DIMENSION AND COLOR CLASSIFICATION OF OLIVE FRUIT WITH IMAGE PROCESSING TECHNIQUES

Fatma Betül KINACI INCE[1,+], Sakir TASDEMIR[2], İlker Ali OZKAN[2]

[1]Bayraklı Gazeteci Çetin Altan VTAH, İzmir Türkiye

[2]Selcuk University, Faculty of Technology, Konya Türkiye

fatmabetulkinaci@hotmail.com, stasdemir@selcuk.edu.tr, ilkerozkan@selcuk.edu.tr

Abstract

The development of image processing technology appears in agriculture as well as in many other fields. Various classifications are carried out for fruits and vegetables. These are processes such as determining the harvest time according to their degree of maturity, deciding the way of collection and performing packaging operations according to their dimension. This study aims to classify the fruit according to its intended use in order to benefit more from the olive fruit that is important in industrial terms. In this study, olive fruit is classified as big, medium, and small according to its dimensions. Also classified as black and green according to their colors. This classification process was made in MATLAB environment and the KNN algorithm and decision trees was used. The results are obtained with Euclid and Manhattan methods used with the KNN algorithm and are given comparatively. According to the application results, 100% success was achieved in both methods in color classification. In dimension classification, 89.2% classification success was achieved in KNN algorithm and 86.7% in decision tree method.

Keywords: Image processing, olive classification, KNN classification algorithm, decision tree.

[+] This paper has been presented at the ICAT'20 (9th International Conference on Advanced Technologies) held in Istanbul (Turkey), August 10-12, 2020.

1. Introduction

Image processing is the processing of pre-recorded or moving images by digitizing. These operations are performed using various software and hardware [1]. Image processing technology is developing rapidly and continues to be used in many areas. Image processing occurs in a wide range of areas such as medical applications, biometric authentication applications like fingerprinting and face recognition, license plate recognition systems, classification of agricultural products, mining exploration [2]. Fuzzy logic study for mass determination on mammography images [3], determination of potato classification parameters with the help of artificial neural networks [4] and application designed to assign direction to smart vehicles [5] are only a few of the hundreds of studies.

A preparation is required to process the digital image. The purpose is to make the image suitable for the field of use [6]. For this, various processes and filters are used on the pixels. These processes are applied by treating with a matrix prepared for application to each pixel in the image. The size of the matrices varies depending on the processing method and the rate of use. This makes the image better interpretable [7].

Mathematical models used to define colors are called color spaces. The numerical value corresponding to each pixel of the color image of the two-dimensional matrix is called the RGB (red, green, blue) space [8]. HSV (hue, saturation, value) color space refers to the color essence, saturation, and brightness of the image [9].

There are programs available for every user, such as Photoshop, Picasa and Paint.NET. In addition to this, there are programs where expert users can perform all image editing operations by their own algorithms. Visual programming languages such as C++, C#, Java, Python, and software programs such as MATLAB perform the expected operations with the help of libraries like OpenCv, EmguCV, OpenFrameworks, Halcon, CIMG, GNU Octave and PIL.

1.1 Olive Farming in Turkey

According to TÜİK death reason statistics, circulatory system diseases were the leading cause of death [10]. If there is a disadvantage of technology development and facilitating daily life, it is that people are living a more sedentary life. This kind of life combined with malnutrition invites to the most important diseases of the age when obesity

and cardiovascular diseases lead. It is known that doctors and dieticians recommend movement and Mediterranean type feeding to prevent these diseases. Because the Mediterranean diet, rich in vegetables and fruits, contains less nutrients than saturated fats and it also contains more than olive oils that do not elevate lipids [11]. In recent years, interest in healthy living and nourishment has increased and olive agriculture, which has become a tradition for thousands of years in Anatolia, has taken its share from this interest.

According to the data of the Republic of Turkey Ministry of Agriculture and Forestry, Turkey's share in world olive planting area is 9.3%. Turkey's rank was the fifth after then Spain, Tunisia, Italy, and Greece. In the grain olives production Turkey's rank is the fourth after Spain, Italy, and Greece. In the processed olive production Turkey's rank is in the third place after Spain and Egypt with 15.9% share in the world [12]. The most important olive oil importer in the world is the United States. In addition, it is known that the producer countries such as Italy and Spain also traded and sold olives. This is an important export revenue for them. Turkey's share of global exports is 4.9%, but it has no import share [12]. In order to increase this share, the olive plant should be given the necessary attention and the research should be replicated.

In this study, in addition to the color classification which enables to determine the maturity of olive fruit, it is aimed to realize the dimension classification. In this way, faster and more efficient classification is aimed.

2. Material and Method

2.1 Olive Fruit

According to the Turkish Food Codex Processed Olive Notification (Notification ID: 2014/33) [13], olive is classified into three groups: green, green to pink and black. The dimensions are determined according to the same regulation and given in Table 1.

In Turkey, the olive extraction process is carried out with sieve type machines. They work by separating light impurities like leaves and separating them according to their size. Kuncan et al. image processing-based olive sifting machine study was prepared in HALCON environment and olives were classified according to their colors [9]. In this

study, the olives were classified as light green, green and black and the olive separation process was moved to an upper level.

Table 1. Grade of olive (units/kg)

Units/kg	**Class**	**Units/kg**	**Class**
60-70	11XL	181-200	2XL
71-80	10XL	201-230	XL
81-90	9XL	231-260	L
91-100	8XL	261-290	M
101-110	7XL	291-320	S
111-120	6XL	321-350	XS
121-140	5XL	351-380	2XS
141-160	4XL	381-410	3XS
161-180	3XL		

2.2 Image Processing

Color spaces enable grouping of colors in color images. RGB color space consists of a mixture of red, green, and blue colors in various proportions. As seen in Figure 1, HSV color space uses the color's essence, saturation, and brightness values. To distinguish an object from within the image, threshold value can be applied using the hue value and the object can be easily separated from the others. Therefore, it is more convenient to use HSV color space in image processing applications [14].

Classification can be performed on images using many methods such as neural networks, deep learning and classification algorithms used in data mining. In this study, hue, saturation, and value attributes for color classification; area, circumference, diameter and center attributes are used for dimension classification. The results of KNN and decision tree classification algorithms are obtained and compared.

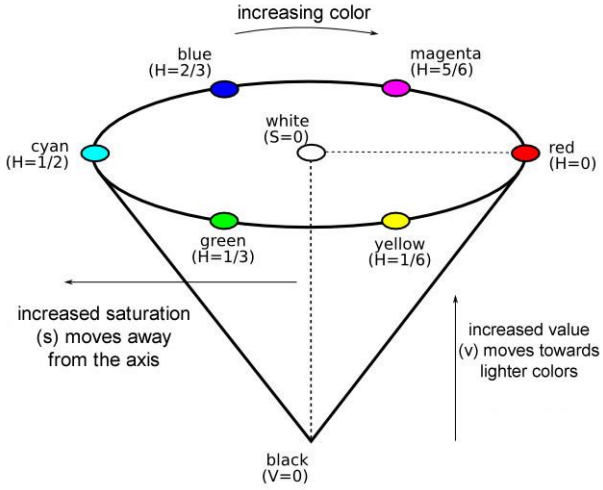

Figure 1. Hue, saturation, and value color model [14]

The KNN algorithm is widely used in various fields because it is easy to perform and is useful in the learning process [15]. In this classification system, the distance of a number of previously classified objects is considered [16]. This classification includes training data. When a new data is added, it is determined which data is included in the data by looking at the data that is closest to it. For this reason, it is very important that the training data is as large as possible in the KNN classification algorithm and that the k value is selected in such a way that it provides the optimum classification. When the k value is increased too much, it is observed that the success of classification decreases. The widely used distance measures with the KNN classification are Euclidean (Eq. 1), the distances of Manhattan (Eq. 2) and Minkowski (Eq. 3). Equations of these distances for two points [15];

Euclidean : $$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \quad (1)$$

Manhattan : $$\sum_{i=1}^{k}|x_i - y_i| \quad (2)$$

Minkowski : $$(\sum_{i=1}^{n}|x_i - y_i|^p)^{1/p} \quad (3)$$

The *x* and *y* values used in the equations show the 2-dimensional pixels, and the *k* value indicates how many closest neighbors will be taken as basis. The value of p in the Minkowski distance formula indicates Manhattan when it is 1, Euclid when it is 2, and Chebyshev when it is ∞.

Decision trees, which are frequently used for classification and estimation, are preferred because of their easily understandable and interpretable results [17]. Decision trees consist of learning and classification steps. Various algorithms such as ID3, C4.5 and C&RT are used in decision trees. By looking at the entropy values of the features in decision trees, it is decided which one is more decisive in classification. The entropy equation can be expressed as follows (equation 4) [18]:

$$E(A) = -\sum_{i=1}^{n}(p_i log_2 p_i) \quad (4)$$

According to the equation n, indicates the number of states whose entropy will be calculated, and p_i indicates the probability of the state i. The attribute that reduces entropy the most is the attribute that contributes the most to classification success. This is defined as gain in the decision tree.

2.3 Application

Green and black olives were used in the study for classification. A data set was prepared by using 20 pieces of large, 20 pieces of medium and 20 pieces of small olives. The images in this data set were taken from a height of 33 cm, with a 12-megapixel, f /2.2 aperture digital camera. A total of 120 olive images were divided into five sections as training and test data using the k-fold cross validation method in the MATLAB program. The images were first converted to grayscale and the threshold was determined. The image is then rendered binary using this threshold. The gaps caused by the glare on the binary images were removed.

The field, environment, center, diameter attributes and average hue, saturation, and value attributes to be used for classification are determined. These properties are classified as small, medium, large, green, and black using the KNN classification algorithm with Euclidean and Manhattan City Blocks methods and decision tree classification method (Fig 2).

BB-Black big	BM-Black medium
	
BS-Black small	GB-Green big
	
GM-Green medium	GS-Green small
	

Figure 2. Olive classification scale.

3. Results

In the study, dimension classification with KNN was carried out using Euclidean and Manhattan methods using all the features. Classification percentages occurring in various k neighborhood degrees are given in Table 2.

Color classification according to Table 2 data was 100% successful with both methods. In dimension classification, the highest classification success was achieved in k=7 neighborhood degrees in Euclidean method and k=11 neighborhood degrees in Manhattan method.

Table 2. KNN classification results in various neighborhood degrees.

	Euclidean			Manhattan City Blocks	
	Color	Dimension		Color	Dimension
k=3	100%	86,7%	k=3	100%	85,8%
k=5	100%	85,8%	k=5	100%	87,5%
k=7	100%	89,2%	k=7	100%	87,5%
k=9	100%	88,3%	k=9	100%	88,3%
k=11	100%	86,7%	k=11	100%	89,2%
k=13	100%	88,3%	k=13	100%	88,3%

As a result of dimension classification with decision tree, 86.7% classification success was achieved. This value does not change when classifying with all attributes and only when using the field attribute. Color classification has shown 100% success as in KNN methods. When the color classification features were examined, it was seen that the hue value achieved the same success. Decision trees for dimension and color classifications are given in figure 3. In the figure, it is seen that the area for dimension classification and hue value for color classification are used.

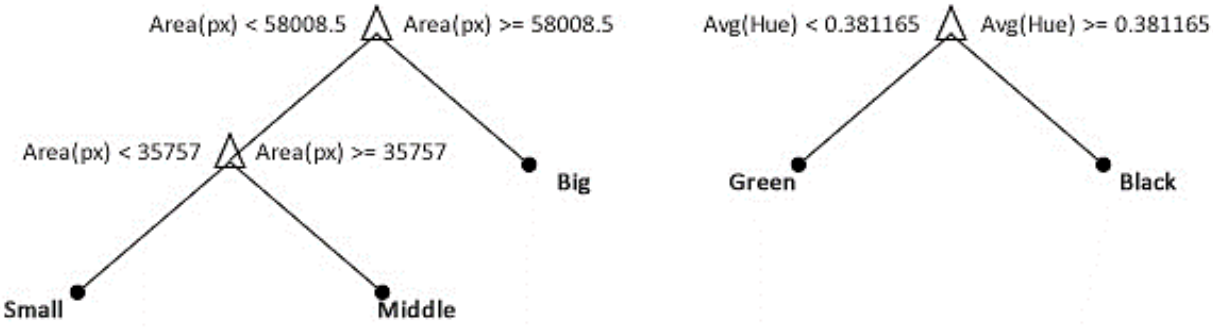

Figure 3. a) Dimension classification tree b) Color classification tree

4. Conclusions and Discussion

In the color classification performed in this study, all methods achieved 100% classification success. According to the KNN dimension classification results used in the study, it can be said that k = 7 in Euclidean method and k = 11 neighborhood degree in Manhattan method. It was observed that the success of both methods decreased after the

optimum k value. In decision tree dimension classification, the algorithm using all attributes and the algorithm using only the field attribute gave the same result. For this reason, it is thought that classification can be made only with the field attribute. It was seen that the hue value in the decision tree color classification could make the classification alone.

In the apple classification made by Sofu et al., the success of color classification was 99% [19]. The size determination of the cherry classification made by Balcı et al. was approximately 85% [8]. Good and bad bean classification was made in the bean classification of Bul et al. and 87% success was obtained by parameter comparison method and 93% by artificial neural network method [1]. 100% classification success was achieved in the nut classification using K-Means clustering algorithm [20]. In the olive extraction machine study carried out by Kuncan et al., 90% success was achieved in HSV method, 80% in Euclidean method and 97% in Mahalanobis method [9]. In a study of image classification based on deformation of olive images, Riquelme et al. found that the color characteristics of the fruit would be sufficient to classify the defects [21]. Again in a defective image detection study in olives, Azarmdel et al. used the RGB, HSV and Lab color parameters to classify olives. In the study performed with two different illumination methods, 90% correct classification success in the method using fluorescent lamp and 75% in the method using LED lamp were obtained [22].

Since the date of the study was not possible to reach the olive fruit and it was not possible to add the "green to pink" class to the color classification. In addition, because of not being found enough variety, "big, medium and small" classification was carried out instead of eleven dimension classes in olive notification. It is thought that all classification types can be reached with a larger data set to be created during the olive harvest period.

References

[1] Bul E, Gelen G, Altun H, 2005, Görüntü İşlemeye Dayalı Tarımsal Ürün Sınıflandırma, *11th National Congress And Fair Of Electrical-Electronic-Computer Engineering*, İstanbul.

[2] Samtaş G, Gülesin M. Digital Image Processing And Its Application In Different Areas, Electronic Journal of Vocational Colleges 2012;2: 85-97.

[3] Göreke V, Uzunhisarcıklı E, Öztoprak B, 2015. A Fuzzy Inference System Design for Computer Aided Mass Detection in Digital Mammogram Images. *Medical Technologies National Congress* 2015, Muğla, 38-41.

[4] Sabancı K, Aydın C, Ünlerşen MF. Determination of Classification Parameters of Potatoes with The Help of Image Processing and Artifical Neural Network, Iğdır University Journal of the Institute of Science and Technology,2012; 2: 59-62.

[5] Oflezer F. An Application Associated With Smart Vehicle Direction Determination, Master of Science Thesis, Yozgat Bozok University, The Institute of Science and Technology, 2015.

[6] Perihanoğlu GM, Feature Extraction From Images By Using Digital Image Processing Techniques, Master of Science Thesis, İstanbul Technical University, The Institute of Science and Technology, İstanbul, 2015.

[7] Altuntaş C, Çorumluoğlu Ö, Improve of Appearance of Digital Image By Filtering Method And A Sampled Code, Journal of Technical-Online 2011; 10:99-107.

[8] Balcı M, Altun AA, Taşdemir Ş, Classification For Napoleon Type Cherries By Using Image Processing Techniques, Journal of Selcuk-Technic 2016; 15:221-237.

[9] Kuncan M, Ertunç HM, Küçükyıldız G, Hızarcı B, Ocak H, Öztürk S, 2013, Görüntü İşleme Tabanlı Zeytin Ayıklama Makinesi, *Auto Control National Meeting*, Malatya, 459-464.

[10] Anonymous, 2018, Death Reason Statistics, 2017 TÜİK [online]. http://www.tuik.gov.tr/PdfGetir.do?id=27620. [Visit Date: 30 April 2019]

[11] Baysal A, Healty Nutrition and Mediterranean Diet, Journal of Nutrition and Diet 1996; 25: 21-29.

[12] Savran MK, 2017, Dünyada ve Türkiye'de Zeytincilik [online], https://www.tarimorman.gov.tr/BUGEM/kumelenme/Belgeler/Budama/Du%CC%88nyada%20ve%20Tu%CC%88rkiye%27de%20Zeytincilik.pdf. [Visit Date: 30 April 2019]

[13] Anonymous, 2019, Turkish Food Codex Olive Notification ID: 2014/33 [online]. https://www.resmigazete.gov.tr/eskiler/2014/08/20140823-7.htm. [Visit Date: 20 April 2019].

[14] Shipman JW, 2012, The hue-saturation-value (HSV) color model [online]. http://infohost.nmt.edu/tcc/help/pubs/colortheory/web/hsv.html. [Visit Date: 26 May 2019].

[15] Taşçı E, Onan A, 2014, The Investigation of Performance Effects of K-Nearest Neighbor Algorithm Parameters on Classification, *16. Academic Information Technologies Conference*, Mersin.

[16] Şeker ŞE. 2008, KNN (K nearest neighborhood)[online]. http://bilgisayarkavramlari.sadievrenseker.com/2008/11/17/knn-k-nearest-neighborhood-en-yakin-k-komsu/. [Visit Date: 12 May 2019]

[17] Çalış A, Kayapınar S, Çetinyokuş T, An Application On Computer And Internet Security With Decision Tree Algorithms In Data Mining, Journal of Industrial Engineering 2014; 25: 2-19.

[18] Larose DT, Discovering Knowledge in Data: An Introduction to Data Mining, New Jersey, USA: Wiley, 2005.

[19] Sofu MM, Er O, Kayacan MC, Çetişli B, Image Processing Method for Determination of Classification and Stain on Apples, Electronic Journal of Food Technologies 2013; 8: 12-25.

[20] Solak S, Altınışık U, Detection and Classification of Hazelnut Fruit By Using Image Processing Techniques And Clustering Methods, Sakarya University Journal of Science 2018; 22: 56-65.

[21] Riquelme M, Barreiro P, Ruiz-Altisent M, Valero C, Olive Classification According to External Damage Using Image Analysis, Journal of Food Engineering 2008; 87: 371-379.

[22] Azarmdel H, Afkari-Sayyah A, Ghaffari H, Alipasandi A, Olive Classification According to RGB, HSV and Lab Color Parameters Using Image Processing Technique, International Journal of Farming and Allied Sciences 2014; 3: 653-658.