# *Implementation of First Order and Second Order Image Feature To Perform Gemstone Classification*

Riyan Hidayat[1]

[1]Teknik Informatika – Universitas Komputer Indonesia
Jl. Dipatiukur 112-114 Bandung
Email : ian_gokilz92@ymail.com[1]

## ABSTRACT

*Gemstone is a stone or mineral that occurs naturally from the results of the procedure geological elements consist of one or several chemical components that have a selling price is very high. Business agate is being rampant in Indonesia, agate many classified wrong, alone - the eye just for the sake of the seller, out there a lot of unscrupulous sellers tried to deceive ordinary people just for the sake of personal interests, lest they do not understand even fooled by agate abal abal valuable.*

*One technique to obtain an image of the pattern is to do a feature extraction. As had been widely grown applications - applications that use the characteristic features of color and texture. A characteristic color and texture itself can be done in various ways, one of them carried out by the method of order 1 and characteristic traits of order 2, and classification using K-Nearest Neighbor algorithm.*

*Based on the test results can be concluded that the implementation of merging methods characteristic features first-order and second-order characteristic features and classification using the K-Nearest Neighbor produces an accuracy of 67.5% using K = 20.*

*Keywords: Gemstone, Feature extraction, Order 1 and Order 2, classification, K-Nearest Neighbor*

## 1. *INTRODUCTION*

*Gemstone is a stone or mineral that occurs naturally from the results of the procedure geological elements consist of one or several chemical components that have a selling price is very high. Agate made from mineral substances crystallize in a very long period of time, can reach millions of years, then this crystal changed into agate [1]. Agate, also known as stone ring or a precious stone is now becoming a trend and was so popular in the community. The beauty and uniqueness as if it has become a magnet that has such great allure to anyone who saw it. Stone ring does have levels or certain classes, such as diamond, emerald, ruby or sapphire is a precious stone that is categorized as a gemstone upscale course can only be found in specialty sellers. As for the types of agate like kind sulaiman, cempaka honey, thorns month, pandan, wine, badar iron and others can be right at the various stalls vending stone ring [2].*

*Business agate is being rampant in Indonesia, almost every front of the alley there is selling agate. Many sellers of agate are popping up because there is still high demand for this precious stone. But as the buyer must be keen in buying genuine agate. Do not let the lack of understanding even fooled with agate Abal Abal valuable. Out there are a lot of unscrupulous sellers tried to deceive ordinary people just for the sake of personal interest only. This fraud may also occur because the seller himself was less understand the ins and outs of agate [3].*

*One technique to obtain an image of the pattern is to do a feature extraction. Feature extraction is based on the visual content of the image of the colors, shapes and textures. Previous techniques for extraction of these characteristics has been widely applied one example is the detection of a skin disease.*

*As had been widely grown applications - applications that use the characteristic features of color and texture. A characteristic color and texture itself can be done in various ways, one of them carried out by the method of order 1 and characteristic traits of order 2, and classification using K-Nearest Neighbor algorithm. Based on research conducted by using a characteristic statistical entitled "Application of image processing to identify the maturity of a cucumber by the texture of the fruit skin using feature extraction statistics" which has the accuracy of 75% of the order of 1 [4], whereas in the order of 2 contained in the study entitled "Classification of oblique fibers in wood using statistical feature extraction based on image processing" which has the accuracy of 92.5791% [5].*

*Of the problems and the solutions that have been described, then this thesis research will classify the types of agate based on the color and texture of the stone by applying the method of order 1 and order 2 and classification algorithm is K-Nearest Neighbor, the expected merger of the two methods and algorithms K-Nearest Neighbor able to classify different types of agate with accuracy. Due to the high degree of accuracy obtained the expected results of*

*this classification can be used as reference material for further support to the experts.*

### 1.1 Problem Formulation

*Based on the above background, the problem can be formulated as follows:*

*How do I apply the methods of extraction characteristics of order 1 and 2 in order to identify or classify different types of agate and analysis of the performance and accuracy of the resulting system as a whole.*

### 1.2 Purpose and Objectives

*Based on the formulation of the problem that has been studied, the purpose of this thesis is to implement a method of extraction characteristics of order 1 and 2 in order to classify agate image based on the extraction of digital imagery.*

*Objectives to be achieved in this research is to ascertain the level of accuracy of the methods of extraction characteristics of order 1 and order 2.*

## 2. BASIC THEORY

### 2.1 Digital Image Processing

*In general, digital image processing refers to the processing of two-dimensional images using a computer. In a broader context, digital image processing refers to the processing of any two-dimensional data. Digital image is an array (array) that contains the values of real or complex which is represented by a specific bit sequence.*

*In the digital image formation color model commonly used are as follows:*

#### 2.1.1 RGB image

*RGB image commonly referred to as true color image, stored in the image size (m x n) x 3 which defines the color red (red), green (green) and blue (blue) for each pixel. Color at each pixel is determined based on the combination of the colors red, green and blue (RGB). A 24-bit RGB image with red components, green, and blue respectively generally worth 8 bits so that the intensity of the brightness of up to 256 level and the color combinations of less than approximately 16 million colors.*

#### 2.1.2 Grayscale

*Citra with different degrees of gray with RGB image, this image is defined by the value of the degree of color. Generally worth 8 bits so that the intensity of the brightness level up to 256 and 256 variants color combinations. The lowest brightness level that is 0 for black and white worth 255. To convert a color image having RGB to the degree of gray can use the formula* $\frac{R+G+B}{3}$ (2.1)

### 2.2 Histogram image

Important information regarding the content of the digital image can be known by making *image histogram. Image histogram is a graph illustrating the spread of pixel intensity values of an image or specific part in the image. From a histogram can be determined relative frequency of occurrence (relative) of intensity on that image. The histogram can also show a lot of things about the brightness (brightness) and kontas (contrast) of an image. Therefore, the histogram is a valuable tool in the image processing job both qualitatively and quantitatively.*

*By equation formula* : $P(fn) = \frac{nk}{n}$ (2.2)

*Where :*
*nk: total value fn emerging*
*n: the total number of pixels*

### 2.3 Texture

*Texture is the properties or characteristics possessed by an area (in the image) is quite large so naturally to these properties can be repeated in the area. Definition of texture in this case is the regularity of certain patterns that are formed from the arrangement of pixels in a digital image [10]. To form a texture of at least two requirements that must be met include:*

1. *Consists of one or more pixels that form the patterns of primitive (the smallest parts). The forms of these primitive patterns can be titk, straight lines, curved lines, size and others which are the basic elements of a form.*
2. *The emergence of primitive patterns repeated at intervals and specific direction so that it can be predicted or discovered characteristic recurrence.*



**Figure 1 Sample Image Texture Of Vistex**

*An image of an interpretation different texture when viewed at a distance and a different point of view, man looking at the texture based on the description that is random, such as smooth, rough, regular, irregular, and other sebgainya. This is an incorrect description and non-quantitative, so it needed a deskrpsi quantitative (mathematical) to facilitate analysis [10].*

### 2.4 Statistics Feature Extraction

*Feature extraction is done to get the pattern of an image to be trained and the image that will be tested. Feature extraction method used is the statistical feature extraction. This method uses a statistical calculation of the degree of gray distribution (histogram) by measuring the level of contrast, granularity, and the roughness of a region of adjacency relationship between pixels in the image. Statistical feature extraction is divided into two first-order statistical feature extraction and second order.*

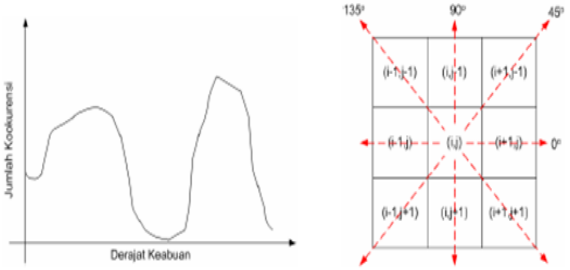



**Figure 2 Illustration statistical feature extraction**

### 2.5 First-order feature extraction

*Feature extraction is a method of collecting a first-order features that are based on characteristics of the image histogram. Histogram shows the probability of occurrence of the value of the degree of gray pixels on an image. Of the values in the histogram is generated, can be calculated several parameters characteristic of the first order, among others, is the mean, variance, skewness, kurtosis, and entropy.*

a. *Mean* (μ)

*Shows the dispersion of the size of an image.*

$$\mu = \sum_{n} fn\, P(fn) \qquad (2.3)$$

*Where fn is a intensity values of gray, while p (fn) indicates the value of the histogram (probability of occurrence of such intensity in the image).*

b. *Variance* ($\sigma^2$)

*Shows the various elements in the histogram of an image.*

$$\sigma^2 = \sum_{n} (fn - \mu)^2 P(fn) \qquad (2.4)$$

c. *Skewness* ($\alpha_3$)

*The relative degree of skewness curve shows a histogram of an image.*

$$\alpha_3 = \frac{1}{\sigma^3} \sum_{n} (fn - \mu)^3 P(fn) \qquad (2.5)$$

d. *Kurtosis* ($\alpha_4$)

*The relative degree of skewness curve shows a histogram of an image.*

$$\alpha^4 = \frac{1}{\sigma^4} \sum_{n} (fn - \mu)^4 P(fn) - 3 \qquad (2.6)$$

e. *Entropy* (H)

Indicates the size of an image of the shape irregularities.

$$H = -\sum_{n} P(fn).^2 \log P(fn) \qquad (2.7)$$

### 2.6. Second-order feature extraction

*One technique to obtain the statistical characteristics of the second order is to calculate the probability of adjacency relationship between two pixels at a certain distance and angular orientation. This approach works by forming a matrix kookurensi of the image data, followed by determining the characteristics as a function of the matrix between the [11].*

### 2.6.1 Methods Co Occurrence Matrix (Gray Level Co Occurrence)

*Methods Co Occurrence Matrix is a matrix whose elements are the number of pairs of pixels that have a certain level of brightness, where the pixel pair separated by a distance d, and with an inclination angle θ. In other words, the matrix is the probability kookurensi graylevel i and j of the two pixels apart at a distance d and angle θ [12].*

*Directions neighboring pixels to represent the distance may be selected, for example,$0^0$, $45^0$ , $90^0$ , $135^0$.*

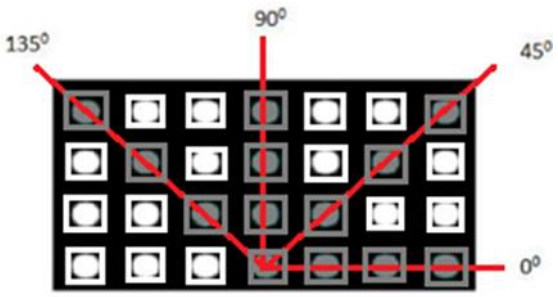



**Figure 3 Directions in calculating GLCM**

*After obtaining the kookurensi matrix, then the next step is to calculate the characteristics of the second-order statistics. Several second-order statistical parameters, among others:*

a. *Angular Second Moment*

*Indicates the size of the image homogeneity properties.*

$$\text{ASM} = \sum_{i} \sum_{j} \{P(i,j)\}^2 \qquad (2.8)$$

*Where p (i, j) the declared value on row i and column j in the matrix kookurensi.*

b. *Contrast*

*Indicates the size of the deployment (moment of inertia) elements of the image matrix. If located far from the main diagonal, the value of great contrast. Visually, the contrast value is a measure of the variation between the degree of gray an image area.*

$$CON = \sum_i \sum_j (i-j)^2 \; P(i,j) \qquad (2.9)$$

c. *Correlation*

*Indicates the size of the linear dependence of the degree of gray image so as to provide an indication of linear structures in the image.*

$$COR = \sum_{i,j} \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma i \; \sigma j} \qquad (2.10)$$

Where :

$\mu_i$ = *value - average line elements of the matrix P (i, j) is expressed by the formula* $\mu_i = \sum_i \sum_j \; i \; p(i,j)$

$\mu_j$ = *value - average column element of the matrix P (i, j) is expressed by the formula* $\mu_j = \sum_i \sum_j \; j \; p(i,j)$

$\sigma_i$ = *standard deviation value element row of the matrix P (i, j) is expressed by the formula*

$\sigma_i = \sqrt{\sum_{i,j}(i-\mu i)^2 p(i,j)}$

$\sigma_j$ = *standard deviation value column element of the matrix P (i, j) is expressed by the formula*

$\sigma_j = \sqrt{\sum_{i,j}(i-\mu j)^2 p(i,j)}$

d. *Variance*

*Shows variations kookurensi matrix elements. Image with a small degree of gray transition will have a small variance anyway.*

$$VAR = \sum_i \sum_j (i-\mu_i)(j-\mu_j)P(i,j) \qquad (2.11)$$

e. *Inverse Different Moment*

*Determining the degree of homogeneity image of a kind of gray. Homogeneous image will have a great price IDM.*

$$IDM = \sum_i \sum_j \frac{1}{1+(i-j)^2} P(i,j) \qquad (2.12)$$

f. *Entropy*

*Showed a measure of disorder form. ENT price is great for the image with uneven degrees of gray transitions and image of little value if the structure is irregular (variable).*

$$ENT^2 = -\sum_i \sum_j P(i,j) \, . \, log \;\; P(i,j) \qquad (2.13)$$

### 2.7 K-Nearest Neighbor

*KNN algorithm is an algorithm that classification is based on the proximity (distance) of the data with other data.*

*In the K-NN algorithm, q-dimensional data, the distance of the data to other data that can be calculated. The value of this distance was used as the value of closeness / similarity between the test data with training data. Rated K on K-means K-NN nearest the data of the test data [13].*

*Euclidian distance is most often used to calculate the distance. Euclidean distance function test that can measure the closeness interpretation dgunakan as the distance between two objects, which are represented as follows:*

$$(x,y) = \sqrt{\sum_{k=1}^{n}(x_k - y_k)^2} \qquad (2.14)$$

Information:

d = distance learning data to the test data.

$x_j$ = test data j, j = 1, 2,. , , n.

$y_j$ = learning data j with j = 1, 2,. , , n.

## 3. RESULTS AND DISCUSSION

### 3.1 Analysis Methods

*The analysis methods used to determine the process flow of a method used can be applied to the applications that are built. Development of this application using order feature extraction feature extraction methods 1 and 2 in order to be useful in order to identify and classify types of agate. The application of the method to the workflow system can be seen in Figure 3.*

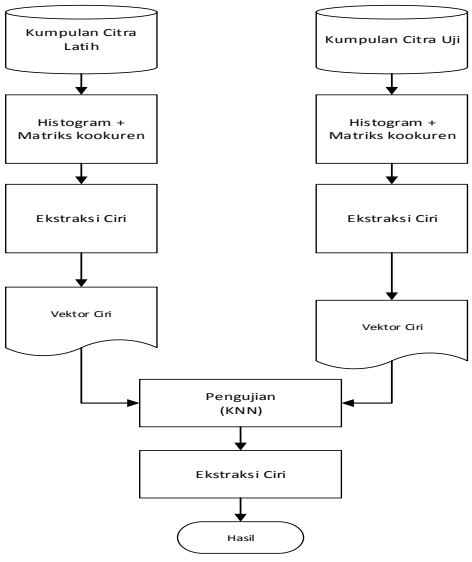



**Figure 4 Process Flow System**

### 3.1.2 Analysis *Pre-Processing Image*

*Prior to extraction characteristics of the image, the image must go through preliminary processing (pre-processing) advance the step sequence is as follows: stage resize, stage resize this image is the process of reducing the size of the digital image, this needs to be done so that all digital images have the same size , Furthermore, changing the image format (conversion) to grayscale.*

*Grayscale phase is necessary for feature extraction process because it has a single pixel value. The process can not be done if the pixel value of more than one like the original value of an image is RGB.*

### 3.1.3. Analysis Feature Extraction

*Extracting characteristic feature is intended to take the features of the image feature vector by using a first-order and second-order begins with segmenting of the image of agate.*

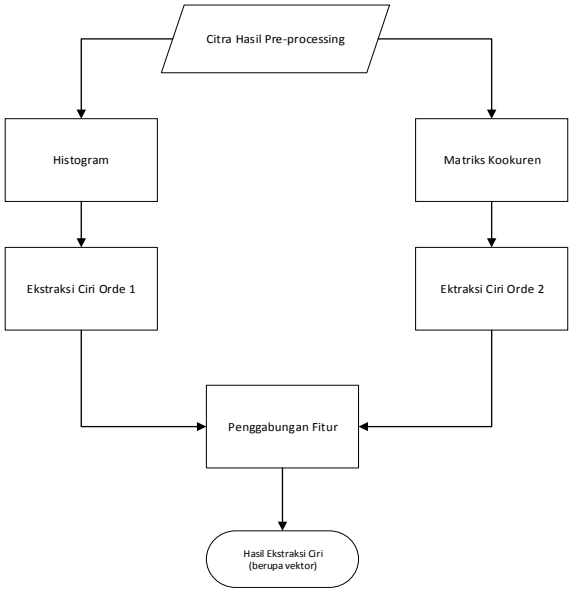



**Figure 5 Feature Extraction Process Flow**

### 3.1.4 Analysis *K-Nearest Neighbor*

*After getting the feature vector is then performed with the classification of K-Nearest Neighbor (K-NN) is to do the classification is based on the proximity (distance) of the data with other data.*

*In the K-NN algorithm, q-dimensional data, the distance of the data to other data that can be calculated. The value of this distance was used as the value of closeness / similarity between the test data with training data. Rated K on K-means K-NN nearest the data from test data.*

### 3.2. System planning

*The design is a representation, planning and making of sketches or arrangement of several separate elements into a unified whole. This stage is to configure the components of software and hardware of a system.*

*Here is an overview of existing interface design of applications to be built, namely:*

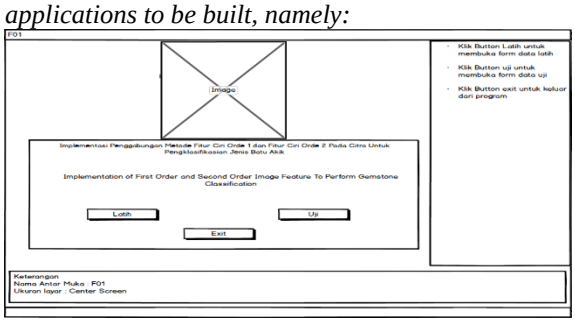



**Figure 6 Interface Main Form**

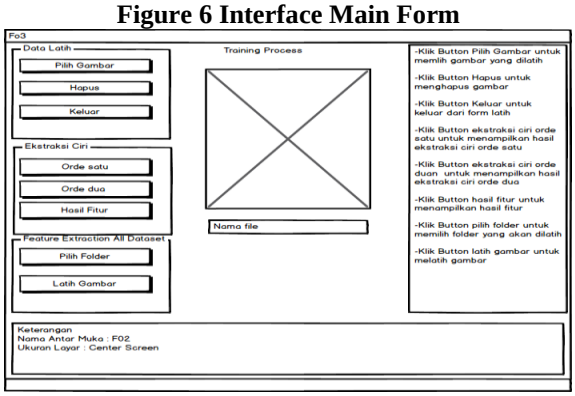



**Figure 7 Interface Form Trains**

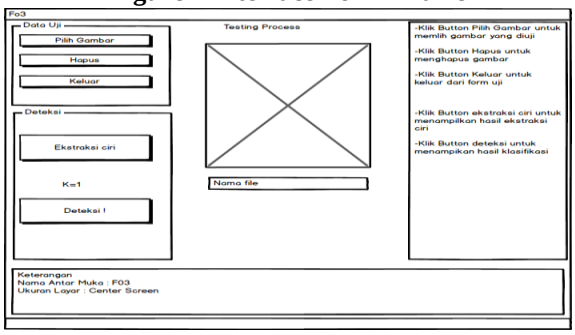



**Figure 8 Interface Form Test**

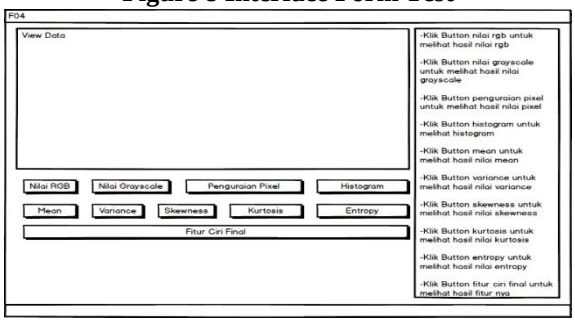



**Figure 9 Interface Order One**

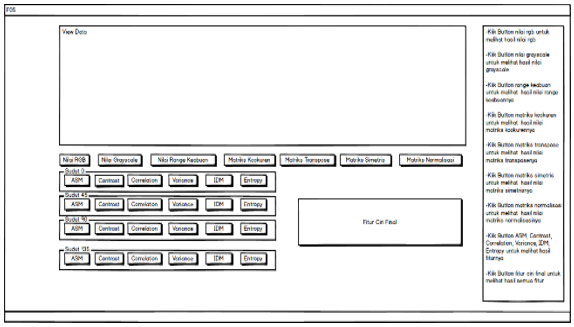



**Figure 10 Interface Order Two**

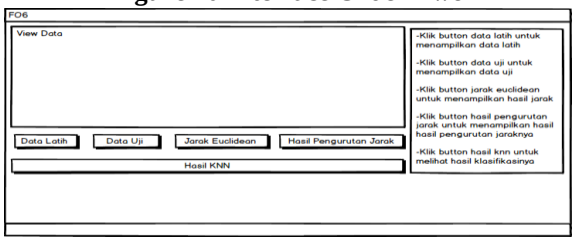



**Figure 11 Interface Form KNN**

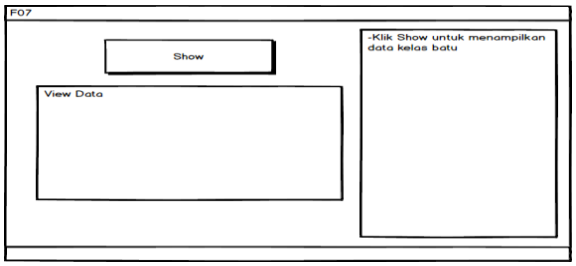



**Figure 12 Interface Form Encoding Class**

### 3.2.1 Procedural Design

*Designing procedural account of how the software runs depicted in flowchart form.*

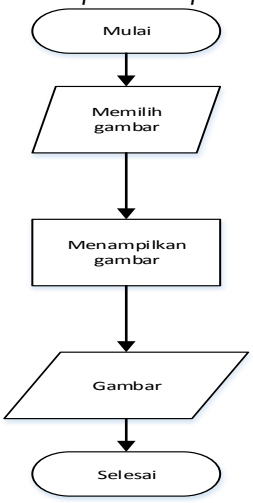



**Figure 13 Procedural Load Image**

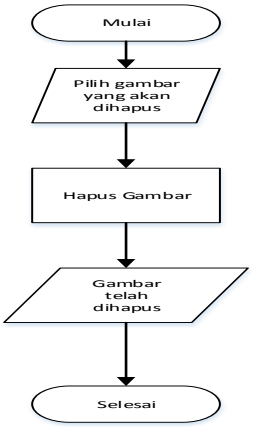



**Figure 14 Procedural Clear Image**

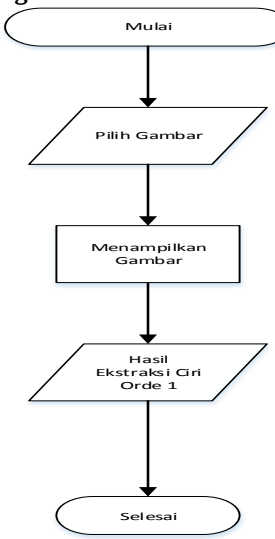



**Figure 15 Feature Extraction Procedure Order 1**

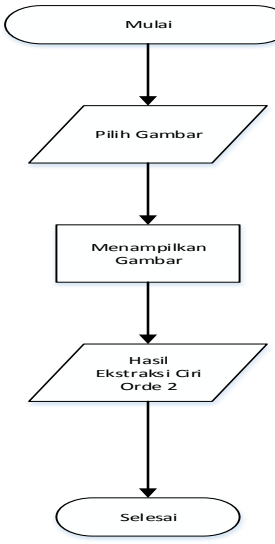



**Figure 16 Feature Extraction Procedure Order 2**

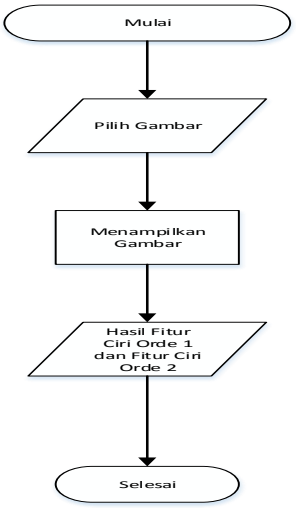



**Figure 17 Merger Procedural Order 1 and Order 2**

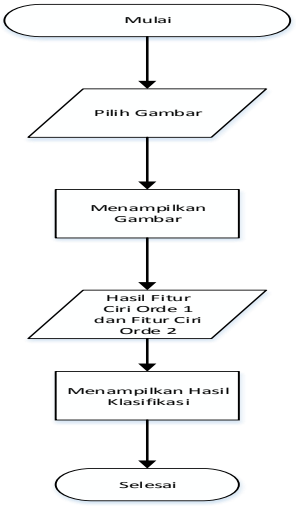



**Figure 18 Procedural Detection (KNN)**

### 3.3 Testing algorithm

*Testing of this algorithm aims to determine the performance of the method of order 1 and order 2, to measure the level of accuracy using methods confusion matrix.*

### 3.3.1 Testing *K-Nearest Neighbor* with K=1

*In this testing is done by testing using K-Nearest Neighbor with K = 1 is the ratio of training data and test data will be used is the ratio of 6: 4. Testing results with K-Nearest Neighbor with K = 1 can be seen in the table below:*

**Table 1 *Confusion Matrix* With K=1**

| Kelas | | Prediksi | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Badar besi | Giok hijau | Giok hitam | Giok merah | Giok putih | Merah siam | Pancawarna | Pecah seribu | Pirus | Sulaiman |
| Target | Badar besi | | | | 1 | | 1 | 1 | 1 | | |
| | Giok hijau | | 4 | | | | | | | | |
| | Giok hitam | 3 | | | | | | | | | 1 |
| | Giok merah | | | | 3 | | | 1 | | | |
| | Giok putih | | | | | 1 | | | | 1 | 2 |
| | Merah siam | 1 | | | 1 | | 1 | | 1 | | |
| | Pancawarna | 2 | | | 1 | | | | | | 1 |
| | Pecah seribu | 2 | | | | | 2 | | | | |
| | Pirus | | | | | | | | | 4 | |
| | Sulaiman | 1 | | | | 1 | | 1 | | | 1 |

$$accuracy = \frac{4+3+1+1+4+1}{40} x\ 100\ \% = \frac{14}{40} x\ 100\% = 35\ \%$$

### 3.3.2 Testing *K-Nearest Neighbor* with K=20

*In this testing is done by testing using K-Nearest Neighbor with K = 20 ie the ratio of training data and test data will be used is the ratio of 6: 4. Testing results with K-Nearest Neighbor with K = 20 can be seen in the table below:*

**Table 2 *Confusion Matrix* With K=20**

| Kelas | | Prediksi | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Badar besi | Giok hijau | Giok hitam | Giok merah | Giok putih | Merah siam | Pancawarna | Pecah seribu | Pirus | Sulaiman |
| Target | Badar besi | 2 | | | | | 1 | 1 | | | |
| | Giok hijau | | 4 | | | | | | | | |
| | Giok hitam | | | 1 | | | 1 | 1 | | | |
| | Giok merah | | | | 3 | 1 | | | | | |
| | Giok putih | | | | | 4 | | | | | |
| | Merah siam | 1 | | | | | 2 | 1 | | | |
| | Pancawarna | 3 | | | | | | 1 | | | |
| | Pecah seribu | | | | | | | | 4 | | |
| | Pirus | | | | | | | | | 3 | 1 |
| | Sulaiman | | | | 1 | | | | | | 3 |

$$accuracy = \frac{2+4+1+3+4+2+1+4+3+3}{40} x\ 100\ \% = \frac{27}{40} x\ 100\% = 67{,}5\ \%$$

### 3.4. Conclusion Testing

*Based on From the testing that was done then didapatkanlah conclusion that K-Nearest Neighbor can classify images with input statistical data directly comparing the closest distance to the training, based on testing with K-Nearest Neighbor with K = 1 has an accuracy rate of 35%, with a ratio of training data test data value = 6: 4. and testing with K-Nearest Neighbor with K = 20 has an accuracy rate of 67.5%, with a ratio of training data test data value = 6: 4. The more training data the greater the degree of accuracy and vice versa.*

## 4. CLOSING

### 4.1 Conclusion

*Based on the discussion of the analysis, design, implementation to the stage of testing for the Merger Implementation Method Feature Feature Feature Feature Order 1 and Order 2 On Citra For the classification of agate type, it can be concluded as follows:*

*Results of testing the accuracy of the implementation of merging methods characteristic features first-order and second-order characteristic features and classification using the K-Nearest Neighbor produces an accuracy of 67.5% using K = 20.*

**4.2 Advice**

*Apart from the advantages of a software must have flaws, and therefore for further development, advice that can be given as follows:*

*Further research is expected to be able to use other methods of feature extraction and classification methods such as using the Decision Tree, SVM, neural networks and other-other in order to compare the results of the accuracy and speed the process.*

**BIBLIOGRAPHY**

[1] "Pusat Infomasi Batu Akik, Pengertian Batu Akik," 19 September 2015. [Online]. Available: http://batu-akik.co/pengertian-batu-akik/.

[2] "Ciri-ciri batu akik asli ,," [Online]. Available: http://www.berjibaku.com/2014/09/ciri-ciri-batu-akik-asli-cara.html. [Diakses 19 September 2015].

[3] ""Tips aman membeli batu akik asli"," 5 Mei 2015. [Online]. Available: http://www.alaurang.com.

[4] M. Yuda Permadi, "APLIKASI PENGOLAHAN CITRA UNTUK IDENTIFIKASI KEMATANGAN MENTIMUN BERDASARKAN TEKSTUR KULIT BUAT MENGGUNAKAN METODE EKSTRAKSI CIRI STATISTIK," *INFORMATIKA*, vol. IX, no. 1, 2015.

[5] D. N. Maharsi, J. Halomoan dan R. D. Atmaja, ""KLASIFIKASI SERAT MIRING PADA KAYU MENGGUNAKAN EKSTRAKSI CIRI STATISTIK BERDASARKAN PADA PENGOLAHAN CITRA"," 2013.

[6] R. S. Pressman, Rekayasa Perangkat Lunak Pendekatan Praktis Buku 1,, Yogyakarta,: Andi,, 2002..

[7] S. Kusumadewi, Artificial Intelligent (Teknik dan Aplikasinya), Yogyakarta: Graha Ilmu, 2003.

[8] R. Munir, Pengolahan Citra Digital, Bandung: Informatika Bandung, 2002.

[9] B. Hendradjaya, "Catatan Kuliah Pengolahan Citra,", in Pengolahan Citra Digital, 1995.

[10] U. Ahmad, Pengolahan Citra Digital & Teknik Pemrogramanya, Yogyakarta: Graha Ilmu, 2005.

[11] A. A. Kusuma, R. R. Isnanto dan I. Santoso, "PENGENALAN IRIS MATA MENGGUNAKAN PENCIRIAN MATRIKS KO-OKURENSI ARAS KEABUAN," 2013.

[12] T. W. A. Putra, "Pengenalan Wajah Dengan Matriks Kookuren Aras Keabuan dan Jaringan Syaraf Tiruan Probabilistik," *Tesis Pasca Program Sarjana Universitas Diponegoro*, 2013.

[13] E. Prasetyo, Data Mining Konsep dan Aplikasi menggunakan Matlab, Yogyakarta: Andi, 2012.

[14] Y. Indra, Pemrograman Terstruktur, Yogyakarta: J&J Learning, Yogyakarta.

[15] A. B. Ladjamudin, Rekayasa Perangkat Lunak, Yogyakarta: Graha Ilmu, 2006..

[16] G. A. Away, The Shortcut of MATLAB Programing, Bandung: Informatika Bandung, 2006.

[17] R. A. d. M. Shalahuddin, Rekayasa Perangkat Lunak, Bandung: Informatika Bandung, 2013.

[18] A. Indriani, "Klasifikasi Data Forum dengan menggunakan Metode Naive Bayes Classifier," 2014.