# Identification using the K-Means Clustering and Gray Level Co-occurance Matrix (GLCM) At Maturity Fruit Oil Head

**Sukemi**
Computer System Departement, Universitas Sriwijaya
Palembang, Indonesia
sukemi@ilkom.unsri.ac.id

**Edi Sukrisno**
Computer System Departement, Universitas Sriwijaya
Palembang, Indonesia
edysk23@ilkom.unsri.ac.id

***Abstract*— This study discusses the identification of its palm fruit camp picture image using K-Means Clustering and identification GLCM. Process fruits traditionally experienced constrained due to human nature that has flaws that the desired results are not effective. Advances in computer technology have come into the world in terms of the farm before harvest and post-harvest. This Dimolished how to recognize the fruit so that it correspond to real conditions. Condition of oil palm fruit is determined by the level of maturity in terms of color, texture and shape of the oil palm fruit. Identification which did classify in the category of mature and not mature. Determination of identification with the K-means clustering method that uses the difference in euclidean distance and GLCM feature extraction as a reference. For the results of the present study is equal to 90% of the 50 test data.**

***Keywords*— K-Means Clustering,GLCM,Euclidean Distance**

## I. INTRODUCTION

Parameter ripeness is usually determined from the parameters of which are of the size, weight, color characteristics, the shape of the fruit, and others. On the side of the fruit skin color is an important feature in the identification of fruit maturity. At maturity level of oil palm fruits can be seen directly from the image of oil palm fruit itself [1][2][3].

The image of oil palm fruit has a dark purple color variant to orange depending on the maturity level of the oil palm fruit. The difficulty in perceiving the ripeness humans use their senses of sight are subjective and inconsistent that can differ from one appraiser with other assessors. The development of increasingly advanced technology allows the identification of ripeness even detection of the type of fruit with the help of computers [4][5][6][7].

In previous research on the identification of the maturity of an apple with GLCM by Maura Widyaningsih, explained that the use of feature extraction GLCM method used is a texture that can perform well identification based on value extraction. In another study conducted using k-means clustering algorithm as the segmentation process with an average success rate to get the accuracy above 60%. Based on the description above, this research will develop a system for the identification of oil palm fruit maturity on digital photo image of oil palm fruit by applying the method of K-means clustering and GLCM. The results of this study are expected to be able to identify on the photo image of oil palm fruit with a high degree of accuracy results. [8][9][10][11].

## II. RESEARCH METHODOLOGY

This study will go through steps such as image acquisition, segmentation, preprocessing, feature extraction and identification. After taking the image of oil palm fruits do image segmentation is done by using the K-Means clustering so that the object divide fruit with a background part. Results of segmentation then performed on the object feature extraction of fruit with GLCM. Last for object recognition with the closest distance calculation (euclidean distance). The value of each object identifier will be offset by contrast, correlation, homogeneity, and energy. Stages methodology in this study can be seen in the block diagram in Figure 1.

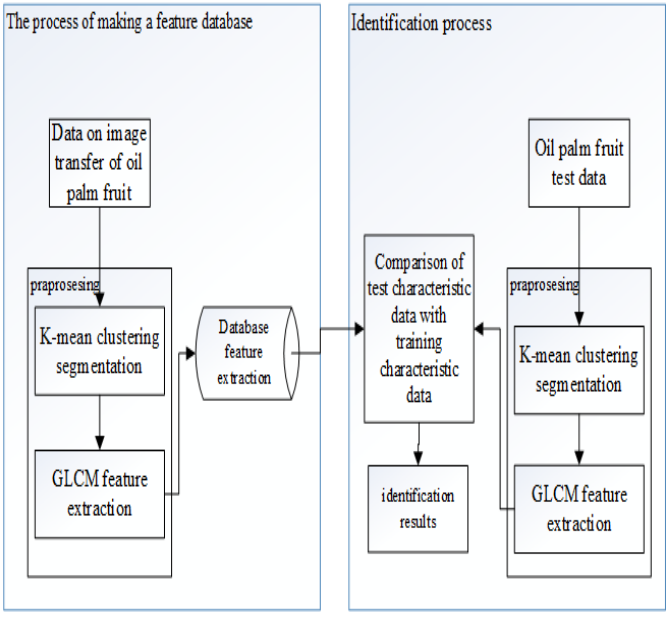


Figure 1. Block the system works

### A. *Image Segmentation K-means clustering*

The initial stage is done is to take the value of RGB (Red, Green, and Blue) of each pixel image Furthermore, each pixel image will be grouped using an algorithm K-Means that will produce fruit part of the image that will be used for further processing: feature extraction [12][13][14][15].

*B. Feature Extraction GLCM*

A feature extraction making process important information contained in the image of oil palm fruits. one important feature of the image of oil palm fruits to be used in the introduction of oil palm fruit maturity is the shape and texture. Based on the explanation, this research will be used Gray Level Co-occurrence Matrices in the process of feature extraction. This process will make retrieval component color pixel grayscale and then calculating the value of Contrast, Correlation, homogeneity, and Energy [16][17][18]. All image data will have a different hash value, so that the test images will be determined value of similarity with the image characteristic database. For the flowchart can be seen in the image below:

In this study GLCM features that are used to perform feature extraction process, namely:

1. *Contrast*

Formula Contrast can be seen in Equation 1.

$$Contrast = \sum_{t=0}^{N-1} n^2 \left\{ \sum_{i=1}^{N} \sum_{j=1}^{N} p_{i,j} \right\}, |i-j| = n \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

2. *Correlation*

Correlation shows a linear dependence of the degree of gray pixels neighboring each other in a gray image. The formula can be seen in Equation

$$Correlastion = \sum_{i=1}^{N} \sum_{j=1}^{N} p_{i,j} \frac{(i-\mu_i)(j-\mu_j)}{\sigma_i \sigma_j} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Where:

N= many degrees of gray of the image
$\mu_i$= the average value of a column of the matrix element Pθ (i,j)
$\mu_j$= the average value of a column of the matrix element Pθ (i,j)
$\sigma_i$= standard deviation value Pθ matrix column element Pθ(i,j)
$\sigma_j$= standard deviation value Pθ matrix column element Pθ(i,j)

3. *Energy*

Indicates the size of the local homogeneity and is the opposite of entropy. To get the energy use Equation 3.

$$ASM = \sum_{i=1}^{N} \sum_{j=1}^{N} \{p_{i,j}\}^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Where: Pi, j = matrix that has been done normalization

4. Homogenitas *Homogeneity*

Homogeneity size of an image can be calculated by Equation 4.

$$Homogeneity = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{1+(i-j)^2} p_{i,j} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4)$$

*C. Identifiication*

The results of the classification of test image data using the Euclidean Distance that uses the closest distance calculation to the similarity of the test data to the training data. The end result is the value of the closest distance from the test image palm oil data to the provisions of the category. Data for the database is determined by 8 data which are divided into 4 mature data and 4 data are not mature. Data testing will be carried out in an introduction one by one, and information will be given as a result of the proximity of the data to the database, so that the closest data will provide information on the category. Identification process can be seen below:

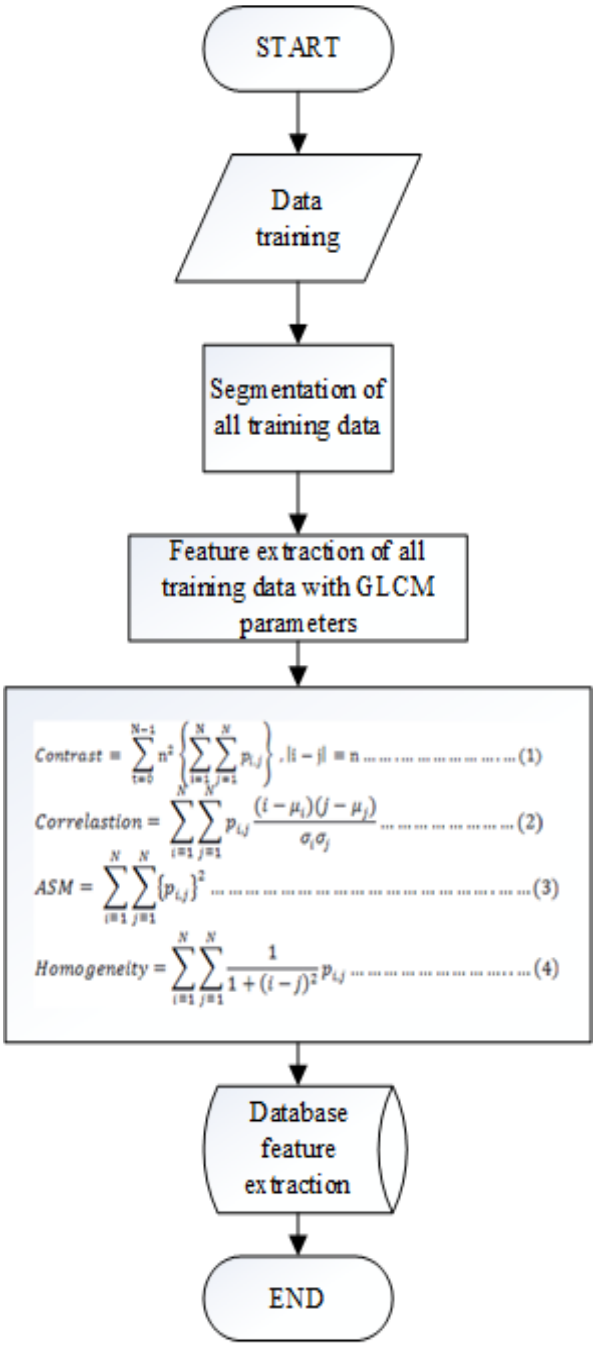


Figure 2. Flowchart the formation of Database feature extraction

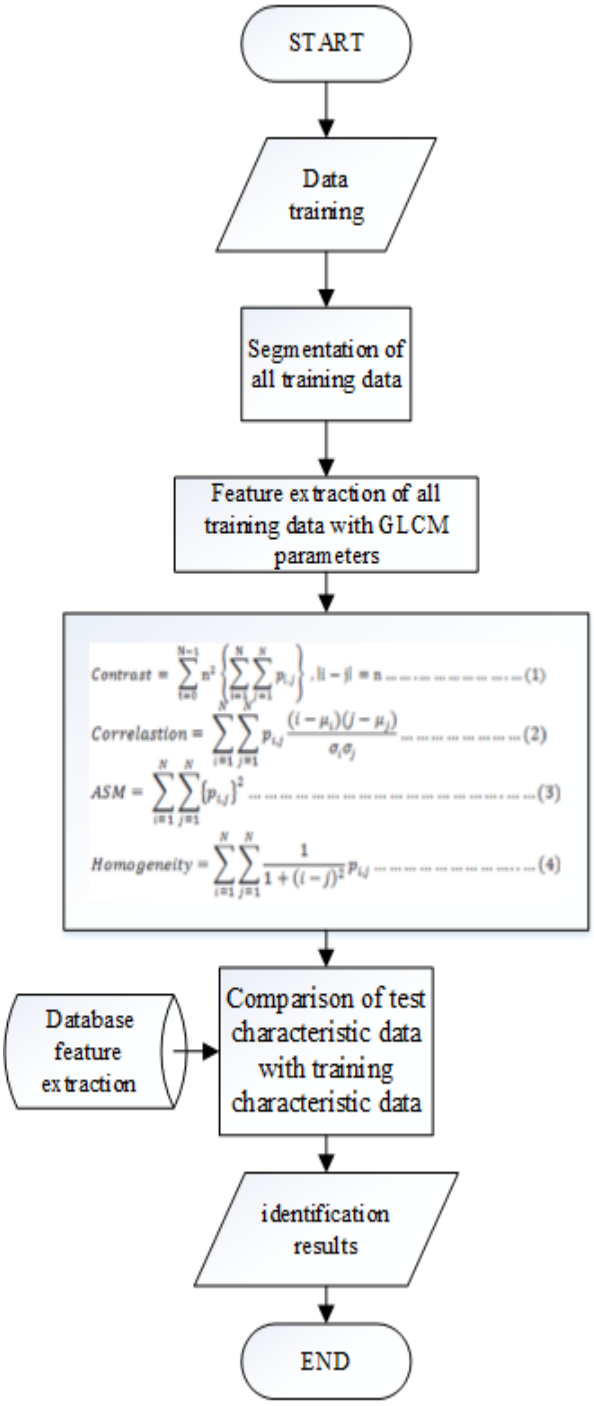


Figure 3. Flowchart identification

## III. Result and Discussion

The first process performed is segmentation. This process of changing the image of oil palm fruits RGB already taken and become the input image for testing data in this study and the image of oil palm fruits RGB segmentation aims to separate between the background and complete picture of the image of the fruit. Then the image will also be extracted using a texture characteristic of GLCM.

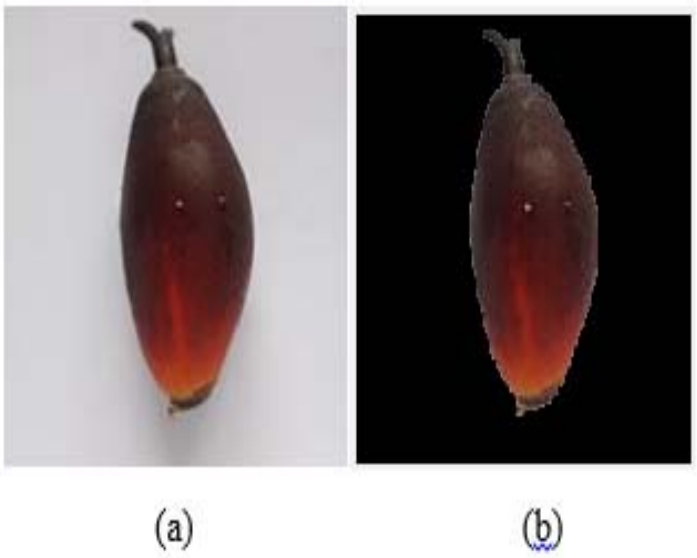


Figure 4. The result of the segmentation process (a) original image, (b) image segmentation K-Means.

Table 3.1 Table database traits

| No | Contrast | Correlation | Energy | Homogeneity |
|---|---|---|---|---|
| 1 | 0,14008 | 0,87280 | 0,72063 | 0,96641 |
| 2 | 0,22420 | 0,84543 | 0,70374 | 0,97058 |
| 3 | 0,13509 | 0,83309 | 0,75817 | 0,97352 |
| 4 | 0,28188 | 0,87603 | 0,67135 | 0,96396 |
| 5 | 0,07888 | 0,86720 | 0,69311 | 0,98346 |
| 6 | 0,09307 | 0,87485 | 0,63253 | 0,97873 |
| 7 | 0,07453 | 0,87592 | 0,71239 | 0,98369 |
| 8 | 0,08203 | 0,88830 | 0,69743 | 0,98271 |

Table 3.2. The identification results

| No. | ripeness | Number of test data | The percentage of success | |
|---|---|---|---|---|
| | | | Right | Wrong |
| 1 | Mature | 25 | 25 | 0 |
| 2 | Not cooked | 25 | 20 | 5 |
| Total Data Test | | 50 | 45 | 5 |
| Accuracy of success | | $Accuracy = \frac{50-5}{50} 100\% = 90\%$ | | |

## IV. COLCLUSION

Feature extraction using cementation k-mean clustering and GLCM can be used as a basis to develop a system to identify successfully developed and can properly classify the maturity of the oil palm fruit based image of oil palm fruits Based on the test results it can be concluded that the level of accuracy in the room with lighting quite reach the figure of 90%.

## REFERENCES


[1] M. Widyaningsih, "IDENTIFIKASI KEMATA-NGAN BUAH APEL DENGAN GRAY LEVEL CO - OCCURRENCE MATRIX ( GLCM )," *SAINTEKOM*, vol. 6, no. 114, pp. 71–88, 2016.

[2] A. Premana, A. P. Wijaya, and M. A. Soeleman, "Image Segmentation Using Gabor Filter and K-Means Clustering Method," *iSemantic 2017-2017 IEEE Int. Semin. Appl. Technol. Inf. Commun.*, pp. 95–99, 2017.

[3] D. M. Agaputra, K. R. R. Wardani, and E. Siswanto, "Pencarian Citra Digital Berbasiskan Konten dengan Ekstraksi Fitur HSV, ACD, dan GLCM," *Telematika*, vol. 8, no. 2, pp. 8–13, 2013.

[4] L. Arsy *et al.*, "Aplikasi Pengolahan Citra Digital Meat Detection Dengan Metode Segmentasi K-Mean Clustering Berbasis OpenCV Dan Eclipse," *J. Teknol. dan Sist. Komput.*, vol. 4, no. 2, pp. 322–332, 2016.

[5] M. H. Purnomo, J. T. Elektro, U. Sultan, and A. Tirtayasa, "Klasifikasi Jenis dan Fase Parasit Malaria Plasmodium Falciparum dan Plasmodium Vivax Dalam Sel Darah Merah Menggunakan Support Vector Machine," vol. 1, no. 2, pp. 1–8, 2012.

[6] M. Metode, A. Tekstur, K. Dan, and M. Sofie, "KLASIFIKASI CITRA REKAMAN SINYAL ELEKTROKARDIOGRAM," vol. 7, no. 1, pp. 233–240, 2016.

[7] F. Neneng, P. W. Shabri, and F. Wahyuni, "IDENTIFIKASI KARAKTERISTIK BUAH KELAPA SAWIT SIAP PANEN DENGAN METODE LASER SPEKEL IMAGING (LSI)," *Photon*, vol. 9, no. 1, pp. 139–142, 2018.

[8] S. Hadianti and D. Riana, "Segmentasi Citra Bemisia Tabaci Menggunakan Metode K-Means," *Senimar Nas. Inov. dan Tren*, vol. 1, pp. 2–7, 2018.

[9] J. Banda, A. Medan, and B. Rata, "Identifikasi tingkat kebulatan buah pepaya berdasarkan luas objek dengan pengolahan citra," *InfoMedia*, vol. 2, no. 2, pp. 41–47, 2017.

[10] N. Sabri, Z. Ibrahim, S. Syahlan, N. Jamil, and N. N. A. Mangshor, "PALM OIL FRESH FRUIT BUNCH RIPENESS GRADING IDENTIFICATION USING COLOR FEATURES," *J. Fundam. Appliend Sci.*, vol. 9, pp. 563–579, 2017.

[11] Sukemi, H. Sudibyo, and A. A. P. Ratna, "Priority based computation: A study on paradigm shift on real time computation," *Proceeding - 2012 IEEE Int. Conf. Comput. Intell. Cybern. Cybern. 2012*, pp. 129–132, 2012.

[12] R. Candra, N. Santi, S. Pd, and M. Kom, "Mengubah Citra Berwarna Menjadi Gray - Scale dan Citra biner," *Teknol. Inf. Din.*, vol. 16, no. 1, pp. 14–19, 2011.

[13] Andri, Paulus, N. P. Wong, and T. Gunawan, "SEGMENTASI BUAH MENGGUNAKAN METODE K-MEANS CLUSTERING DAN IDENTIFIKASI KEMATANGANNYA MENGGUNAKAN METODE PERBANDINGAN KADAR WARNA," *SIFO Mikroskil*, vol. 15, no. 2, pp. 91–100, 2014.

[14] F. Syuhada, I. G. P. S. Wijaya, and F. Bimantoro, "Pengenalan Wajah Untuk Sistem Kehadiran Menggunakan Metode Eigenface dan Euclidean Distance," *J-COSINE*, vol. 2, no. 1, pp. 64–69, 2018.

[15] W. K. Pratt, *PROCESSING DIGITAL IMAGE PROCESSING*, vol. 5. 2001.

[16] D. Rohpandi, A. Sugiharto, and M. Y. S. Jati, "Klasifikasi Citra Digital Berbasis Ekstraksi Ciri Berdasarkan Tekstur Menggunakan GLCM Dengan Algoritma K-Nearest Neighbor," *Voice of Informatics*, vol. 7, no. 2, pp. 79–86, 2018.

[17] F. Shofrotun, T. Sutojo, D. R. Ignatius, and M. Setiadi, "Identifikasi Tumbuhan Obat Herbal Berdasarkan Citra Daun Menggunakan Algoritma Gray Level Co-occurence Matrix dan K-Nearest Neighbor," *J. Teknol. dan Sist. Komput.*, vol. 6, no. 2, pp. 51–56, 2018.

[18] E. Nurraharjo and B. D. Santoso, "Teknik pengambilan warna terpilih pada citra motif batik dengan metode color space adaptif," *Proseeding SENDI*, vol. 1, no. 1, pp. 978–979, 2018.