

# Fuzzy C-Means segmentation of mango fruit

Linda Marlinda<sup>1)\*</sup>, Muhamad Fatchan<sup>2)</sup>, Widiyawati<sup>3)</sup>, Faruq Aziz<sup>4)</sup>, Wahyu Indrarti<sup>5)</sup>

<sup>1,4)</sup> Universitas Nusa Mandiri, <sup>2)</sup> Universitas Pelita Bangsa, <sup>3)</sup> STMIK Bani Saleh, <sup>5)</sup> Universitas Bina Sarana Informatika

<sup>1)</sup>[linda.ldm@nusamandiri.ac.id](mailto:linda.ldm@nusamandiri.ac.id), <sup>2)</sup>[fatchan@pelitabangsa.ac.id](mailto:fatchan@pelitabangsa.ac.id), <sup>3)</sup>[widiyawati.zul@gmail.com](mailto:widiyawati.zul@gmail.com),

<sup>4)</sup>[14002420@nusamandiri.ac.id](mailto:14002420@nusamandiri.ac.id), <sup>5)</sup>[wahyu.wii@bsi.ac.id](mailto:wahyu.wii@bsi.ac.id)

**Submitted** : Mar 31, 2021 | **Accepted** : Apr 8, 2021 | **Published** : Apr 17, 2021

**Abstract:** In measuring the maturity of mangoes traditionally, it can be seen from image analysis based on texture and smooth skin color. Problems found in mango fruit segmentation are images that are influenced by several noise factors and environmental objects. The mango peel segmentation process is needed so that the classification process can be carried out better. The segmented image will read the feature extraction value of an object that has been separated from the background. The procedure on the image that has been analyzed is the pattern recognition process. In this process, the segmented image is divided into several parts according to the desired object acquisition. The clustering technique is used to segment the image by grouping data according to its class and partitioning the data into mango datasets. Fuzzy C Means (FCM) method is used to produce optimal results in determining clustering-based image segmentation. Based on the results of the trials that have been carried out in this study, it is proven that the proposed method can be used to classify mangoes with an accuracy rate of 87%. The final result of Fuzzy C-based mango segmentation processing means that the available feature extraction value with a maximum number of iterations is 31 iterations, error (x) = 0.00000001, and image computation testing time is 2444.913636. Digital image processing using the FCM method results in a classification of mango ripeness so that it can be used as a tool in analyzing mangoes..

**Keywords:** Clustering; Fuzzy C Means; Image; Mango; Segmentation

## INTRODUCTION

In Indonesia, the mango fruit is a seasonal fruit that has many types and flavors. This plant can grow in the lowlands and highlands( et al., 2018). This fruit plant has a strong tree trunk and many branches, the bark is thick and rough and has many leaves. The fruit maturity problem of manga has a bright yellow color but sour taste, but there is also bright green but sweet taste(Siswanto & Utama, 2017). From these problems, a method is needed so that consumers can find out the level of maturity of the mango fruit with a sweet taste(soffiana agustin, 2011). Many problems arise when doing traditional maturity identification. For mango growers, this level of ripeness is very easy for them distinguish, but for ordinary people will certainly experience a lot of difficulties. This problem is due to human nature which has several weaknesses, including weaknesses caused by physical limitations and fatigue. With the advancement of computer technology makes work humans are getting faster and easier. The problem of classifying mango fruit ripeness can be solved by applying computer vision science, enabling devices to recognize and analyze objects in the form of images taken to recognize ripeness conditions. This ability will obviously be very helpful, especially for those who do not have knowledge about the selection of mango ripeness. The method according to the success rate of mango fruit maturity is based on color normalization using the Fuzzy C Means method(Lv et al., 2019).

To analyze the image of mango fruit, it can be recognized from the pattern recognition, image analysis is used to divide an image into several parts and take some of the desired object. In the field of image processing. The process of dividing the image by digital segmentation into multiple regions with a set of pixels. This segmentation process can simplify an image image so that the image is easily analyzed using the clustering method(Bhargava & Bansal, 2018). Cluster-based grouping is done by grouping data by class and partitioning the data in the form of a dataset.

Problems will arise in the image segmentation process with several contributing factors so that the resulting image is unsatisfactory. Some of the factors that experience noise in the image of mango images have blurry or

\* Corresponding Author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

blurry colors, causing color degradation, this is due to noise in the lighting image when shooting mangoes (Lv et al., 2019)(Siswanto & Utama, 2017).

Research in the field of image processing is related Fruit maturity through color detection of fruit skin, using either a computer or smartphone, has been done by several previous researchers. Among them is a study entitled Classification of Mango Fruit Based on the Level of Maturity Using the Least-Squares Support Vector Machine, the results of using the GLCM and K-Means Clustering Methods for Feature Extraction combined with the LS-SVM method for classification can be used to obtain maturity levels for more than one type of fruit. Mango with high accuracy values. From the test results on the Harum Manis, Manalagi, and Kent mango with the classification label Matang and Rawah, the accuracy, precision and recall results are obtained with the same value, namely 98.33%. The number of features used for the classification process greatly affects the level of accuracy and computation time, the more features used, the higher the accuracy rate but makes the computation time longer, whereas if you use fewer features, the computation time tends to be faster but will get an accuracy value. relatively lower ( et al., 2018).

Then there is also research with the same method but different objects entitled Implementation of Image Segmentation with Using the Generalized Fuzzy C-Means Clustering Algorithm with Improved Fuzzy Partitions method, the results of using the FCM method for segmenting an object in terms of validity functions (partition coefficient and partition entropy) get good results and the number of iterations is less than other methods (Hardiyanto et al., 2012).

From the above problems, to overcome the disturbance in the mango fruit image based on color, the clustering technique uses the Fuzzy C-Means algorithm so that it can determine the maturity of the mango fruit with an accuracy rate of 87%.

## LITERATURE REVIEW

### Image Segmentation

Image segmentation is an image processing that studies the improvement of images that have image noise so that they can process images into quality and clean(Xiong et al., 2020). The division of an image can use the imaging technique into several regions (regions) so that it has similar colors based on the attributes that follow. This segmentation can improve digital image processing and remove noise related to image quality, including image restoration, contrast enhancement, and color transformation. Image processing is based on transformation by processing images in rotation, translation, scale, geometric transformations(Bhargava & Bansal, 2018). This process will select an image based on the feature images so that it will produce an optimal image that can be analyzed. Information processing in image recognition can be done using data compression or reduction, this processing can store and generate data transmission for optimal processing time. The fuzzy c mean method for segmentation in the input process uses image data and outputs the processed image(Hardiyanto et al., 2012).

Image segmentation grouping consists of low and high-level segmentations. The division of the image into several parts of the smallest area using mean shift, watershed, level set, and a superpixel is called low segmentation level, usually dividing the image into several small areas and is over-segmented and produces good segmentation image processing. The high-level segmentation method, which is region merging and graph cut, produces images that are clean from noise(Veling, 2019).

### Clustering

The process of grouping data in segmentation processing uses clustering(Andika & Hafiz, 2018). The clustering process has a high degree of similarity between data with one another, but it will be different from image data with other clusters. Clustering depends on the image data and the application used, this process is a form of compression of the sample data which is converted into a small number of representative prototypes or clusters. In the process of class identification in the figure, the size and similarity of controls on how clusters are formed will be seen, including distance, connectivity, and intensity(Mahardika, 2019).

Image images that use clustering-based segmentation have multi-dimensional data in grouping image pixels into several clusters, this data has many attributes and components in the preparation of an image. The processed image has three RGB dimensions with a segmentation image based on one cluster of mango images which are grayscale images. This image has pixels that have proximity and then clustered(Wu & Liu, 2020). This research uses Fuzzy C means because of the success rate of the feature grouping process in one cluster where one image pixel can be owned by several clusters in determining its members.

### Fuzzy C-Means

One of the techniques used for image segmentation processing using the Fuzzy c-means (FCM) algorithm, this algorithm can determine the centroid and calculate the average position of the cluster data iteratively when

evaluating an objective function (Jia et al., 2020). Fuzzy C-Means is a clustering technique or data grouping in several classes or clusters so that it has a level of similarity between data but is very different from data in other clusters. The existence of points in the data is determined by the degree of membership in a cluster (Zhang et al., 2020). To obtain accurate data, processing is required by determining cluster points based on the degree of membership. The minimized objective function based on the repetition process describes the distance and data points to the center of the cluster weighted by the degree of membership. To minimize the function due to the degree of membership at a data point that has more than one group. Fuzzy C-Means can group large data, is more robust against data outliers and is very simple, and easy to implement (Wu & Chen, 2020)

## METHOD

### Research data

The method used in this research is the Fuzzy C-Means (FCM) method to obtain the color separation results in the image. In this study, the test data used were the mango fruit image. The sizes used in these images are 256 X 256 pixels respectively (Hardiyanto et al., 2012). FCM clustered data in several class based on degree of data point to obtain accurate data.

### Data analysis

The stages of the research stages in data analysis are as follows (Veling, 2019):

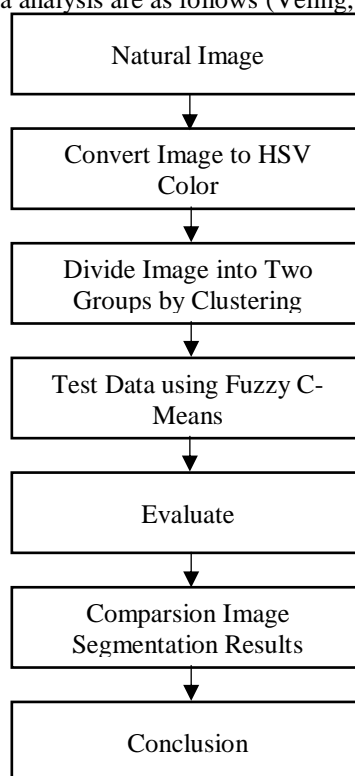


Fig. 1. Research workflow

This study uses the Fuzzy C-Means Clustering (FCM) method. FCM was chosen because it has good segmentation results. The algorithm of this method is as follows (C-, 2015):

1. Input data to be clustered (X), with an  $n \times m$  matrix, where:
  - $n$  = number of data samples.
  - $m$  = attribute of each data.
  - $X_{ij}$  = data sample  $i$ - $i$  ( $i = 1, 2, \dots, n$ ), attribute  $j$  ( $j = 1, 2, \dots, m$ ).
2. Specify some variable initial value of the parameter
  - The number of clusters ( $c$ ) that will be used
  - The amount of data ( $n$ ) that will be used
  - Maximum iteration = MaxIter

Error (e)

The initial objective function (P0) = 0;

Initial iteration (t = 1) (Xiong et al., 2020)

- The random values used are  $\mu_{ik}$ ,  $i = 1, 2, \dots, n$ ;  $k = 1, 2, \dots, c$ ; as  $u$  initial partition matrix.  $\mu_{ik}$  is a membership that has degrees and will show the probability of data being a member of a cluster. The random collection will determine the position and value of the matrix. Internal 0 - 1 is assigned to the cluster membership value. Generating a random number ( $\mu_{ik}$ ) where  $i = (1, 2, 3, \dots, n)$ ;  $k = (1, 2, 3, \dots, c)$  as the element of the initial partition matrix  $U$  (Rosyani & Saprudin, 2020).

The number of each attribute:

$$Q_i = \sum_{k=1}^c u_{ik}$$

Where:  $Q_j$  is the number of degrees of membership per column = 1, where  $j = 1, 2, \dots, m$  then count the number of each column (attribute) (Lei & Ouyang, 2019).

Calculate:

$$u_{ik} = \frac{u_{ik}}{Q_i} \quad (1)$$

- Calculating the center of the cluster to  $-k$ :  $v_k$  where  $k = 1, 2, \dots, c$ , using the equation (2)

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \quad (2)$$

- Calculating the objective function of the  $-t$  iteration, using the equation (3)

$$P_t = \sum_{i=1}^n \sum_{k=1}^c ([\sum_{i=1}^m (X_{ii} - V_{kj})^2] (\mu_{ik})^w) \quad (3)$$

- Calculate the change in the partition matrix, where  $h = 1, 2, \dots, c$ , using the equation (4)

$$\mu_{ik} = \frac{[\sum_{i=1}^m (X_{ii} - V_{kj})^2]^{-\frac{1}{w-1}}}{\sum_{k=1}^c [\sum_{i=1}^m (X_{ii} - V_{kj})^2]^{-\frac{1}{w-1}}} \quad (4)$$

- Check condition means stop, if  $P_t - P_{t-1} < \epsilon$  ) atau ( $t > \text{MaxIter}$ ) then stop, if not:  $t = t + 1$ , repeat step 4

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \quad (5)$$

## RESULT

### Image test data

The test image used is the image of the mango fruit, the size used in the image is 256 x 256 pixels each.



Fig. 2. Original image mango

### Test Image

The test image used is divided into 2 clusters following the criteria to be carried out.

### Transformation of Test Image to HSV Color Space

The test image that has been divided into 3 color images is then transformed into the HSV color space. The transformation process to the HSV color space is carried out by finding the average value of the red, green, and blue matrix data in the image and then changing it into the form of hue, saturation, and value values which will

then be processed using the clustering method. The image transformation process in the RGB to HSV color space uses equations.

### Image Testing The test for mango.jpg using HSV

This section will divide the area (clustering) of the test image, namely mangga.jpg by using the HSV color space. The initialization process begins with transforming into the HSV color space, with a test image can be seen in the image

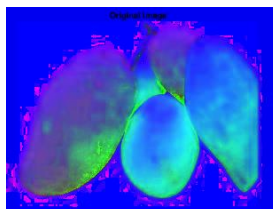


Fig. 3 Image Testing The test for mango.jpg using HSV

The observations from Figures 2 and 3 show that the HSV color space can distinguish parts of the image area more accurately than the RGB color space. The Hue (H) component, the Saturation component (S), and the Value (V) component of the test image can be seen in Figure 2, while the Red (R) component, the Green (G) component, and the Blue (B) component can be seen in Figure 3. The components will become inputs for the FCM method. In the first experiment enter R, G, B, and the second experiment H, S, V. Is a big challenge in the image segmentation process. With the development of image segmentation methods, clustering of images with detailed color density is deliberately raised as a challenge to obtain better segmentation methods. A good segmentation method is expected to be able to handle increasingly complex segmentation problems.

Testing the results of clustering The mango.jpg test image uses the Color Space.



Fig. 4. Image Labeled by cluster Index

In this section, the clustering of the test image 1, namely mango.jpg, will be carried out using the FCM method in the HSV and RGB color spaces. The first step in this experiment is that the 1st test image is transformed into the HSV color space, then the clustering process is carried out using the FCM method, then the 1st test image is transformed into RGB color space. In this experiment, clustering was carried out into 3 parts of the area which were expected to be able to distinguish green leaves, disease-affected leaves, and dry leaves in Test Image 1. Test Image 1 can be seen in Figure 5



Fig. 5. Object in cluster 1

The observations show that the FCM method in the HSV color space detects parts of the area more accurately than the FCM method in the RGB color space. The results of clustering using the FCM method are shown in Image 4. In this figure, it can be seen that the division of regions is more accurate in each region. The results of clustering with the FCM method using the HSV color space show that some parts are not detected properly, as in figure 6.



Fig. 6. Object in cluster 2

Areas that have different colors in the image have several areas that experience detection errors on objects, it is very difficult to find differences in the results of the two methods with the naked eye, but when using the help of the HSV color space, the differences will be seen, so the detection results of the FCM method are more accurate.

### DISCUSSION

The computation time required for clustering in the HSV color space by the FCM method has almost as fast results. However, in this image, the time required is slightly faster than the FCM method in the RGB color space. The observations in figure 7 show that the computation time required by the FCM method in the HSV color space is 2444.913636 seconds and requires 31 iterations.

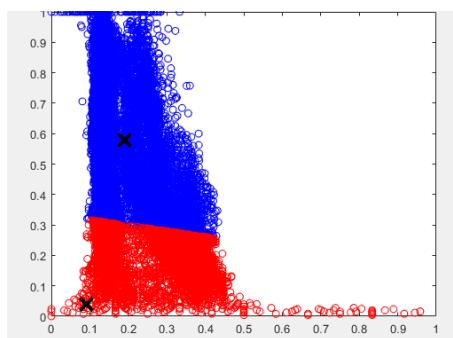


Fig. 7. Fuzzy c-means clustering

Parameters determined in the segmentation process are using Fuzzy C-Means, where the number of clusters ( $c$ ) is set from 2 clusters. With the maximum number of iterations (MaxIter) of 31 iterations, the smallest error ( $\xi$ ) = 0.00000001, and testing with ground-truth with the iteration table as follows. Observation Results of the Clustering Process on the Test Image computation time 2444.913636 with 31 iterations

### CONCLUSION

The conclusion of this study is the Clustering Technique, which can produce differences in each area according to the degree of a gray color. The Fuzzy C Means method in the HSV color space and information on the division of regions (clusters) can produce more detailed and accurate areas in the HSV color space. Fuzzy C Means method in RGB color space and information on the division of regions (clusters) several results divided regions cannot be detected properly in segmentation. The experimental results show that the FCM method in the HSV color space can provide more accurate segmentation results than the RGB color space.

### REFERENCES

- Andika, T. H., & Hafiz, A. (2018). Analisis Perbandingan Segmentasi Citra Menggunakan Metode K-Means dan Fuzzy C-Means. *Seminar Nasional Teknologi Dan Bisnis 2018*, 237–246.
- Bagus, C., & Imron, M. (2018). Klasifikasi Buah Mangga Berdasarkan Tingkat Kematangan Menggunakan Least-Squares Support Vector Machine. *Explore IT : Jurnal Keilmuan Dan Aplikasi Teknik Informatika*, 10(2), 1–8. <https://doi.org/10.35891/explorit.v10i2.1255>
- Bhargava, A., & Bansal, A. (2018). Fruits and vegetables quality evaluation using computer vision: A review. *Journal of King Saud University - Computer and Information Sciences*. <https://doi.org/10.1016/j.jksuci.2018.06.002>
- C-, F. (2015). *Implementasi Algoritma Fuzzy C Means Dan Statistical Region Merging Pada Segmentasi Citra*. 9–10.
- Hardiyanto, I., Purwananto, Y., Kom, S., Kom, M., & Soelaiman, R. (2012). Implementasi Segmentasi Citra dengan Menggunakan Metode Generalized Fuzzy C- Means Clustering Algorithm with Improved Fuzzy

Partitions. *Teknik Pomits*, 1(1), 1–5.

- Jia, X., Lei, T., Du, X., Liu, S., Meng, H., & Nandi, A. K. (2020). Robust Self-Sparse Fuzzy Clustering for Image Segmentation. *IEEE Access*, 8, 146182–146195. <https://doi.org/10.1109/ACCESS.2020.3015270>
- Kaswar, A. B., Palopo, U. C., Morfologi, O., & Pendahuluan, I. (2018). *itra buah mengkudu membentuk klaster hyperellipsoid pada ruang fitur*. 8, 13–20.
- Lei, X., & Ouyang, H. (2019). Image segmentation algorithm based on improved fuzzy clustering. *Cluster Computing*, 22, 13911–13921. <https://doi.org/10.1007/s10586-018-2128-9>
- Lv, J., Ni, H., Wang, Q., Yang, B., & Xu, L. (2019). A segmentation method of red apple image. *Scientia Horticulturae*, 256(March), 108615. <https://doi.org/10.1016/j.scienta.2019.108615>
- Mahardika, F. (2019). Penerapan Segmentasi Warna pada Gambar di Media Sosial dengan Algoritma Fuzzy K-Means Cluster. *Simetris : Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 10(2), 631–638.
- Rosyani, P., & Saprudin, S. (2020). Deteksi Citra Bunga Menggunakan Analisis Segmentasi Fuzzy C-Means dan Otsu Threshold. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 20(1), 29–36. <https://doi.org/10.30812/matrik.v20i1.715>
- Siswanto, & Utama, G. P. (2017). Segmentasi Pada Citra Buah Mangga Menggunakan Aplikasi Matlab. *Bit*, 14(2), 9–17.
- soffiana agustin, eko prasetyo. (2011). *Klasifikasi jenis pohon mangga gadung dan curut berdasarkan tesktur daun*. 58–64.
- Veling, P. S. (2019). Mango Disease Detection by using Image Processing. *International Journal for Research in Applied Science and Engineering Technology*, 7(4), 3717–3726. <https://doi.org/10.22214/ijraset.2019.4624>
- Wu, C., & Chen, Y. (2020). Adaptive entropy weighted picture fuzzy clustering algorithm with spatial information for image segmentation. *Applied Soft Computing Journal*, 86, 105888. <https://doi.org/10.1016/j.asoc.2019.105888>
- Wu, C., & Liu, N. (2020). Suppressed robust picture fuzzy clustering for image segmentation. *Soft Computing*, 8(2015). <https://doi.org/10.1007/s00500-020-05403-8>
- Xiong, J., Liu, Z., Chen, S., Liu, B., Zheng, Z., Zhong, Z., Yang, Z., & Peng, H. (2020). Visual detection of green mangoes by an unmanned aerial vehicle in orchards based on a deep learning method. *Biosystems Engineering*, 194, 261–272. <https://doi.org/10.1016/j.biosystemseng.2020.04.006>
- Zhang, X., Jian, M., Sun, Y., Wang, H., & Zhang, C. (2020). Improving image segmentation based on patch-weighted distance and fuzzy clustering. *Multimedia Tools and Applications*, 79(1–2), 633–657. <https://doi.org/10.1007/s11042-019-08041-x>