

Research Article

Classification of Types of Dates Using Extraction of Shape and Texture Features with K-Nearest Neighbors Method

Muhammad Ichsan¹, Abdul Syukur², Affandy³, Moch. Arief Soeleman⁴

^{1,2,3,4}Faculty of Computer Science, Dian Nuswantoro University, Semarang, Indonesia

Email: ichsan29061997@gmail.com

Academic Editor: Nguyen Ngoc Anh

Copyright © 2022 Muhammad Ichsan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract: Dates are one type of plant that comes from palm trees, dates can also be called in Latin *arecaceae*. There are several types of dates available in the market making it difficult for buyers (consumers) to recognize the types. The purpose of this study was to determine the accuracy of the results of dates from ajwa, grape and mejol dates using the classification of texture and shape features using the K-NN (K-Nearest Neighbors) method. The stages of image processing, namely data processing, segmentation, and extraction using the K-Nearest neighbors method, from this research it is known that the accuracy value is 96.33%.

Keywords: Classification, Texture and Shape Feature Extraction, K-Nearest Neighbors.

A. INTRODUCTION

Dates are a type of plant that comes from palm trees, dates can also be called in Latin *Arecaceae*. Dates are one of the oldest plants that still exist in the world, the harvest from these dates is a major source of livelihood in North Africa and the Middle East, although date palm plants also grow in several parts of the world. Dates production in the world has increased significantly three times from 30 years ago which reached 6.57 million tons in 2011. Dates also contain various properties and very important nutrients that are useful as a treatment in several diseases, dates also increase hemoglobin which contains iron. Dates have various types and various forms, sometimes dates have similarities but there are differences, the uniqueness of dates lies in their great taste and usefulness. Therefore, to distinguish dates is in terms of the color and shape of the dates which are very diverse, so that they can be used as a characteristic or identification of each of these dates [1].

Consumers' knowledge in distinguishing various types of dates still does not understand it. Although there has been so much information about the characteristics and forms of each type of date, the physical test is often wrong and wrong by the limited human vision of the level of subjectivity of the examiner.

Therefore, it is urgently needed a classification of types of dates with form and texture feature extraction is expected to be able to conduct research and testing of the types of dates. To find a solution to a problem that exists today, a study is sought that functions to identify an image of Sukari dates, Ajwa dates, and others using feature extraction and shapes and textures with the K-Nearest Neighbors method. The ultimate goal of this research is to produce and determine a classification that can be used to distinguish and identify several types of dates by using the shape and texture feature extraction method with K-Nearest Neighbors.

Many previous research studies have applied various methods or methods and feature extraction in the identification of dates. Like the research about "Classification of various kinds of dates by using digital image processing techniques using artificial neural networks" [2].

Another study conducted identified the use of date palms with the GLCM method of the android application system, the steps of the image processing process consisted of before processing the fetal level image, how to perform floating segmentation, and also the extraction of second order statistical features using the GLCM method, where the results of his research resulted in a success rate of 83.31% which has 91 training data images and 31 test data images [3].

Research related to feature and shape extraction is a study conducted using onion types as a classification for feature extraction and feature extraction, where in this study there are 5 types of onions: namely onions, red onions, Dayak onions, garlic and Bawanglanang, with the types of onions have 10 images and the total data is 50 images [4].

The next research is research conducted research on noni fruit in classification based on texture and color, where in their research using a process in processing a digital image of the noni type, in their research can prove that the results from the classification of the noni species by using the Support Vector Machines (SVM) algorithm.

Research related to the K-Nearest Neighbors (KNN) method was carried out how to classify various kinds of apples using the K-Nearest Neighbors method, which in this study used various types of apple objects, with the Hue saturation value (HSV) feature. and local binary pattern (LBP) features as extraction of color and shape features of the apples to be studied. Then the results will be classified using the KNN method then the evaluation results can see the average in Precision produced in the study of 93% with an Accuracy of 93% [5].

In this study, the K-Nearest Neighbor method will be used. For shape feature extraction using metric parameters, eccentricity, and texture feature extraction using contrast, energy, correlation, and homogeneity features. Meanwhile (KNN) is a method which aims to be able to classify objects based on data and their closest distance or have similarities to the most significant characteristics with the object.

B. LITERATURE REVIEW

1. Related Research

Mochamad Fandi, Oky Dwi Nurhayati and Rizal Isnantoet. al conducted research on the types and types using the GLCM method, this study identified the types of dates with GLCM using the android application, this image processing process has several stages contained in the processing before the fetal level image method, using development segmentation, and feature extraction. the second order statistic, using the GLCM method, where the results of the research found the result rate of 83.33% which used 91 training image data and 31 tested image data [1].

Novan P. Wijaya and Anugerah adi Ridwan conducted a study on the classification of apples using the K-Nearest Neighbors method. This study identified research related to the K-Nearest Neighbors (KNN) method. The research uses various types of apples as objects, Hue saturation value (HSV) feature with local Binary pattern (LBP) as the extraction of color and shape features on the apples to be studied. Then the results will be classified using the KNN method then the results of the evaluation can see the average generated by 94% with an Accuracy of 94% [2].

Jhoan Angelina W., Herman Santoso P., Edy Budiman Havaluddin & Maratus Sholeha conducted a study on the classification of onion species using the KNN method based on the shape and texture feature extract. From this paper, metric and eccentricity parameters were used. In this extraction, the parameters contrast, correlation, energy, and homogeneity are used. In the feature extraction, 6 parameters will be produced which are processed through classified stages. There are 6 types of onions used in this study, namely Dayak onions, garlic, red onions, onions, garlic, and onions by using a sample of 10 images

with a total sample of 50 images. The result of a thorough 5x test can be observed that in the 50% test of 50 onion samples it produces an accuracy of 87% by having k3, k5, and k7. The results of the accuracy of the use of the value (k) in 5x research are the best, namely on k7 which has an accuracy of 83.13%, but the results of accuracy with 5 times tested are good and the use of (K) is good at 87.56%.

2. K-Nearest Neighbor

K-nearest neighbor or KNN is an algorithm method that has a function to find out the results of data classification by which includes learning data (train data sets), and has been owned by (K) the nearest neighbor (nearest neighbor). K is the nearest neighbor. Proximity can be defined as the distance matrix, between Euclidean distances which can be solved by using equation 1 as follows:

$$D_{xy} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Information:

D: Nearest distance

x: Training data

y: test data

n: the number of attributes from the value 1 to n

f: similarity attribute i between case x and case y

i: attribute from 1 to n

K-Nearest Neighbors can process classification through the projection of material data into a space that has sufficient dimensions. This research can be divided into parts that can be stated in the criteria of the material data. Each of the existing data is applied to points c to many dimensional spaces.

3. Shape Feature Extraction

This research implements extracting shape and texture features in finding pattern recognition. Shape is one of the characteristics that can be extracted from an object so that it can be distinguished from other objects. This shape pattern recognition is a method that uses a mixture of parameters from the eccentricity and metric of the object to the binary image. Eccentricity and metric parameters are methods of feature extraction that mean to be able to take and extract from eccentricity and metric values. Texture is a property that has a character that exists at a point (in the image), which is very large so that naturally the previous properties can repeat at that point. The explanation of texture in this case can be understood as the regularity of certain points on the image formed from a series of pixels in a digital image.

Shape features can be grouped depending on the technique used. The category is based on boundaries (boundary-based) which is based on the region (region-based). Boundary-based technique depicted in the form of a region by the use of external characters, such as pixels along the boundary of the object.

4. Classification

Classification is an algorithm method that can perform and classify and even cluster an object that refers to the character and characteristics it has. In this study, the classification process uses the K-Nearest Neighbor (KNN) method. KNN is a method that can perform the classification stage on objects that refer to theoretical data and the distance is closest or has significant similarity to the object. Near and far neighbors can be calculated from a Euclidean

distance. This method is easy and can provide accurate and good results on the results of the Algorithm K Means classification, and is a non-hierarchical method that was initially taken as part of the many components of the population to be used as a center.

5. Date

Dates are a fruit from the Middle East which is one of the fruits that can be found in Indonesia. This date fruit has a characteristic with the month of Ramadan because it is often found. Dates are a type of fruit that can bear fruit anytime and anywhere, especially in the Middle East with a time of year. But in general, these dates are often harvested in the fall or early winter in their respective countries. the reason is because in winter these dates have fresh and cold temperature conditions. In addition, it also has a delicious taste with a characteristic sweet taste, these dates have many health benefits. Dates contain lots of calcium, vitamin K, iron, folic acid, and anti-oxidants

6. Image Acquisition

Is an analog image that is converted and made into a digital image that can be taken from the environment and used by several tools including: digital camera, webcam, smartphone, and scanner, to be applied to the processing step. Things that can affect a digital image whose results are determined at the time of acquisition is related to the resolution of the tool used. The distance and angle of view an image is taken, its light, size, and magnification, the object or camera in a state of movement does not even move in the resulting image.

C. METHODS

There are two data in a study, where the first is primary data obtained by searching, collecting the data that will be needed, such as observations (observations), interviews and questionnaires. The second is secondary data, meaning that the data obtained comes from sources that already exist on a regular basis, such as data contained in an agency. While this study uses primary data of the first type. Because the data obtained comes from search data collected and observed. The data used are:

1. Ajwa dates
2. Grape Dates
3. Mejol Dates



Fig. 1 Ajwa dates



Fig. 2 Grape dates



Fig. 3 Mejol dates

In this research, the method used at this stage is the K-Nearest Neighbors (KNN) method through the extraction of shape and texture features that can be seen in the stages carried out by creating a computer vision system (image processing).





Figure 4 Stages of Research Methods

D. RESULTS AND DISCUSSION

1. Data Collecting

The dataset obtained from dates taken using a cellphone camera. By shooting 25 cm which is outdoors and free from camera light (flashlight). The position of the dates is vertical with the sides of the dates in the lower position, the background of the picture is white. The date set of image data used in this study found 141 with 50 objects of ajwa dates, 50 objects of grapes, and 41 of mejol.



Figure 5
 Leaf image before
 preprocessing



Figure 6
 Leaf image after
 preprocessing

After obtaining the leaf image after preprocessing, then changing the preprocessed image into a gray image, which is then converted into a binary image.



Figure 7. Result Gray



**Figure 8. Binaryimage
 Treshold**

2. Dataset

The data set used consisted of 141 samples of dates from 3 types of dates, 50 data sets of Ajwa dates, 50 grape dates and 41 Mejol dates.

3. Test

It is known that with the best accuracy value, all testing data that has been prepared is tested. This test used 3 types of dates with a total of 141 images. The following table is obtained from a calculation of confusion matrix data that has been recognized and not recognized, in general it can be concluded that the average precision value obtained is 90-100%, Recall is 88-92% with an accuracy of 93%.

No	Jenis kurma	True ajwa	True anggur	True mejol	Class Precision
1	Pred ajwa	38	4	4	90.48%
2	Pred anggur	5	51	51	91.07%
3	Pred mejol	0	0	0	100.00%
4	Class Recall	88.37%	92.73%	92.73%	

Figure 9 Confusioan Matrix

Confusion Matrix (x: true class, y: pred. class, z: counters)

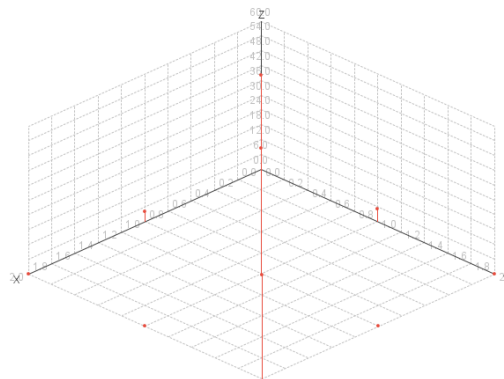


Figure 10 Plot Diagram Forms

E. CONCLUSION

From the explanation that has been described previously, conclusions can be drawn by reading an image of a date fruit for data so that it can be carried out on the next stage using the image processing method. The stages are using shape and texture feature extraction. Attributes that can be taken as information are diameter, width, area, roundness, slenderness and perimeter. while the texture features of the attributes taken are mean, deviation, skewness, energy entropy and smoothness. On all attribute components generated from a form and texture feature extraction process, classification is carried out using the K-Nearest Neighbors method with a value of (k) 3 having an accuracy value of 92.36%, and an RMSE value of 0.292.

REFERENCES

1. Al-Farsi*, M. A., & Lee, C. Y. (2008). Nutritional and functional properties of dates: a review. *Critical reviews in food science and nutrition*, 48(10), 877-887.
2. Andono, N. P, Sutojo, & Muljono. (2017). *Pengolahan Citra Digital*. Andi: Yogyakarta.
3. El-Bendary, N., Hariri, E.E ., Hassanien, A. E., & Snasel, V.(2011). PCA-Based Home Videos Annotation System. *International Journal of Reasoning-based Intelligent System*, 3(2), 71-79.
4. Fandi, M., Nurhayati, O. D., & Isnanto, R. (2020). Aplikasi identifikasi jenis buah kurma dengan metode GLCM berbasis android. *Jurnal Pengembangan Rekayasa dan Teknologi*, 16(1), 34-44.
5. Fazrini, D. (2018). Klasifikasi Penyakit Kanker Usus Besar (Kanker Kolorektal) Menggunakan Probabilistic Neural Network.
6. Gawande, A. P., & Dhande, S. S. (2014). *Implementation of fruit Grading System by Image Processing and Data Classifier-A Review*. *International Journal of Engineering Research and General Science*, 2(6), 411-413.

7. Guo, Y., Han, S., Li, Y., Zhang, C., & Bai, Y. (2018). K-Nearest Neighbor combined with guided filter for hyperspectral image classification. *Procedia Computer Science*, 129, 159-165.
8. Guru, D. S., Sharath, Y. H., & Manjunath, S. (2010). Texture features and KNN in classification of flower images. *IJCA, Special Issue on RTIPPR (1)*, 21-29.
9. Halela, I. A., Nurhadiyono, B., & Rahmanti, F. Z. (2016). Identifikasi Jenis Buah Apel Menggunakan Algoritma K-Nearest Neighbor (KNN) dengan Ekstraksi Fitur Histogram. *Techno. COM*, 1-8.
10. Ibrahim, A., Eissa, A., & Alghannam, A. (2014). Image processing system for automated classification date fruit. *International Journal of Advanced Research*, 2(1), 702-715.
11. Listia, R., & Harjoko, A. (2014). Klasifikasi Massa pada Citra Mammogram Berdasarkan Gray Level Cooccurrence Matrix (GLCM). *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 8(1), 59-68.
12. Manickavasagan, A., Ganeshmoorthy, K., Claereboudt, M. R., Al-Yahyai, R., & Khriji, L. (2014). Non-destructive measurement of total soluble solid (TSS) content of dates using near infrared (NIR) imaging. *Emirates Journal of Food and Agriculture*, 970-976.
13. Novianto, H. (2019). *Kurma Kian Deras Masuk ke Indonesia, Mayoritas dari Mesir*. Retrieved from <https://beritagar.id/artikel-amp/berita/kurma-kian-deras-masuk-ke-indonesia-mayoritas-dari-mesir>
14. Nurpadillah, D. I., Haviluddin, H., Pakpahan, H. S., Islamiyah, I., & Setyadi, H. J. (2019). Pengenalan Karakter Tulisan Menggunakan Metode Learning Vector Quantization. *Sains, Aplikasi, Komputasi dan Teknologi Informasi*, 1(2), 23-29.
15. Prasetyo, E. (2011). *Pengolahan Citra Digital dan Aplikasinya Menggunakan Matlab*. Yogyakarta: Andi.
16. Putra, D. (2010). *Pengolahan Citra Digital*. Yogyakarta: Andi.
17. Putri, Z. (2019). Cara mengetahui perbedaan kurma ajwa asli dan palsu. Retrieved from <https://www.kurmasehat.com/perbedaan-kurma-ajwa-asli-dan-palsu/>
18. Ratnasari, E. K. (2016). Pengenalan jenis buah pada citra menggunakan pendekatan klasifikasi berdasarkan fitur warna lab dan tekstur co-occurrence. *Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, 1(2).
19. Sonata, V. (2015). *Klasifikasi Jenis Buah Kurma Menggunakan Teknik Pengolahan Citra Digital dengan Metode Jaringan Saraf Tiruan* (Doctoral dissertation, Universitas Gadjah Mada).
20. Widians, J. A., Pakpahan, H. S., Budiman, E., Haviluddin, H., & Soleha, M. (2019). Klasifikasi jenis bawang menggunakan metode k-nearest neighbor berdasarkan ekstraksi fitur bentuk dan tekstur. *Jurnal Rekayasa Teknologi Informasi (JURTI)*, 3(2), 139-146.
21. Wijaya, N., & Ridwan, A. (2019). Klasifikasi Jenis Buah Apel Dengan Metode K-Nearest Neighbors Dengan Ekstraksi Fitur HSV dan LBP. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, 8(1), 74-78.