four fluorescent lamps, a conveyor belt in which fruits are placed (Fig. 1).

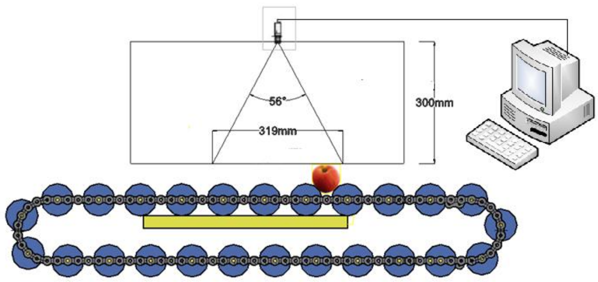


Figure 1. Computer vision system

Manuscript received June 30, 2013; revised August 22, 2013.

doi: 10.12720/ijoee.1.4.264-267

B. *Image Segmentation*

Segmentation is one of the hardest and most important applications of the image processing. Segmentation means classification of the areas which have different features. On the other hand it means grouping the areas which have similar characteristics. According to Due and Sun, image segmentation technique is divided into four sections for food's quality evaluation. These are thresholding based, region based, gradient based and classification based techniques (Fig. 2). While thresholding based and region based techniques are commonly used applications it is thought that the other two ways are used rarely [9].

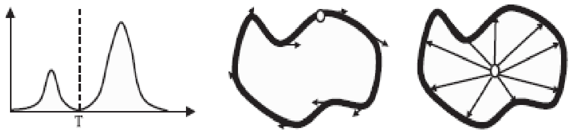
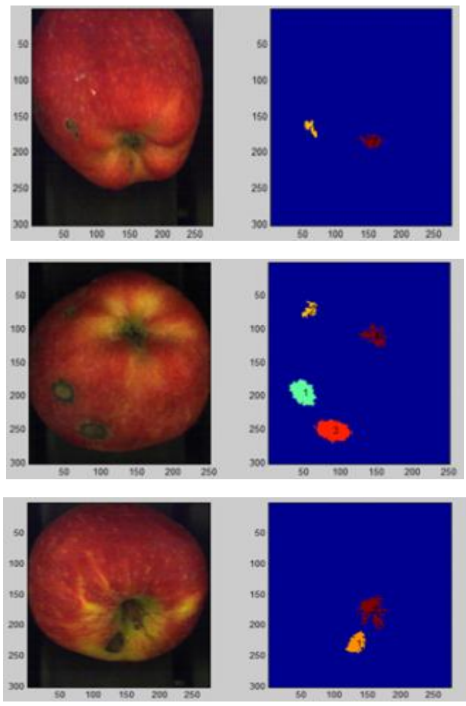


Figure 2. Typical segmentation techniques. (a) Thresholding (b) Edge-based segmentation (c) Region-based segmentation [10]

In this study, segmentation process was performed by using thresholding method without the need for camera filter, by the help of "CIE L*C*h" as a device independent color space.

$$f(x,y)=\begin{cases}1 & If \quad C>T\\ 0 & ELSE\end{cases} \tag{1}$$

After examining the different color spaces, it was decided to carry on thresholding process on "C" component. As a result of thresholding process, small parts were deleted and the resulting patterns were numbered (Fig. 3).



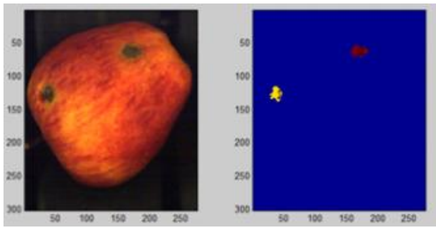


Figure 3. Examples of segmentation on apple images

C. *Feature Extraction*

Patterns obtained as a result of the segmentation process must be specified according to the basic characteristics in which they are separable each other. Hence, in this study, twenty five features were specified for each of the patterns containing statistical, textural and shape information (Table I).

TABLE I. NUMBERS OF FEATURES

1	R_μ	6	B_σ	11	GG_σ	16	*Min*	21	*S*
2	G_μ	7	RG_μ	12	BG_σ	17	*Max*	22	*P*
3	B_μ	8	GG_μ	13	*Hu*	18	*Max-Min*	23	*Y*
4	R_σ	9	BG_μ	14	*Mean(μ)*	19	*Skewness*	24	*R*
5	*G*	10	RG_σ	15	*Std.dv.(σ)*	20	*Kurtosis*	25	*N*

Mean of *R*, *G*, *B* color space components;

$$Mean(\mu)=\frac{1}{N}\sum_{j=1}^{N}Pj \tag{2}$$

Standard Deviation of *R*, *G*, *B* color space components

$$Standard\ Deviation(\sigma)=\sqrt{\frac{1}{N-1}\sum_{j=1}^{N}(Pj-\mu)^2} \tag{3}$$

Mean and standard deviation of *R*, *G*, *B* color space components were recomputed after gradient.

$$\nabla F=\frac{\partial F}{\partial x}i+\frac{\partial F}{\partial y}j \tag{4}$$

One of the features that uses commonly in pattern recognition is invariant moments. Seven invariant moments have been developed firstly by Hu [11]. In this study the first moment was used developing by Hu. These moments are fixed in changes of pattern scale or as a result of rolling process. Accordingly, central moment of *f(x, y)* pattern is;

$$\mu_{pq}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\bar{x})^p(y-\bar{y})^q f(x,y)\partial(x-\bar{x})\partial(y-\bar{y}) \tag{5}$$

$$\bar{x}=m_{10}/m_{00},\quad \bar{y}=m_{01}/m_{00} \tag{6}$$

$\bar{x}$, $\bar{y}$ are coordinates of the center of pattern gravity. Central moment and normalized Hu invariant moment [11];

$$\phi_1=\mu_{20}+\mu_{02} \tag{7}$$

Mean, standard deviation, minimum, maximum and difference between minimum and maximum of all pixels of pattern are as follows:

$$minimum(\min) = \min(1, \ldots ..N) \tag{8}$$

$$maximum(\max) = \max(1, \ldots ..N) \tag{9}$$

$$max\min - gradient = \max - \min \tag{10}$$

$$Skewness = \frac{\sum_{J=1}^{N}(Pj - ?)^3}{N\sigma^3} \tag{11}$$

$$Kurtosis = \frac{\sum_{J=1}^{N}(Pj - ?)^4}{N\sigma^4} \tag{12}$$

Area; S

Circumference; P

Roundness;

$$Y = \frac{P^2}{4\pi S} \tag{13}$$

Maximum radius; R

The actual area; N, It is find by multiplying a pixel area and a pixel number covered pattern.

D. Feature Selection

In order to increase the performance and also to decrease the fault rates of the classifier, all of the o features which are obtained are not used in the classifier. Instead of this the most effective selective features are determined and classification is preceded according to the based features. Feature selection is made to determine the most suitable feature subset. Process restarts when new features come or education datas are changed.

If all feature groups contains N features, number of possible subsets are equal to 2^N. Different type of research strategies were used in order to find most suitable feature subsets. Some of them are Sequential Forward Selection (SFS), Sequential Backward Selection (SBS) and Sequential Forward Floating Selection (SFFS). In these methods, erasing or adding features are the only way of change. As implies by the name, SFS searches the features space sequentially. It starts from an empty group and in every new search it looks for the best single feature space in order to add group. Conversely, SBS starts from a full feature group and it erases the worst single feature for every single search. SFFS algorithm is the method of search from below to above and it expands the SFS algorithm. According to SFFS procedure as long as subset results are better than the previous results it moves one step reverse for each step forward. As a result if there is no performance there is no step reverse. SBFS has the same principle like SFFS. Hereby reverse watching is controlled dynamically and eventually parameter set is not needed [12] and [13]

III. Results

A database was created while given the labels of patterns (calyx, stem, and defect) on the apple image which was obtained from segmentation.

TABLE II. Patterns Observed for Segmentation Process

Group	Number of pattern
Calyx	113
Stem	86
Defect	387
Healthy tissue	140
Total	726

As seen on Table II, in database patters divided into three groups (stem, calyx, and defect) as a result of segmentation on starking apple images. Besides, there is a pattern which is called blank "healthy tissue" arise from dark pieces of image. Blank label is an undesired situation of segmentation.

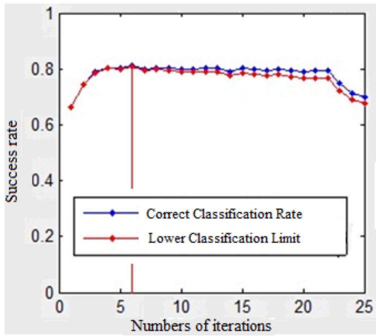


Figure 4. SFS feature selection success curve

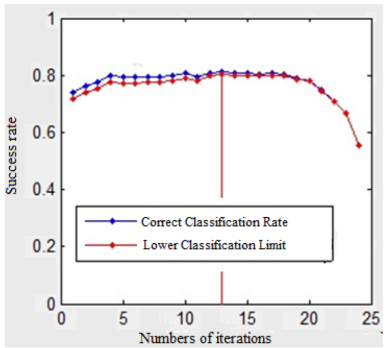


Figure 5. SBS feature selection success curve

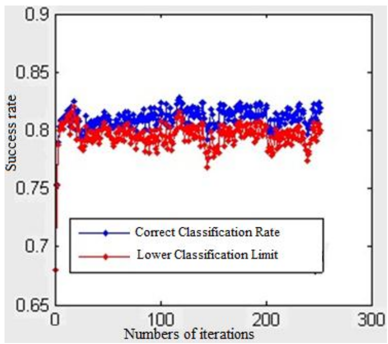


Figure 6. SFFS feature selection success curve

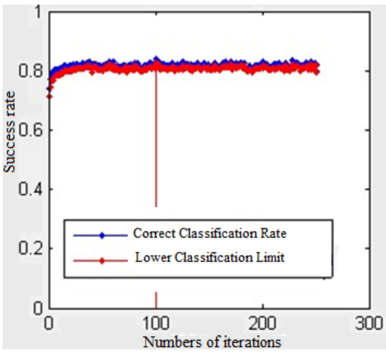


Figure 7. SBFS feature selection success curve

Created database was used for testing the feature selecting algorithm. To determine the features of a subset, SFS, SBS, SFFS, SBFS methods compared with the features selection programme which was developed by Ververidis and Kotropoulos [14]. Success curves of SFS, SBS, SFFS, and SBFS methods obtained by this programme can be seen respectively in Fig. 4, Fig. 5, Fig. 6 and Fig. 7.

TABLE III. DISCRIMINATORY SUCCESS RATES AND SELECTED FEATURES SUBSETS OF FEATURE SELECTION METHODS

Feature Selection Method	Success Rate (%)	Selected Features
SFS	81,1	[8, 10, 15, 22, 23, 24]
SBS	81,3	[1, 2, 5, 6, 10, 11, 14, 16, 19, 22, 23, 24]
SFFS	82,9	[1, 3, 5, 6, 10, 11, 14, 15, 16, 18, 23, 24]
SBFS	83,1	[1, 2, 4, 5, 6, 10, 14, 16, 18, 23, 24]

The highest success rates and selected features by feature selection methods are presented in Table III. As a result of this study SBFS method selected the best feature selection method with the %83,1 success. The feature subset which has 11 elements [1], [2], [4]-[6], [10], [14], [16], [18], [23] and [24] was obtained from this method.

IV. CONCLUSION

In this study, firstly segmentation performed on images which obtained from computer vision system. After the segmentation, 25 features obtained which contain statistical, textural and shape information. In order to increase the performance and also to decrease the fault rates of the classifier, all of the features which are obtained are not used in the classifier. Because of that reason, SFS, SBS, SFFS and SBFS methods compared in order to find the features subset which can do the best selection. To perform this, a database used which has 726 patterns of 100 apples. As a conclusion, with SFBS method, feature subset which has 11 elements has %83,1 success rate. It was observed that SBFS is more successful when compared with other methods.

ACKNOWLEDGMENT

This project is supported in the framework of Trakya University Scientific Research Projects under project no 2010/140

REFERENCES

[1] V. Leemans and M. F. Destain, “A real-time grading method of apples based on features extracted from defects,” *J. Food. Eng*, vol. 61, no. 1, pp. 83–89, Jan. 2004.
[2] Q. Li, M. Wang, and W. Gu, “Computer vision based system for apple surface defect detection,” *Comput. Electron. Agric*, vol. 36, no. 2-3, pp. 215- 223, Nov. 2002.
[3] I. Kavdır and D. E. Guyer, “Evaluation of different pattern recognition techniques for apple sorting,” *Biosystems Eng.*, vol. 99, no. 2, pp. 211 – 219, Feb. 2008.
[4] Z. Wen and Y. Tao, “Building a rule-based machine-vision system for defect inspection on apple sorting and packing lines,” *Exp. Sys. Appl*, vol. 16, no. 3, pp. 307–313, Apr. 1999.
[5] J. Blasco, N. Aleixos, J. Go ḿez, and E. Molto´, “Citrus sorting by identification of the most common defect using multispectral computer vision,” *J. Food Eng*, vol. 83, no. 3, pp. 384–393, Dec. 2007.
[6] O. Kleynen, V. Leemans, and M.-F. Destain, “Development of a multi-spectral vision system for the detection of defects on apples,” *Journal of Food Engineering*, vol. 69, no. 1, pp. 41–49, July 2005.
[7] 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, Jan. 2007
[8] H. K. Ekenel and B. Sankur, “Feature selection in the independent component subspace for face recognition,” *Pattern Recognition Letters*, vol. 25, no. 12, pp. 1377-1388, Sept. 2004
[9] C. J. Du and D. W. Sun, “Recent developments in the applications of image processing techniques for food quality evaluation,” *Tren. Food Sci. Tech.*, vol. 15, no. 5, pp. 230–249, May 2004.
[10] D.-W. Sun, “Inspecting pizza topping percentage and distribution by a computer vision method,” *Journal of Food Engineering*, vol. 44, no. 4, pp. 245–249, June 2000
[11] M. K. Hu, “Visual pattern recognition by moment invariants,” *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, Feb. 1962
[12] P. Pudil, J. Novovicova, and J. Kittler, “Floating search methods in feature selection,” *Pattern Recognition Letters*, vol. 15, no. 11, pp. 1119-1125, Nov. 1994
[13] F. Tan, “Improving feature selection techniques for machine learning,” Ph.D. dissertation, Georgia State University, USA, 2007.
[14] (2010). Feature Selection Using Matlab. Available: http://www.mathworks.com/matlabcentral/fileexchange/22970-feature-selection-using-matlab

Hayrettin TOYLAN was born in Kirklareli/Turkey on 3rd June, 1982. He studied the primary school in Kirklareli and then the high school in Kirklareli, at Technical and Industrial Vocational High School (Lycee degree). He joined to the Electricity Department of the Faculty of Technical Education of Marmara University in 2000. He graduated in 2004. In 2005, he started to his MSc at Electricity Department in Institute of Science and became a research assistant at Electricity Education Department in Faculty of Technical Education in Trakya University in 2007. He completed his MSc in March 2008. He majored in Mechanical Engineering as a PhD student in Institute of Science at Department of Mechanical Engineering, Faculty of Engineering and Architecture in Trakya University in 2008. After completing his PhD education in 2012, was appointed as an assistant professor in Kirklareli University. He still continues this task.

Hilmi Kuscu is an Assistant professor in Mechanical Engineering Department at the University of Trakya. He received his B.Sc. degree in Mechanical Engineering from the Istanbul Technical University in 1992. He obtained his M.Sc. (1996) and Ph.D. (2002) degrees from the University of Trakya. His current research interests include theory of machinery, system Dynamics and control, computer and PIC programming and robotics.