ATLANTIS PRESS

Advances in Engineering Research, volume 208

Proceedings of the International Conference on Innovation in Science and Technology (ICIST 2020)

Improving Image Quality Using Color Intensity Modification to Determine the Ripeness of Avocado

Budi Hartono*, Yunus Anis, Veronica Lusiana
Department of Informatics Engineering
Stikubank University
Semarang, Indonesia
*budihartono@edu.unisbank.ac.id

***Abstract*—Image with better quality will make the next processing easier. This study aims to analyze the result of image quality improvement using color intensity modification to determine the ripeness of avocado. Color intensity modification used two methods, i.e. Histogram Equalization (HE) and Contrast Stretching (CS). The test data was 96 images 100x100 pixel in JPEG format. Image feature extraction used Gray Level Co-occurrence Matrix (GLCM) by selecting texture feature. The results of K-means clustering with two centroids showed that there is error, ranging from 2.08% to 4.17%. After improving image quality, determination of the ripeness of avocado became better. Both techniques can improve the quality or clarity of image object. The research result could be used for consideration in selecting image pre-processing method.**

Keywords—adaptive histogram equalization, contrast stretching, GLCM, texture feature

I. INTRODUCTION

An image with good contrast is an image that has a wide intensity distribution. The histogram does not show a dominant peak. On the other hand, an image with low contrast is an image that has a narrow intensity distribution. Improving image quality can be done by stretching the contrast and equalizing the histogram [1,2]. After obtaining an image in a better contrast condition, the next step is to perform quantization. In this study, quantization aims to reduce the number of variations in the pixel values appearing in the image. Images with limited pixel value variations are expected to facilitate the process of finding image objects.

Digital image feature extraction is a way to obtain image features through the contents of the image [3]. Characteristics that can describe an image can also be obtained through the file name or description of the contents of the image through key words and sentences. The gray level co-occurrence matrix (GLCM) can be used to calculate several texture features [4]. This study used three features, namely energy, contrast, and homogeneity. All features used the 0° direction. These three features produced three pairs of feature combinations, namely energy with contrast, energy with homogeneity, and contrast with homogeneity.

This research analyzed the results of the image quality improvement of avocado using color intensity modification. The color intensity modification process used the contrast stretching method and the histogram equalization method. It can be seen that the image resulting from the histogram equalization has a better visual appearance with clearer object. The image after going through these two processes is in better condition, so that the identification of ripe avocado and raw avocado is better. K-means is a clustering algorithm, the purpose of this algorithm is to divide data into several groups, with input in the form of data without class labels. In this algorithm, the computer groups the data into its own input without first knowing the target class [5]. The research materials were avocado fruit images obtained from datasets by Škrjanec [6] and Oltean [7].

II. RELATED WORK

A. Image Enhancement

Color intensity modification in an image is related to the distribution of pixel intensity. The modification here is intended to clarify the existence of an object contained in the image. This process is needed to assist image processing related to visual objects. Two methods that are often used to modify color intensity are the histogram equalization and contrast stretching [3].

Contrast in an image is related to the distribution of pixel intensity, namely the process of expanding the range of intensity [2]. Contrast stretching (CS) is one way to improve image quality through point surgery. The technique that is often used for histogram processing is histogram equalization (HE) [8,9]. It is used to produce a uniform or even histogram so it is often referred to as histogram alignment.

B. Feature Extraction

After the image features are obtained, there are many benefits that can be obtained from this information, including recognizing objects in the image. Three image features commonly used to obtain image features are color, object shape, and texture. The description and application of feature extraction in more depth is explained by Nixon and Aguado [10].

The energy feature is used to measure the concentration of intensity pairs. The energy value enlarges when the pixel pair is concentrated on several coordinates, and vice versa will decrease if the location is spread out. The contrast feature is used to measure the strength of the difference in intensity in the image. The contrast value increases when the variation in intensity in the image is high. On the other hand, it becomes smaller when the variation is low. The homogeneity feature is the inverse of the contrast value, which is to measure the homogeneity of intensity variations in the image [11,12]. The homogeneity value increases if the intensity variation in the image is low. Table 1 lists some research related to image enhancement, feature extraction, and gray level co-occurrence matrix (GLCM).

TABLE I. SEVERAL PUBLICATIONS RELATING TO THIS RESEARCH

No.	Researchers	Purpose	Object	Method
1	Mustaghfirin, et al. [13]	Detection	Iris	Histogram Equalization and Adaptive Histogram Equalization
2	Widodo, et al. [14]	Improve Image Quality	Dental radiology	Contrast Stretching
3	Syahputra, et al. [15]	Recognition	Leaf, Indonesian medicinal plants	GLCM, Moment invariant, and Shape morphology
4	Dewi, et al. [16]	Classification	Patchouli plantation	Color moment, GLCM
5	Sutarno and Fauliah [17]	Classification	Durian Fruit	GLCM, LVQ
6	Andrian, et al. [18]	Identification	Butterfly	GLCM, KNN

C. *Gray Level Co-occurrence Matrix (GLCM)*

The co-occurrence intensity matrix is a matrix that describes the frequency of occurrence of pairs of two pixels with a certain intensity over a certain distance and direction in the image. The GLCM matrix contains gray level co-occurrence data for neighbors in the 0°, 45°, 90°, and 135° directions [12]. These directions are obtained from the pixel intensity as the center with the closest neighboring pixel which has 1 distance.

III. RESEARCH DESIGN

The stages in this research are convert rgb image to gray, image enhancement, Quantization, texture feature extraction, and Clustering. The complete research stages and the methods used can be seen in Figure 1. The gray values are obtained from the average values of R, G, and B from RGB data image. The process of changing the original image or RGB image to gray is aimed at reducing the vector dimensions so that the feature extraction computation process becomes lighter.

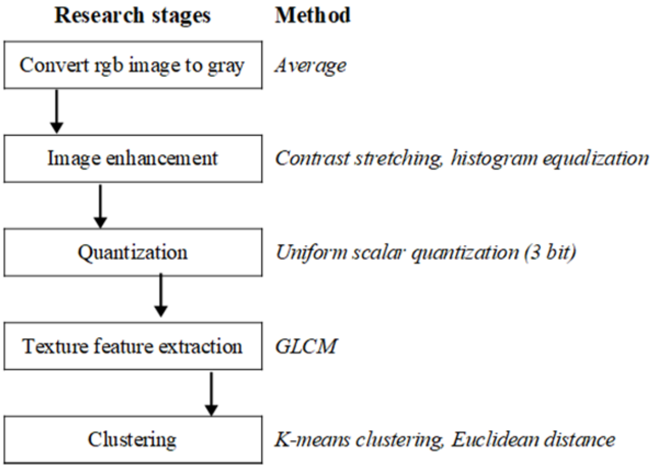

Fig. 1. Research stages.

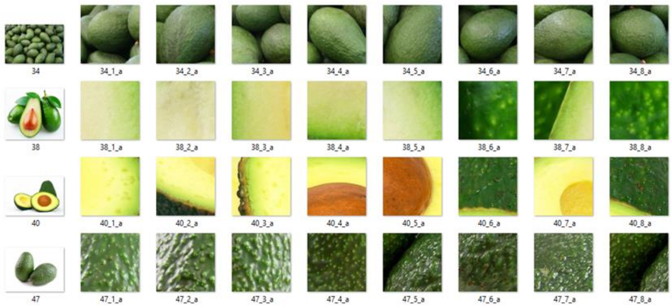

Fig. 2. Raw Avocado data.

In this study, we used uniform scalar quantization with an output of 3 bits, which was obtained from the input 8-bit grey value. The quantization output provided an alternate grey value of 8 units. Experimental image data obtained from datasets [6,7]. In Figure 2 we can see 32 images of raw avocados. In Figure 3 and Figure 4 there are 64 ripe avocados images.

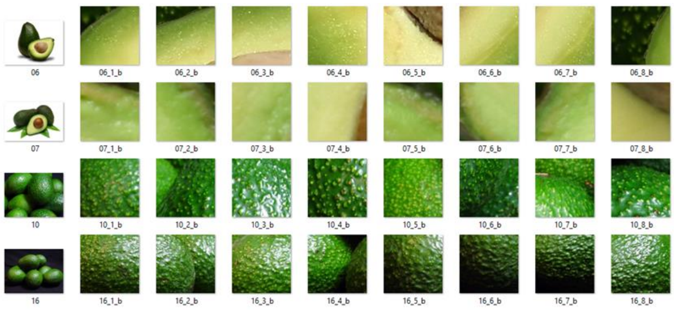

Fig. 3. Ripe Avocado data (i).

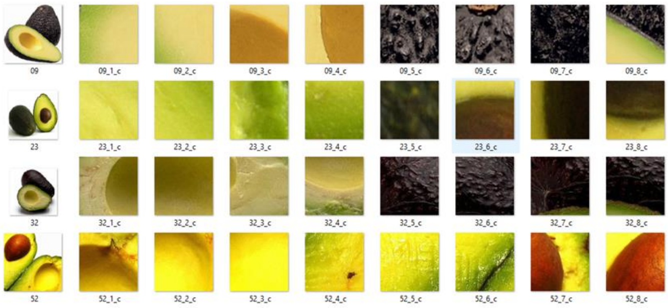

Fig. 4. Ripe Avocado data (ii).

A. Quantization

There are two kinds of quantization, namely scalar quantizers and vector quantizers. In grouping input values, scalar quantization uses a single parameter, for example via gray values. Meanwhile, vector quantization is used to classify input values using two or more values. Vector quantization requires a larger amount of computation [19].

Scalar quantization differentiates the ways to classify input values by means of a uniform (uniform scalar quantizer) or by a distribution function (nonuniform scalar quantizer). Uniform scalar quantization groups the input values with the same width so that the division resembles a rung shape, where the group count will correspond to the selected quantization output bit [19]. The number of the group here becomes the maximum variation number of the quantization result or the quantization output.

B. K-Means Clustering

There are two types of data clustering that are often used in the data grouping process, namely Hierarchical and Non-Hierarchical. K-Means is a non-hierarchical or Partitional Clustering method. The K-Means Clustering method seeks to classify existing data into several groups, where the data in one group have the same characteristics and have different characteristics from the data in other groups. The K-Means Clustering method aims to minimize the objective function set in the clustering process by minimizing variations between data in a cluster and maximizing variation with data in other clusters [20].

C. Euclidean Distance

The method of measuring the similarity of the image can be done by determining the level of similarity (similarity degree) or dissimilarity (dissimilarity degree) of the two feature vectors of the image. The image feature vector is obtained through the image feature extraction process. The degree of similarity is a value that indicates whether the two vectors are similar or not. The similarity measurement method can be applied in general to various types of data, not limited to image data [3].

Some of the commonly used similarity measurement methods are Minkowski Distance, Euclidean Distance, Cityblock (Manhattan) Distance, Hamming Distance, Mahalanobis Distance, Canberra Distance, and others. Euclidean distance is a measurement method that is often used to calculate the similarity of 2 vectors. Euclidean distance computes the roots of the squares of the difference between 2 vectors [5,20,21]. The Euclidean distance between the two vectors p and q can be calculated using the following equation (1).

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + ... + (p_n - q_n)^2} \quad (1)$$

IV. Results and Discussion

The image from histogram equalization and contrast stretching process has a better visual appearance and clearer object. In the histogram, after the histogram equalization process, the pixel intensity redistribution process occurs so that it becomes more evenly distributed. Intensities 0 to 255 are almost all represented, as can be seen in Figure 5 and Figure 6. When viewed from the x-axis of the histogram, the pixel intensity values are relatively evenly distributed and there is an even distribution of the frequency of the appearance of the pixels.

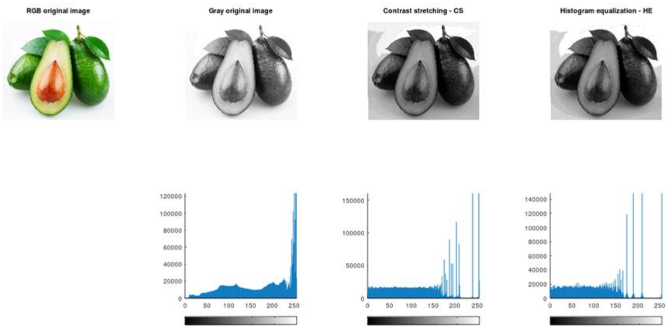

Fig. 5. Improve the image quality of raw Avocado.

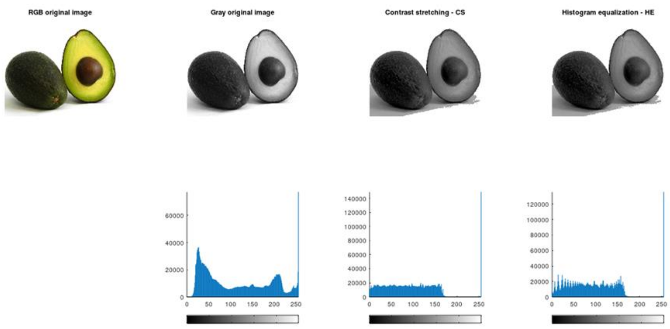

Fig. 6. Improve the image quality of ripe Avocado.

The Euclidean distance is the straight-line distance between two points in Euclidean space. The application of Euclidean distance can be implemented to calculate the distance between data and centroids in the K-means clustering. This is often used because the distance calculation in this distance space is the shortest distance that can be obtained between two calculated points. The results of the Euclidean distance calculation show that the smaller the d (p, q) score, the more similar the two matched feature vectors are.

Through Figure 7 and Figure 8 can be observed graphic K-means clustering with energy and homogeneity feature from contrast stretching and histogram equalization. Meanwhile, Figure 9 and Figure 10 can be observed graphic K-means clustering with contrast and homogeneity feature from contrast stretching and histogram equalization. All features are obtained from GLCM matrix for neighbors in the 0° directions.

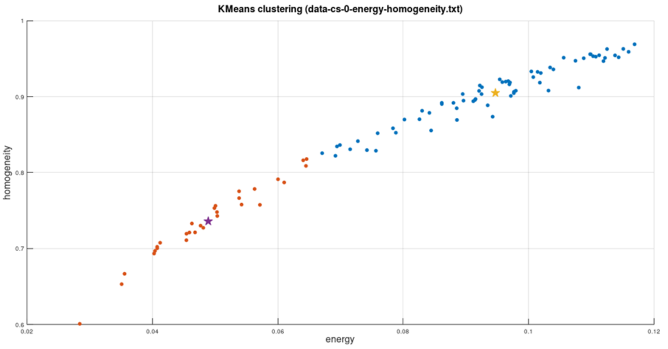

Fig. 7. Graphic K-means clustering energy and homogeneity feature from contrast stretching.

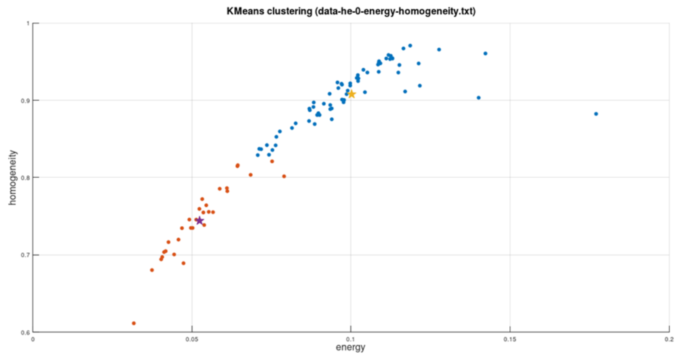

Fig. 8. Graphic K-means clustering energy and homogeneity feature from histogram equalization.

This learning is included in unsupervised learning. Inputs received are data or objects and the desired clusters. This algorithm will classify data or objects into these k groups. In each cluster, there is a center point (centroid) which represents the cluster [5]. Table 2 records all feature value data. In this study, the researcher intended to know to what extent the selection of the combination energy, homogeneity, and contrast features with histogram equalization and contrast stretching method can increase recognize of ripe avocado and raw avocado. The total number of data is 96 images, consisting of 64 pieces of ripe avocado image and 32 pieces of raw avocado image. The results of K-means clustering with two centroids showed that 3, 2, 3, and 4 raw avocado images were included in the ripe avocado image group. When compared with the total data, there is cluster error from 2.08% to 4.17%.

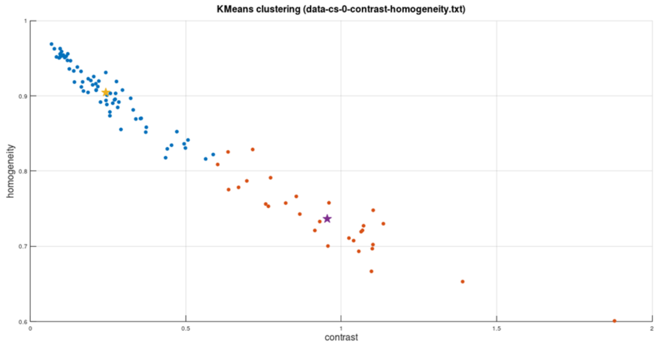

Fig. 9. Graphic K-means clustering contrast and homogeneity feature from contrast stretching.

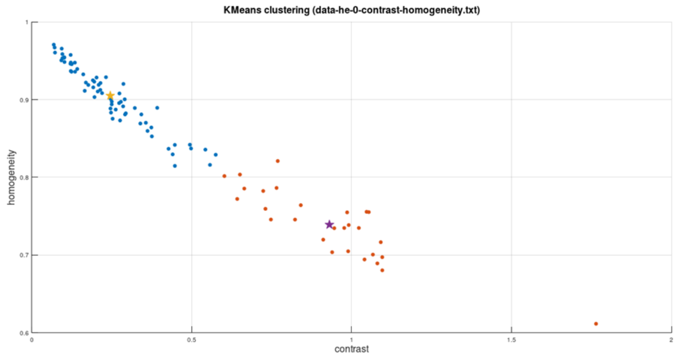

Fig. 10. Graphic K-means clustering contrast and homogeneity feature from histogram equalization.

TABLE II. RESUME OF FEATURE AND METHOD OF ENHANCING IMAGE QUALITY

No.	Feature	Method of enhancing image quality	Ripe Avocado (C1)	Raw Avocado (C2)	Error (%)
1	Energy and homogeneity	Contrast stretching	67	29	3.125
2	Energy and homogeneity	Histogram equalization	66	30	2.080
3	Contrast and homogeneity	Contrast stretching	67	29	3.125
4	Contrast and homogeneity	Histogram equalization	68	28	4.170

V. CONCLUSION

This study aims to analyze the result of image quality improvement using color intensity modification to determine the ripeness of avocado. Color intensity modification used two methods, i.e. Histogram Equalization (HE) and Contrast Stretching (CS). Image feature extraction used Gray Level Co-occurrence Matrix (GLCM) by selecting texture feature. The results of K-means clustering with two centroids showed that there is cluster error, ranging from 2.08% to 4.17%. After improving image quality, determination of the ripeness of avocado became better. Both techniques can improve the quality or clarity of image object. The research result could be used for consideration in selecting image pre-processing method. Further development of this research can be applied on recognition the ripeness of other fruits.

ACKNOWLEDGMENT

This work was supported by research institutions and community service (LPPM) and scientific publication technical unit, Universitas Stikubank Semarang, Indonesia.

REFERENCES

[1] Y. Xie, L. Ning, M. Wang, and C. Li, "Image Enhancement Based on Histogram Equalization," IOP Conf. Series: Journal of Physics: Conf. Series, vol. 1314, p. 012161, 2019.

[2] R.C. Gonzalez and R.E. Woods, Digital Image Processing, Global Edition, Fourth Edition. London, UK: Pearson Education, 2018.

[3] D. Putra, Pengolahan Citra Digital. Yogyakarta: Penerbit Andi, 2010.

[4] A. Kadir dan A. Susanto, Teori dan Aplikasi Pengolahan Citra. Yogyakarta: Penerbit Andi, 2013.

[5] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques 3rd Edition. US: Morgan Kaufmann, 2011.

[6] M. Škrjanec, Automatic fruit recognition using computer vision. Bsc Thesis, (Mentor: Matej Kristan), Fakulteta za računalništvo in informatiko, Univerza v Ljubljani, 2013.

[7] M. Oltean, "Fruits 360 dataset," Mendeley Data, V1, 2018.

[8] W.A. Mustafa and M.M.M.A. Kader, "A Review of Histogram Equalization Techniques in Image Enhancement Application," IOP Conf. Series: Journal of Physics, vol. 1019, p. 012026, 2018.

[9] J. Lee, S.R. Pant, and H.S. Lee, "An Adaptive Histogram Equalization Based Local Technique for Contrast Preserving Image Enhancement," International Journal of Fuzzy Logic and Intelligent Systems, vol. 15, no. 1, pp. 35-44, 2015.

[10] M.S. Nixon and A.S. Aguado, Feature Extraction and Image Processing for Computer Vision, Fourth Edition. Amsterdam, Netherlands: Elsevier, Academic Press, 2020.

[11] R.M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," IEEE Transactions on Systems Man & Cybernetics, vol. 3, pp. 610-621, 1973.

[12] U. Ahmad, Pengolahan Citra Digital dan Teknik Pemrogramannya, 1st ed. Yogyakarta: Graha Ilmu, 2005.

[13] F. Mustaghfirin, Erwin, H.K. Putra, U. Yanti, and R. Ricadonna, "The Comparison of Iris Detection Using Histogram Equalization and Adaptive Histogram Equalization Methods," IOP Conf. Series: Journal of Physics, vol. 1196, p. 012016, 2019.

[14] H.B. Widodo, A. Soelaiman, Y. Ramadhani, and R. Supriyanti, "Calculating Contrast Stretching Variables in Order to Improve Dental Radiology Image Quality," IOP Conf. Series: Materials Science and Engineering, vol. 105, p. 012002, 2016.

[15] H. Syahputra, Z. Indra, D. Febrian, and D.P. Adriani, "Leaf feature extraction using glcm, moment invariant and shape morphology for indonesian medicinal plants recognition," IOP Conf. Series: Journal of Physics: Conf. Series, vol. 1317, p. 012008, 2019.

[16] C. Dewi, M.A. Grahadhuita, and L. Muflikhah, "Color moment and gray level co-occurrence matrix in classification of soil organic matter for patchouli plantation," Indonesian Journal of Electrical Engineering and Computer Science, vol. 19, no. 2, pp. 983-991, 2020.

[17] Sutarno and S.P. Fauliah, "Implementation of Learning Vector Quantization (LVQ) Algorithm for Durian Fruit Classification Using Gray Level Co-occurrence Matrix (GLCM) Parameters," IOP Conf. Series: Journal of Physics: Conf. Series, vol. 1196, no. 012040, 2019.

[18] R. Andrian, D. Maharani, M.A. Muhammad, and A. Junaidi, "Butterfly identification using gray level co-occurrence matrix (glcm) extraction feature and k-nearest neighbor (knn) classification," Jurnal Register, vol. 6, no. 1, pp. 11-21, 2020.

[19] P.D. Symes, Video Compression Demystified. NY: McGraw-Hill Professional Publishing, 2000.

[20] I.H. Witten, E. Frank, M.A. Hall, and C.J. Pal, Data Mining Practical Machine Learning Tools and Techniques 4th Edition. US: Morgan Kaufmann, Elsevier, 2020.

[21] S.K. Ali, Z.M. Aydam, and B.M. Rashed, "Similarity metrics for classification: A Review," IOP Conf. Series: Materials Science and Engineering, vol. 928, p. 032052, 2020.