

CLASSIFICATION OF MANGO FRUIT QUALITY BASED ON TEXTURE CHARACTERISTICS OF GLCM (GRAY LEVEL CO-OCCURRENCE MATRICES) WITH ALGORITHM K-NN (K-NEAREST NEIGHBORS)

Wahyu Wijaya Widiyanto¹, Eko Purwanto², Kusri³

¹²³Department of Informatics, University AMIKOM Yogyakarta, Jl. Ring Road Utara, CondongCatur, Sleman, Yogyakarta, Central Java, 55283, Indonesia

Informasi Makalah	INTISARI
Dikirim, 7 Januari 2019 Revisi, 11 Januari Diterima, 30 April 2019 <hr/> Kata-kata kunci : Klasifikasi GLCM K-Nearest Neighbour Mangga	Proses klasifikasi kualitas mutu buah mangga dengan cara konvensional menggunakan mata manusia memiliki kelemahan di antaranya membutuhkan tenaga lebih banyak untuk memilah, anggapan mutu kualitas buah mangga antar manusia yang berbeda, tingkat konsistensi manusia dalam menilai kualitas mutu buah mangga yang tidak menjamin valid karena manusia dapat mengalami kelelahan. Penelitian ini bertujuan untuk klasifikasi kualitas mutu buah mangga ke dalam tiga kelas mutu yaitu kelas Super, A, dan B dengan computer vision dan algoritma k-Nearest Neighbor. Hasil pengujian menggunakan jumlah k tetangga 9 menunjukkan tingkat akurasi sebesar 88,88%.
<hr/> Keywords <i>Classification</i> <i>GLCM</i> <i>K-Nearest Neighbor</i> <i>Mango</i>	<hr/> ABSTRACT The classification process of the quality of mangoes in the conventional way using the human eye has disadvantages including requiring more energy to sort, the assumption of the relative quality of different human mangoes, the level of human consistency in assessing the quality of mangoes that do not guarantee validity because humans can experience fatigue. This study aims to classify the qualities of mangoes into three quality classes, namely the Super, A, and B classes with computer vision and the k-Nearest Neighbor algorithm. The test results using the number of k neighbors 9 shows an accuracy rate of 88.88%.

Korespondensi Penulis:

Wahyu Wijaya Widiyanto¹

Department of Informatics, University AMIKOM Yogyakarta,
 Jl. Ring Road Utara, CondongCatur, Sleman, Yogyakarta,
 Central Java, 55283, Indonesia

1 INTRODUCTION

Mango is one of the fruits that are popular in Indonesia and the mangoes are a type of fruit that has many varieties because it is currently cultivated. Some of the superior varieties of mango fruit include Arumanis, Gedong, Golek, and Manalagi Mangoes (Anonim1, 2015). In addition to having a sweet and refreshing taste, it turns out the mango fruit contains many nutrients that are beneficial to the health of the body. This fruit is also rich in vitamin E which makes the body more healthy and prevents various diseases. Sometimes mango farmers still use manual methods to distinguish the quality of the mango's maturity, while the methods carried out by human labor are often inaccurate and different in their determination. This difference is caused by different perceptions in everyone. Mango is included in tropical fruit, the maturity condition of tropical fruit will be very visible from the color, whether the fruit is still raw, half cooked, ripe or

already rotten. Therefore color feature extraction from tropical fruit will be used to determine the maturity level of the fruit for industrial purposes (Noviyanto, 2009).

The need for mangoes for export purposes requires handling quality so that they can compete with mango producers from other countries. One of the problems in handling quality is postharvest problems. Postharvest includes cleaning, sorting, storing, packaging and marketing. The sorting process is one of the processes that greatly determines the quality of the mangoes to be released to consumers. So far the quality identification in the mango sorting process is still done conventionally using human eyes that have limitations. The process of identifying quality in this way has several disadvantages including requiring more energy to sort, the level of human perception in terms of different fruit bleaching, the level of human consistency in assessing fruit quality does not guarantee because humans can experience fatigue (Syaefullah et al. 2011).

Texture Characteristics Analysis

Texture can be defined as a mutual relationship between the value of the intensity of repetitive neighboring pixels in an area wider than the distance of the relationship (Kadir and Susanto, 2013). In general, applications with texture characteristics can be divided into two categories, namely applications for the sake of segmentation and applications for texture classification, which use texture features for object classification. The method used to obtain texture characteristics can be divided into three groups, namely statistical methods, structural methods, and spectral methods. According to Kadir and Susanto (2013) Statistical methods use statistical calculations to form traits. An example that is included as a statistical method is GLCM (Gray Level Counseling Matrices). The structural method describes the arrangement of de elements in texture, an example of a structural method is shape grammar. The spectral method is a method based on the spatial-frequency domain, an example of a spectral method is the flourier domain energy distribution.

Ahmad's (2002) study states that texture analysis is quite useful for detecting the presence of defects in mango skin. Texture analysis can be used in conjunction with examining the size of the mango through the area of the projection on the flat plane, to do the screening of Arumanis mangoes based on their size and the presence of defects on their skin. There are several texture traits including energy and entropy descriptors. An energy descriptor is a measure that states the intensity distribution of pixels against the range of gray levels. Entropy indicates the complexity of the image, the higher the entropy value in an image, the more the image complex will be. The definition of energy and entropy as shown in equation (1) and equation (2) below:

$$energy = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} [p(i, j)]^2 \quad (1)$$

$$entropi = - \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p(i, j) \log_2(p(i, j)) \quad (2)$$

Other characteristics of texture include contrast, homogeneity, inverse difference moment (idm), variance, and dissimilarity shown in equations (3) to (7) below:

$$contrast = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - j)^2 (p(i, j)) \quad (3)$$

$$homogeneity = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{P(i, j)}{1 + |i - j|} \quad (4)$$

$$idm = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{1}{1 + (i - j)^2} p(i, j) \quad (5)$$

$$variance = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - \mu)^2 (p(i, j)) \quad (6)$$

$$dissimilarity = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} |i - j| p(i, j) \quad (7)$$

Texture characteristics to be sought are based on the intensity matrix gray level coefficient of matrix (GLCM). The cooccurrence matrix is defined by two steps, the first is to determine the distance between points in a vertical and horizontal direction, the second step is to calculate pairs of pixels that have the same intensity value according to the distance between predetermined

points. The results of the calculation of each pair of intensity values are arranged on the matrix according to the coordinates (Ahmad, 2005).

2 METHODS

Research Flow

This study focuses on digital image processing and the K-Nearest Neighbor algorithm. The input data is in the form of mango image data which amounts to 192 image data.

The flow of research carried out in this study are as follows:

1. Collect data in the form of 192 image mango image data.
2. The process flow of the quality class quality classification system which includes several processes, namely the image acquisition process, then the digital image processing which includes preprocessing, segmentation, and feature extraction processes, then the characteristic data is stored in the database. Characteristic data will be used as input data during the classification process of quality class quality using the K-Nearest Neighbor algorithm. The design process flow of the classification system is shown in Figure 1:

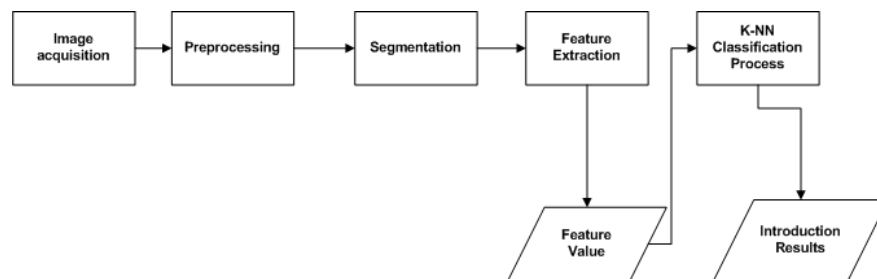


Figure 1. Research Flow

The image acquisition process is carried out on the mangoes that have been picked from the tree. The acquisition process is carried out on the mangoes that have been classified according to the quality class, namely the Super class, class A, and class B. The quality classification process is carried out by mango farmers who are experts in assessing the quality of mangoes.

1. The preprocessing process is carried out with the aim of increasing the likelihood of success at the stage of further processing of an image. Operations performed at the preprocessing stage include resizing and gray scaling operations. Resizing is the process of changing resolution or changing the horizontal and vertical size of an image. The resizing operation is done is to change the size of the image from its original size, which is 3264 x 1836 pixels to 1024 x 575 pixels with the aim of speeding up the processing of digital images, because the larger the size of the image it will affect the time needed for digital image processing.
2. Image segmentation is a process that is intended to get objects contained in the image or divide the image into several regions with each object or region having similar attributes. In objects that contain only one object, the object is distinguished or separated from its background, so that what is to be processed in the next stage is the part of the image that includes the object image only. Segmentation by selecting the threshold value automatically can be done using the Otsu method. This method determines the threshold value by distinguishing two groups, namely objects and backgrounds, which have parts overlapping, based on histograms (Kadir and Susanto, 2013).
3. GLCM texture feature extraction to be searched uses distance (1.0), which means one pixel to the right and zero pixels to the bottom, or pixel pairs that must be calculated are those that are one pixel apart from the angle of 00, as shown in Figure 2:

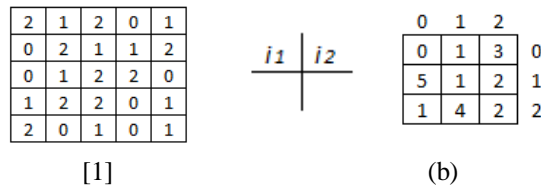


Figure 2 (a) Image measuring 5x5 three intensities (0, 1, 2); (b) GLCM distance (1,0)

In Figure 5a, the image is 5x5 in size and has three intensity values, namely 0, 1 and 2, then the GLCM will be 3x3. The distance between the specified pixels is $d = (1, 0)$, which means one pixel to the right and zero pixels down. Then count the number of pairs of pixels, where the first pixel has an intensity value of i_1 and the pair spaced d has an intensity value of i_2 , and is entered into column i_1 and row $k-i_2$ on GLCM. For example, in the first pixel pair which has an intensity value (2, 1), there are 2 pairs so that the value 2 is entered into column 2 row 1 on the GLCM. Figure 5b. shows the GLCM resulting from calculating the entire pixel pair. After the GLCM matrix is formed, texture characteristics can be calculated based on the matrix.

4. The process of estimating or classifying the quality grade of mangoes using the k-Nearest Neighbor algorithm will be carried out several times using a variable k value to determine the best value of accuracy.

K-Nearest Neighbour(K-NN)

KNN is a classification method that determines categories based on the majority of categories in the nearest k -neighbor. If D is a set of training data then when the test data d is presented, the algorithm will calculate the distance between each data in D with test data d . distance calculation is done using the Euclidian distance. Then k the data in D which has the closest distance to d is taken. The set k is the k -nearest neighbor. Furthermore the category of test data d is determined based on the label of the majority of categories in the closest set of k -neighbors (Farsiah, et al, 2013).

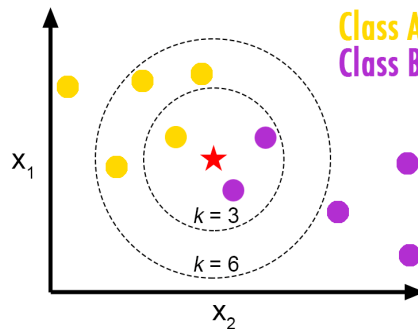


Figure 3. KNN where $k = 3$ and $k = 6$

Or simply the steps of the KNN algorithm are as follows:

1. Determine k .
2. Calculate the distance between the new data to each labeled data.
3. Determine k labeled data that has the least distance.
4. Classify the new data into majority labeled data.

Calculation of distance can be searched using the euclidian distance shown in equation (8):

$$Di = \sqrt{\sum_{i=1}^k (x_1 - y_1)^2} \tag{8}$$

3 SEARCH RESULT

All quality class classification system processes that will be built include several processes, namely beginning with the image acquisition process, digital image processing that is equipped with pre-processing, segmentation, and feature extraction processes, then the data displayed is in accordance with the database. Characteristic data will be used as input data during the quality class classification process using the K-Nearest Neighbour algorithm. Image taking is done on the mangoes that have been picked from the tree. The acquisition process is carried out on mangoes that have been done based on the quality class in the Super class, class A, and class B. The process of quality classification is carried out by mango farmers who are experts in assessing the quality of mangoes.

A. Pre-processing

The preprocessing process is carried out with the aim of increasing the likelihood of success at the stage of further processing of an image. Operations performed at the preprocessing stage include resizing and grayscale operations. Resizing is the process of changing resolution or changing the horizontal and vertical size of an image. The resizing operation is done is to change the size of the image from its original size, which is 3264 x 1836 pixels to 1024 x 575 pixels with the aim of speeding up during the process of digital image processing. The preprocessing stage is shown in Figure 4. below:

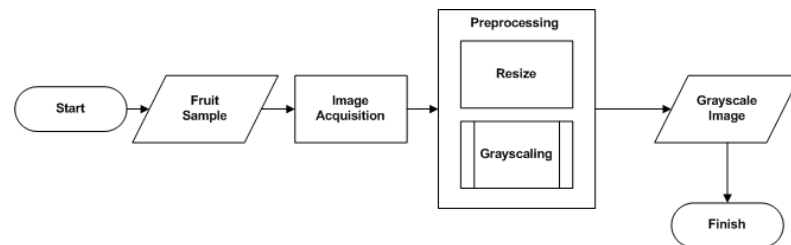


Figure 4. Image Processing

B. Segmentation

Image segmentation is a process that is intended to get objects contained in the image or divide the image into several regions with each object or region having similar attributes. In objects that contain only one object, the object is distinguished or separated from its background, so that what is to be processed in the next stage is the part of the image that includes the object image only. Segmentation by selecting the threshold value can automatically be done by the Otsu method. This method determines the threshold value by distinguishing two groups, namely objects and backgrounds, which have parts overlapping, based on histograms (Kadir and Susanto, 2013). The principle of the Otsu method to obtain a threshold value is shown by the flow diagram in Figure 5 below:

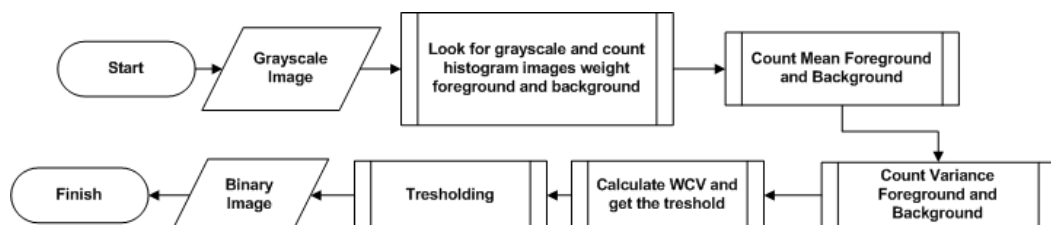


Figure 5. Segmentation with The Otsu Method

C. Extraction of GLCM texture features

GLCM texture feature extraction to be searched uses distance (1.0), which means one pixel to the right and zero pixel to the bottom, or pixel pairs that must be calculated are those that are one pixel apart from the angle of 00. The process of getting the GLCM is shown in Figure 6 below:

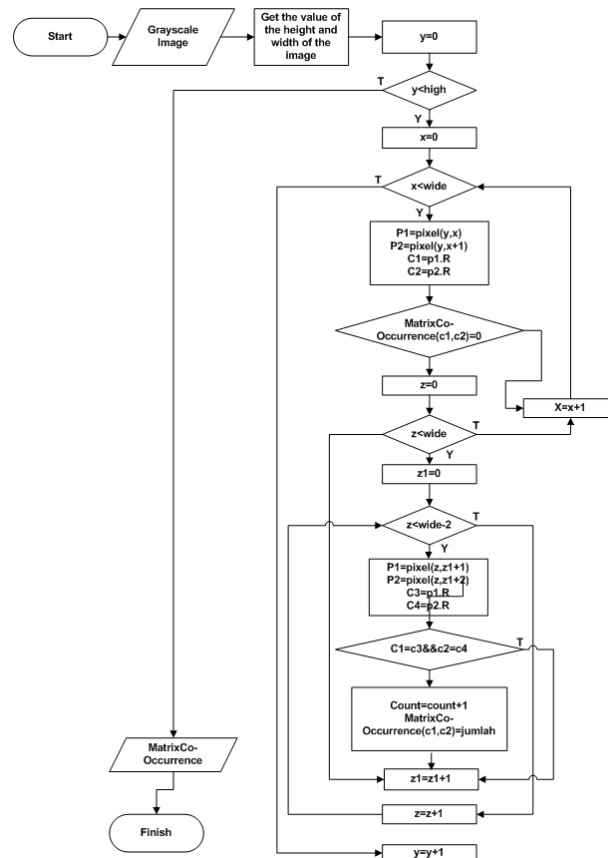


Figure 6. GLCM Process

After obtaining GLCM, the next step is to look for energy features, entropy, contrast, homogeneity, Idm, variance, and dissimilarity using equation (1) to equation (7). The process of obtaining the texture features obtained from the GLCM is shown in Figure 7 below:

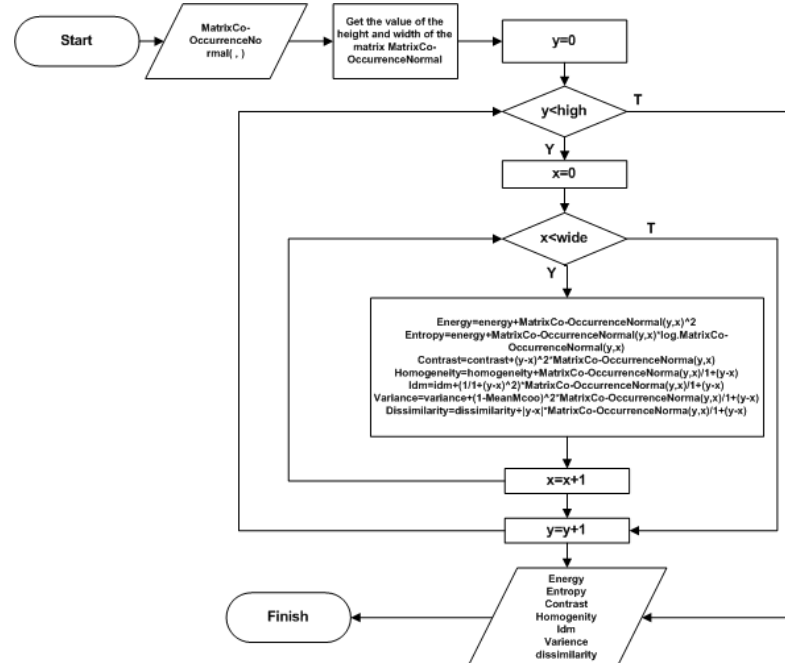


Figure 7. The Process Of Getting A Texture Feature From GLCM

D. K - Nearest Neighbor

The classification process of the mango fruit quality class using the k-NearestNeighbour algorithm will be conducted several times using a variable k value to determine the best value of accuracy. The classification process with k-Nearest Neighbour is shown in Figure 9 below:

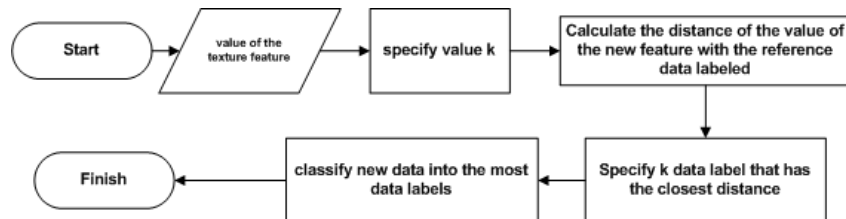


Figure 8. Process Classification With K – Nearest Neighbor

E. Results Of Implementation Of The Prototype Interface For Users

User interface prototype is the stage of making prototypes in the display section of the prototype, or in this case is the creation of a user interface that functions as a page to communicate between users and systems. Figure 9 shows the implementation of the prototype main interface page:

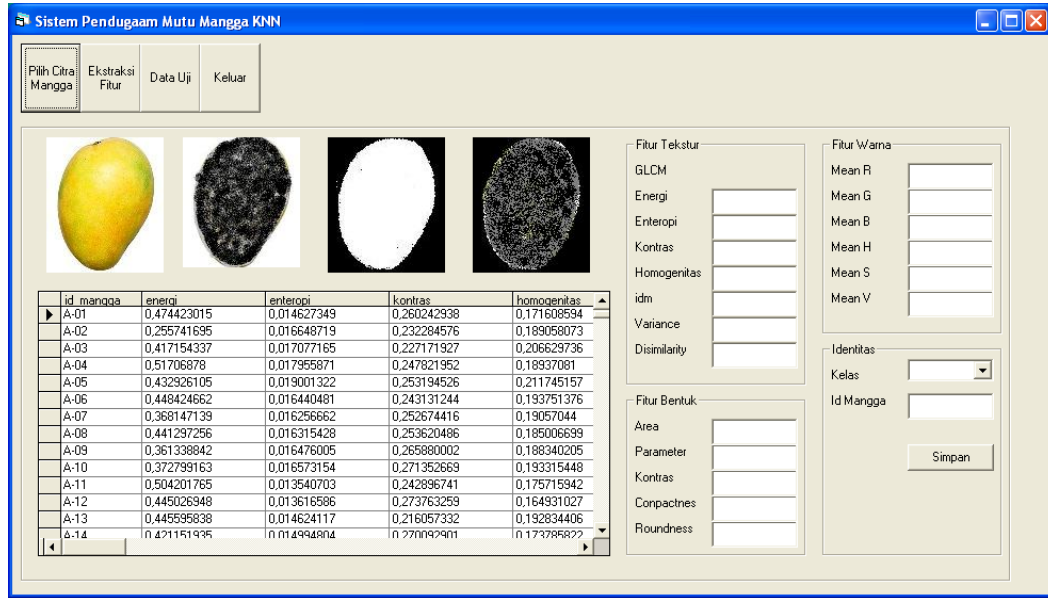


Figure 9. User Interface Prototype

The interface page of the testing process serves to display the testing process, from the test results it will be known how many percentages of success in the classification, it is known how many and in which mango quality class there is a classification failure. The interface page of the testing process is shown in Figure 10 below:

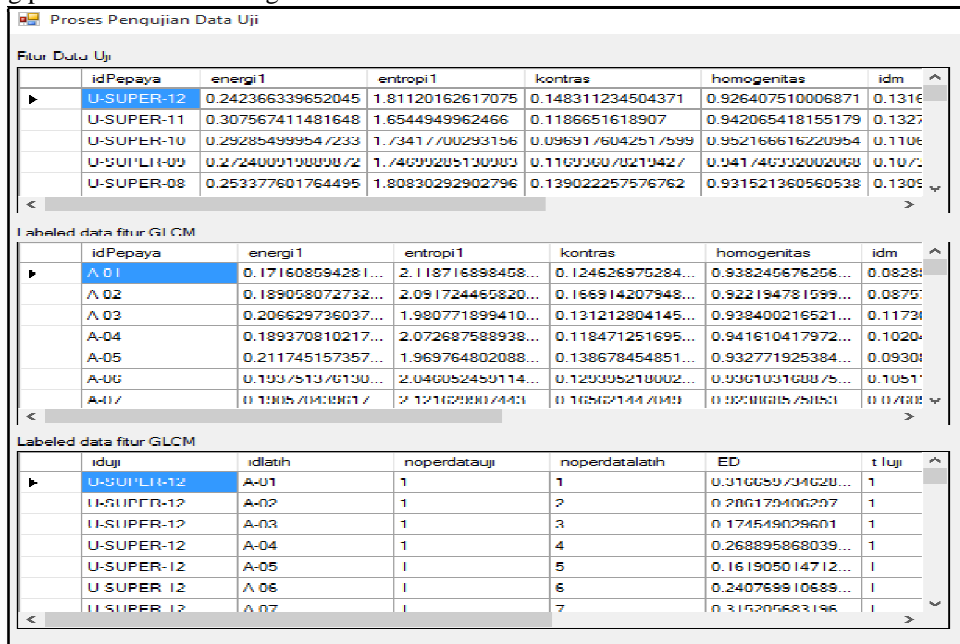


Figure 10 Testing Interface Page

F. Feature Extraction Results

The feature extraction process will produce GLCM texture feature values including energy features, entropy, contrast, homogeneity, idm, variance, dissimilarity. Table 1 shows the average value of GLCM texture feature extraction for each class of mango quality obtained from the feature extraction process through prototype experiments.

Table 1. The Average Value Of GLCM Texture Feature Extraction Results

No.	Feature	SUPER	A	B
1	<i>Energy</i>	0.282	0.196	0.148
2	<i>Entropy</i>	0.169	0.201	0.226
3	<i>Contras</i>	0.127	0.130	0.154
4	<i>Homogeneity</i>	0.937	0.936	0.925
5	<i>Idm</i>	0.132	0.113	0.081
6	<i>Variance</i>	0.154	0.133	0.124
7	<i>Dissimilarity</i>	0.125	0.128	0.150

G. Test Results

Testing the quality classification system for mangoes is based on test data. The classification process of the mango fruit quality class using the k-Nearest Neighbour algorithm will be conducted several times using the k value varies for know the value of the best accuracy. The test data is 36 data consisting of 12 test data for each quality class.

The accuracy of the classification system of the mango quality class is calculated by equation (9).

$$\text{accuracy}(\%) = \frac{\text{the number of images of mango is recognized according to class}}{\text{total number of mango images}} \times 100 \quad (9)$$

Table 2. Test Results For Test Data

No	Number of neighbor k	Known by Class	Not According to Class	Recognized (%)
1	3	31	5	86,11%
2	5	31	5	86,11%
3	7	31	5	86,11%
4	9	32	4	88,88%
5	11	31	5	86,11%
6	13	31	5	86,11%

In Table 2 it can be seen the number of images recognized according to the class. It can be concluded that by using the number of k neighbors 9 produces the best accuracy value, namely 32 test image data identified and 4 test image data not recognized according to the quality class or an accuracy rate of 88.88%.

4 CONCLUSION

Based on the results of the research classification of the mango quality class based on the characteristics of the GLCM texture using the k-Nearest Neighbour algorithm can be concluded as follows: The classification process of the mango quality class was carried out using a prototype that had been successfully designed and made using digital image processing technology for extracting the GLCM texture features of mango images and the K-Nearest Neighbor algorithm for its classification method. homogeneity, idm, variance and dissimilarity, GLCM texture features and k-Nearest Neighbor algorithm can be used to classify the quality of mangoes. The test results in Table 2 show the best accuracy value is the number of k Neighbor 9 with an accuracy value of 88.88%.

Suggestion for the sake of further research is the classification process using other methods or algorithms in the hope that the results of the classification obtained will be better and more accurate, Trying to extract other features from form features, texture features or color features, in the hope of getting discriminatory features the high for each quality class and the time in the required image feature extraction process is faster.

REFERENCE

- [1] Adnan, 2011, Karakteristik sifat jeruk manis berdasarkan tingkat ketuaan, *Prosiding seminar teknologi inovatif pascapanen pertanian ISBN: 978-979-116-32-9*, Bogor
- [2] Ahmad, U., Tjahjohutomo, R., & Mardison, 2008, Perancangan dan Konstruksi Mesin Sortasi dan Pemutuan Buah Jeruk dengan Sensor kamera CCD, *Junal Keteknikan Pertanian (JTEP)*, ISSN 0216-3365, Bogor.
- [3] Ahmad, U., 2005, *Pengolahan Citra Digital & Teknik Pemrogramannya*, Graha Ilmu, Yogyakarta.
- [4] Ahmad, U., 2002, Pengolahan Citra untuk Pemeriksaan Mutu Buah Mangga, *Buletin Keteknikan Pertanian*, Fakultas Teknologi Pertanian IPB, Bogor
- [5] Arifin, A.D., Arieshanti, I., & Arifin, A.Z., 2012, Implementasi algoritma k-nearest neighbor yang berdasarkan one pass clustering untuk kategorisasi teks, ITS, Surabaya.
- [6] BLST (Bogor life Science and Technology), 2015, Pepaya Calina IPB 9, <http://blst.co.id/pepaya-calina-ipb-9>, diakses 17 November 2016.
- [7] Dinar, L., Suyantohadi, A., dan Fallah, M.A.F., 2012. Pendugaan Kelas Mutu Berdasarkan Analisa Warna dan Bentuk Biji Pala (*Myristica Fragnans Houtt*) Menggunakan Teknologi Pengolahan Citra Digital dan Jaringan Syaraf Tiruan, *Junal Keteknikan Pertanian (JTEP) Vol. 26. No.1 April 2012*, ISSN 0216-3365, Bogor.
- [8] Economic and Social Development Department, FAO., 2010, Medium-term prospects for agricultural commodities (Tropical Fruits), <http://www.fao.org/docrep/006/y5143e/y5143e1a.htm>, diakses 28 April 2016.
- [9] Farsiah, L., Abidin, T.F., & Munadi, K., 2013, Klasifikasi gambar berwarna menggunakan k-nearest neighbor dan support vector machine, *SNASTIKOM*, Banda Aceh.
- [10] Kadir, A. & Susanto, A. 2013, *Teori dan Aplikasi Pengolahan citra*, Penerbit Andi, Yogyakarta.
- [11] Kementrian Pertanian RI., 2014, Basis Data Statistik Pertanian, <http://aplikasi.pertanian.go.id/bdsp/newkom.asp>, diakses 27 April 2016.
- [12] Sugiyanto, S., & Wibowo, F., 2015, Klasifikasi Tingkat Kematangan Buah Pepaya (*Carica Papaya L*) California (*Callina-Ipb 9*) Dalam Ruang Warna Hsv dan Algoritma K-Nearest Neighbors, *Seminar Nasional Hasil-Hasil Penelitian dan Pengabdian LPPM Universitas Muhammadiyah Purwokerto*.
- [13] Syaefullah, E., Purwadaria, H.K., & Sutrisno, 2011, Pengolahan Citra Digital dan Jaringan Saraf Tirua untuk Identifikasi Tingkat Ketuaan Pepaya, *Prosiding Seminar Nasional Teknologi Inovatif Pascapanen Pertanian III*, ISBN: 978-979-1116-32-9, Bogor.