Identification of Dayak Onion with Shape and Texture Feature Extraction by Image Processing

Joan Angelina Widians
Informatics Department, Faculty of Engineering Mulawarman University
Samarinda, East Kalimantan, Indonesia
angel.unmul@gmail.com

Anindita Septi*arini*
Informatics Department, Faculty of Engineering Mulawarman University
Samarinda, East Kalimantan, Indonesia
anindita@unmul.ac.id

Masna Wati
Informatics Department, Faculty of Engineering Mulawarman University
Samarinda, East Kalimantan, Indonesia
masnawatissi@gmail.com

Andi Tejawati
Informatics Department, Faculty of Engineering Mulawarman University
Samarinda, East Kalimantan, Indonesia
tejawatiandi@gmail.com

Maratus Soleha
Informatics Department, Faculty of Engineering Mulawarman University
Samarinda, East Kalimantan, Indonesia
maratussoleha@gmail.com

Tjikoa, Ade Fiqri
Informatics Department, Faculty of Engineering Mulawarman University
Samarinda, East Kalimantan Indonesia
adefiqri5@gmail.com

***Abstract*—Dayak Onion (Eleutherine Palmifolia (L.) Merr) is a medicinal plant that has expanded in East Kalimantan. Humans select the type of onion very quickly, but not for computers. Human perception tends to be subjective to an object because there are factors in the composition of color, shape, and texture owned by the entity thing. The study aims to identify Dayak onions with digital images and apply the K-Nearest Neighbor (KNN) method to extract shape and texture features. There are five types of onions: Dayak onions, Shallot, Bombai onions, Garlic, and Lanang onions. K-Nearest Neighbor (KNN) is a method for objects classification on data learning that is the closest distance or has the most characteristic similarities to an object. This method is for the Dayak Onion's identification and classification. The accuracy level of the KNN method is on shape and texture feature extraction to identifying Dayak onions using 100 image data. Overall, the best accurate result with the KNN method is 86.66%.**

Keywords—Dayak Onion, Shape Feature, Texture Feature, Feature Extraction, K-Nearest Neighbor (KNN), Image Processing.

I. INTRODUCTION

Traditional medicine has become one of the alternatives in demand by the People of Indonesia. One of the plants useful for herbal medicine in East Kalimantan is Dayak Onion (*Eleutherine Palmifolia (L.) Merr*). Onion is a common term for a group of plants belonging to the genus Allium[1] [2] [3]. Tubers, leaves, or onions are used as vegetables or as spices [4]. Select the type of onion based on the characteristics is very easy for humans but not for computers. Human perception tends to be subjective to an object because there are factors in the composition of color, shape, and texture owned by the entity thing [5] . The purpose of a system is to identify Dayak onion, which is done automatically with digital images. Dayak onion (*Eluetherine* Palmifolia (L.) Merr) is native to South America growing in Java, Kalimantan, and Sumatra. Locals in the area already use this plant as traditional medicine. Dayak Onion can utilize an ingredient as traditional medicine [6] [7] . This plant has a different shape and type and has hereditary people used Dayak people as a cure for various types of diseases [6] [8] [9].

The potential of onions as a multi-functional medicinal plant is so great that enhanced information about Dayak onions is necessary [10] [11]. This study aims to identify the characteristics of the image of Dayak Onion using the K-Nearest Neighbor (KNN). The KNN extracts features of the shape and texture of the image of five types of onions, namely Dayak onions, shallots, Bombai onions, garlic, and Lanang onions. K-Nearest Neighbor (KNN) is a method for objects classification on data learning that is the closest distance or has the most characteristic similarities to an object. Calculate the near or far of neighbors with Euclidean distance [12][13][14]. This technique provides accuracy to both identification and classification results. Feature extraction will analyze a feature retrieval of a shape or texture with the value obtained for the identification process [15] . The general steps of designing a computer vision system (image processing and pattern recognition) are in Fig 1.

Engineering Faculty, Mulawarman University .

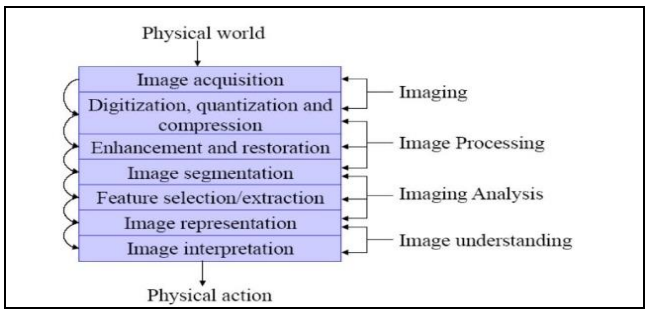

Fig. 1. Image Processing Results.

A. *Image Processing*

Image-processing is a way that converts an image into another image, having better quality for a particular purpose. Image processing is a process of image input, and output can be an image or a set of characteristics or parameters related to the pictures [16] [17] [18].

B. *Image Acquisition*

The Image acquisition is converted process analog images into digital images taken from the real world using digital cameras, webcams, smartphones, and scanners. Factors that affect digital imagery produced during the acquisition process are the resolution of the tool used, the distance and angle of view, the lighting factor, magnification and reduction, the object or camera in motion or not, and the resulting image format [5] [15].

C. *Digitation, quantization and compression*

Image compression reduces the size of the photo to produce a solid or compact digital representation that can still represent the quantity of information contained in the data. In images, video, and audio, compression leads to a reduced bit rate for digital representation [5] [19].

D. *Image Enchancement and Restoration*

Image enhancement is part of the preprocessing stage to improve image quality [16]. This project uses the technique of changing the color level of the image, and the image color becomes a grayscale image. Converting color levels to gray lowers the computing level at the feature capture stage. Grayscaling is simplifying of RGB image models. The three matrix layers are red, green, and blue matrix layers become one grayish matrix layer [20].

E. *Image Segmentation*

Segmentation is part of the preprocessing stage. The purpose of segmentation is to separate a specific foreground object from another object (background). Segmentation output is a binary image, where the foreground's symbol is 1, and the background's symbol is 0 [20]. This research uses thresholding segmentation. Manner to extract the object from the background is to select the threshold T. The global thresholding method is that all pixels in an image convert to black and white with one thresholding value [21]. In this research that global thresholding uses the automatic function of the Otsu method. The Otsu method determine a variable by distinguishing between two or more groups naturally [20].

F. *Feature Selection / Extraction*

The shape is one of the characteristics that can extract from an object to distinguish it from other things. The shape pattern recognition method uses two combinations of eccentricity and metric parameters of an object in a binary image. The eccentricity and measured parameters aims are retrieving or extracting eccentricity and measured values [22] [23].

The texture is a characteristic that an area in the imagery has so that naturally, those properties can recur in that area. The texture is the regularity of specific patterns formed from the arrangement of pixels in the digital image [15]. We use a matrix of intensity pairs (Gray Level Co-occurrence Matrix (GLCM). GLCM is a matrix that describes the frequency at which two pixels pairs appear at a certain intensity in a certain distance and direction in an image. GLCM texture feature extraction produces several features, including contrast, correlation, energy, and homogeneity [19].

G. *Image Representation*

Image representation is the conversion of data results segmentation necessary more suitable images processing [5][18].

H. *Image Interpretation / Pattern Recognition*

Pattern recognition is one of the fields in computer vision on object classification methods into classes to solve the problems. The pattern recognition application performs the glory of the image in one particular based on its [20] [24].

II. RESEARCH METHOD

This research applies the KNN method for the object's classification based on the closest distance data. The method is also for things with the most characteristic similarities. Calculate Neighbor distance by Euclidean. This technique is simple and expected to provide good accuracy to the classification results and determine the identification of the characteristics of Dayak onion. Identification of Dayak onions (*Eleutherine Palmifolia (L.) Merr*) by classifying the type of onion by the K-Nearest Neighbor (KNN) method based on the extraction of shape and texture feature requires image input [12]. This study uses onions, namely Dayak onion, Shallot, Bombai onions, Garlic, and Lanang onions. Onions that identified five types in the Fig 2, 3, 4, 5, 6.

Fig. 2. Dayak Onion (*Eleutherine palmifolia (L.) Merr*).

Fig. 3. Shallot (*Allium cepa L. Var. Aggregatum).*

Fig. 4. Bombai Onion *(Allium Cepa Linnaeus).*

Fig. 5. Garlic (*Allium sativum*).

Fig. 6. Lanang Onion (*Allium sativum linn*).

The initial data processing way is converting images into grayscale images. This stage facilitates the image process, namely the segmentation process. Segmentation using thresholding where the process on the difference in the grayness's degree is a grayscale image. Grayscale images of onions become input in the segmentation process. The threshold value, the threshold value (T), is automatically determined by the Otsu method. Grayscale and segmentation processes in Fig 7.

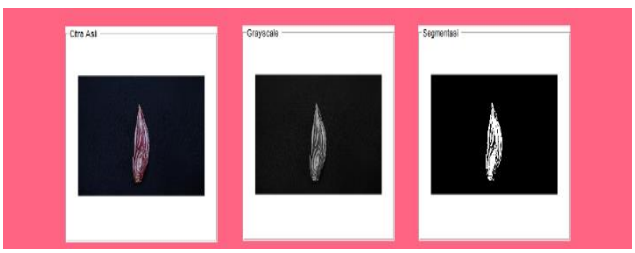

Fig. 7. Preprocessing Results.

A. Feature Extraction

The extraction parameters of texture features used in this assessment are contrast, correlation, energy, and homogeneity. Feature extraction of the onion images using grayscale images. The grayscale images will produce a Grayscale matrix is calculated by Gray Level Co-occurrence Matrix (GLCM) method. The results GLCM method with parameters contrast, correlation, energy, and homogeneity. Based on five times the test that often arises errors and may not be recognized, namely the type of garlic in code D01 using k3 always errors. This condition occurs due to garlic has a feature extraction parameter value that is almost similar to onions. The results of shape and feature texture extraction are in Table 1.

TABLE I. RESULT OF FEATURE EXTRACTION

No	Code	Metric	Eccen-tricity	Contrast	Correlation	Energy	Homo-geneity	Onion
1	A01	0.0342	0.9498	0.0972	0.9010	0.8760	0.9772	Dayak Onion
…	…	…	…	…	…	…	…	…
20	A10	0.9365	0.9595	0.0764	0.8716	0.8849	0.9783	Dayak Onion
21	B01	0.3141	0.5230	0.1003	0.9297	0.7833	0.9562	Shallot
…	…	…	…	…	…	…	…	…
40	B10	0.9652	0.6928	0.0056	0.9776	0.9734	0.9979	Shallot
41	C01	0.6241	0.6053	0.0344	0.9934	0.5683	0.9859	Bombai onion
…	…	…	…	…	…	…	…	…
60	C10	0.8463	0.2963	0.0415	0.9933	0.7029	0.9858	Bombai onion
61	D01	0.7287	0.9010	0.0122	0.9823	0.9546	0.9963	Garlic
…	…	…	…	…	…	…	…	…
80	D10	0.9276	0.6405	0.0131	0.9867	0.9310	0.9967	Garlic
81	E01	0.5523	0.6114	0.0162	0.9862	0.9388	0.9956	Lanang onion
…	…	…	…	…	…	…	…	…
100	E10	0.9467	0.5433	0.0074	0.9860	0.9569	0.9978	Lanang Onion

According to the Table 1, the metric parameter is a magnitude that indicates the degree of roundness of an object's shape. Metric results obtained by Equation (1).

$$M = \frac{4\pi \times A}{C^2} \quad (1)$$

The eccentricity parameter is a symbolic comparison value between the distance of minor ellipse foci and the significant foci ellipse of an object. Eccentricity results are obtained by Equation (2).

$$e = \sqrt{1 - \frac{b^2}{a^2}} \quad (2)$$

The contrast parameter is a feature that represents the difference between the color level or grayscale that appears in an image. Contrast results by Equation (3).

$$Contrast = \sum_i \sum_j (i-j)^2 \, p_{(i,j)} \quad (3)$$

The correlation parameter represents the linear association of degrees of the grayish image. The correlation results by Equation (4).

$$Correlation = \sum_i \sum_j \frac{(i-\mu i)(j-\mu j)p_{(i,j)}}{\sigma i \, \sigma j} \quad (4)$$

The energy parameter states the measure of uniformity in the image. The higher the image likeness, the higher the energy value. Energy results by Equation (5).

$$Energy = \sum_i \sum_j p(i,j)^2 \quad (5)$$

The homogeneity parameter describes the size of the uniformity. Homogeneity will be a high value if all pixels have a uniform value. Calculation of energy results contained in Equation (6).

$$Homogeneity = \sum_i \sum_j \frac{p(i,j)}{1+|i-j|} \quad (6)$$

The result of the extraction parameter value features shapes and textures in Table 1. Implementation of the equation (1) - (6) is a graph in Fig 8.

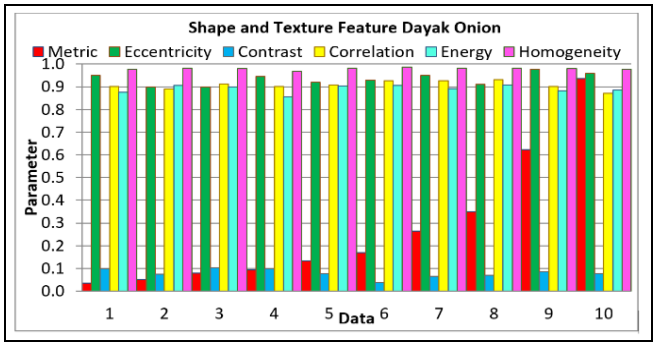

Fig. 8. Extraction Feature Results.

Fig 8 describes the extraction shape and texture feature of Dayak Onion. The chart indicates the highest metric is 0.9365, and the lowest is 0.0342. Nevertheless, the highest eccentricity parameter is 0.9777, and the lowest is 0.8993. The highest contrast is 0.1030, and the lowest is 0.0379. While the highest correlation parameter is 0.9304, and the lowest is 0.8716. Whereas the highest is 0.9078, and the lowest is 0.8547. Furthermore, the highest homogeneity parameter is 0.9872, and the lowest is 0.9685.

B. *Method Accuracy*

The accuracy level of the KNN method is on shape and texture feature extraction to identifying Dayak onions using 100 image data from onions using k3, k5, k7, in Fig 9.

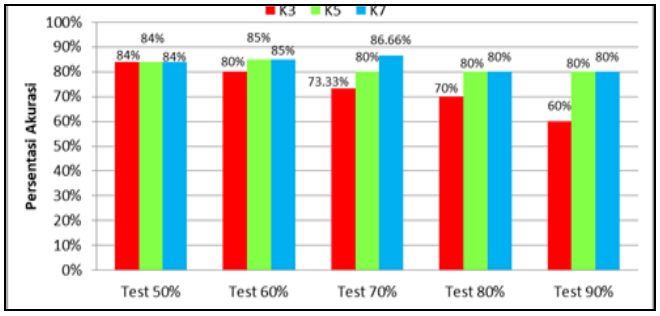

Fig. 9. Testing Graph.

Fig 9 shows the highest accuracy level of the KNN method exists in k3 and 50% testing with an accuracy score of 83.56%. While, the highest accuracy of k7 in 70% testing is 86.66% and the lowest on 80% and 90% testing with score 80%. Overall, the best accurate KNN method is 86.66%.

III. Conclusion

The conclusions are:

1. Dayak onion shape and texture features extraction produces six parameters: metric, eccentricity, contrast, correlation, energy, and homogeneity.
2. The K-Nearest Neighbor (KNN) method extracts shape and texture features to identify Dayak onions. The best accurate identify Dayak Onion using the KNN method is 86.66%.

Acknowledgment

The Engineering Faculty, Mulawarman University, fund this research. The author would like to thank BKSDA East Kalimantan, Indonesia, for providing the Dayak Onion and clinical data to test our proposed.

References

[1] E. Budiman, U. Hairah, Haeruddin, A. Tejawati, S. Darmawan, and S. Wahyuni, "Biodiversity information system of medicinal plants from tropical rainforest Borneo based on traditional knowledge ethnic of Dayak," *Adv. Sci. Lett.*, vol. 24, no. 11, pp. 8668–8673, 2018, doi: 10.1166/asl.2018.12321.

[2] F. Falah and N. Hadiwibowo, "Species identification of traditional medicine plants for women's health in East Kalimantan: lesson learned from local wisdom," *Indones. J. For. Res.*, vol. 4, no. 1, pp. 49–67, 2017.

[3] J. Angelina Widians, M. Wati, A. Tejawati, and E. Budiman, "Biodiversity information system for management of medicinal plants data tropical rainforest Borneo," *Int. J. Eng. Technol.*, vol. 7, no. 4.44, p. 31, 2018, doi: 10.14419/ijet.v7i4.44.26858.

[4] S. H. Noorcahyati, "Tumbuhan berkasiat obat etnis asli Kalimantan," *Balai Penelit. Teknol. Konserv. Sumber Daya Alam Balikpapan*, 2012.

[5] A. Septiarini, H. Hamdani, H. R. Hatta, and A. A. Kasim, "Image-based processing for ripeness classification of oil palm fruit," in *2019 5th International Conference on Science in Information Technology (ICSITech)*, 2019, pp. 23–26.

[6] J. A. Widians, N. Puspitasari, and A. Febriansyah, "Disease diagnosis system using certainty factor," in *2019 International Conference on Electrical, Electronics and Information Engineering (ICEEIE)*, 2019, vol. 6, pp. 303–308.

[7] E. M. Kuntorini and M. D. Astuti, "Penentuan Aktivitas Antioksidan Ektrak Etanol Bulbus Bawang Dayak," *Sains dan Terap. Kim.*, vol. 4, no. 1, pp. 15–22, 2010.

[8] M. Apuy, A. M. Lahjie, B. Simarangkir, Y. Ruslim, and R. Kristiningrum, "Traditional plants in forest gardens of West Kutai, Indonesia: Production and financial sustainability," *Biodiversitas J. Biol. Divers.*, vol. 18, no. 3, pp. 1207–1217, 2017.

[9] E. Budiman, M. Wati, U. Hairah, F. Alameka, and M. Jamil, "Intelligent dDecision support systems of medicinal forest plants for kkin disease," in *2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, 2021, pp. 354–359.

[10] N. Dengen, E. Budiman, J. A. Widians, M. Wati, U. Hairah, and M. Ugiarto, "Biodiversity information system: Tropical rainforest borneo and traditional knowledge ethnic of dayak," *J. Telecommun. Electron. Comput. Eng.*, vol. 10, no. 1–9, 2018.

[11] N. Puspitasari, J. A. Widians, E. Budiman, M. Wati, and A. E. Ramadhan, "Dayak Onion (Eleutherine palmifolia (L) Merr) as an alternative treatment in early detection of dental caries using certainty factor," *2020 3rd Int. Semin. Res. Inf. Technol. Intell. Syst. ISRITI 2020*, no. L, pp. 482–487, 2020, doi: 10.1109/ISRITI51436.2020.9315469.

[12] J. A. Widians, H. S. Pakpahan, E. Budiman, H. Haviluddin, and M. Soleha, "Klasifikasi jenis bawang menggunakan metode k-nearest neighbor berdasarkan ekstraksi fitur bentuk dan tekstur," *J.*

Rekayasa Teknol. Inf., vol. 3, no. 2, p. 139, 2019, doi: 10.30872/jurti.v3i2.3213.

[13] A. Gokhale, S. Babar, S. Gawade, and S. Jadhav, "Identification of medicinal plant using image processing and machine learning," in *Applied Computer Vision and Image Processing*, Springer, 2020, pp. 272–282.

[14] A. Septiarini, A. Harjoko, R. Pulungan, and R. Ekantini, "Automatic detection of peripapillary atrophy in retinal fundus images using statistical features," *Biomed. Signal Process. Control*, vol. 45, pp. 151–159, 2018.

[15] M. Wati, N. Puspitasari, E. Budiman, and R. Rahim, "First-order feature extraction methods for image texture and melanoma skin cancer detection," in *Journal of Physics: Conference Series*, 2019, vol. 1230, no. 1, p. 12013.

[16] A. Septiarini, H. Hamdani, H. Rahmania, and K. Anwar, "Scientia horticulturae automatic image segmentation of oil palm fruits by applying the contour- based approach," *Sci. Hortic. (Amsterdam).*, vol. 261, no. November 2019, p. 108939, 2020, doi: 10.1016/j.scienta.2019.108939.

[17] A. Septiarini, A. Sunyoto, H. Hamdani, A. A. Kasim, F. Utaminingrum, and H. R. Hatta, "Machine vision for the maturity classification of oil palm fresh fruit bunches based on color and texture features," *Sci. Hortic. (Amsterdam).*, vol. 286, p. 110245, 2021.

[18] A. Marion, *Introduction to image processing*. Springer, 2013.

[19] A. Septiarini, A. Harjoko, R. Pulungan, and R. Ekantini, "Automated detection of retinal nerve fiber layer by texture-based analysis for glaucoma evaluation," *Healthc. Inform. Res.*, vol. 24, no. 4, pp. 335–345, 2018.

[20] M. M. P. Petrou and C. Petrou, *Image processing: the fundamentals*. John Wiley & Sons, 2010.

[21] A. Septiarini, H. Hamdani, H. R. Hatta, and K. Anwar, "Automatic image segmentation of oil palm fruits by applying the contour-based approach," *Sci. Hortic. (Amsterdam).*, vol. 261, p. 108939, 2020.

[22] A. A. Kasim, R. Wardoyo, and A. Harjoko, "Batik classification with artificial neural network based on texture-shape feature of main ornament," *Int. J. Intell. Syst. Appl.*, vol. 11, no. 6, p. 55, 2017.

[23] Y. F. Riti, H. A. Nugroho, S. Wibirama, B. Windarta, and L. Choridah, "Feature extraction for lesion margin characteristic classification from CT Scan lungs image," in *2016 1st International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 2016, pp. 54–58.

[24] L. M. Abualigah and A. T. Khader, "Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering," *J. Supercomput.*, vol. 73, no. 11, pp. 4773–4795, 2017.