IMPLEMENTATION OF ORDER STATISTICS EXTRACTION METHOD ONE AND TWO AND CLASSIFICATION K-NEAREST NEIGHBOR TO IDENTIFY THE IMAGE OF MINERAL ROCK SEDIMEN

Muhammad Rizky[1]

[1] Teknik Informatika – Univesitas Komputer Indonesia
St. Dipatiukur 112 - 114 Bandung
E-mail : kidurizky@yahoo.co.id[1]

ABSTRACT

One way to identify the image is to give the image texture. If the image is said to have a texture image pattern occurs repeatedly meet all of the image field. A different image have different characteristics. The characteristics are the basis for the classification of the image based on the texture. There are several methods to obtain the characteristics of texture in an image, several methods to obtain the characteristics of image texture was statistically order one and two. The characteristics of texture obtained from these methods include the mean, variance, skewness, kurtosis, energy, entropy, dissimiliarity, contrast, autocorrelation, correlation and homogenity. From the results of these characteristics are then used for classification by using K-Nearest Neighbor classification that determines the classification results based on the greatest probability. The object being tested is the image of a thin slice of rock mineral fotomikroskop results in the form of files * .jpg.

From the research that has been done, it can be deduced as follows: K-Nearest Neighbor method can perform image classification based on the texture extracted by the extraction method of order one and two. Because the data is in the form of feature extraction results continue the data, or so-called nominal data, so that the process of data classification feature extraction results can be directly used as input in the classification K-Nearest Neighbor.

based on test results, the conclusions obtained are K-Nearest Neighbor can classify images properly, because the data of the image texture feature extraction mineral rocks with statistical methods of order one and two have a distance interval between class apart. So naïve Bayes classification can work well when making classification.

Keywords: texture images, feature extraction, order one and two, classification, K-Nearest Neighbor

1. PENDAHULUAN

Sandstone or arenite is a type of sedimentary rock composed of mineral grains of sand and other organic materials in small sizes, these stones contain materials such as silica and carbonate cement that binds the sand grains together and containing granules of clay that occupies the space between the sand grains. Sandstone is the most common type of sedimentary rock is often mined for use as groundwater aquifers, and as a reservoir of oil and natural gas, is a great reservoir of oil and gas accumulation [8]. Some of the most common minerals found in sandstones is feldspar and quartz [7].

To determine the type and mineral content in the sandstone, geologists are still using the interpretation of experts to determine whether the types of minerals contained in mineral rocks and still images using the point counting method to determine what percentage of the mineral content in the image of the rock incisions microscope image results. This method is a statistical method which is done manually to measure the percentage of mineral deposits in the sedimentary rock, especially sandstone, the shortcomings of this method is that this method takes a long time to be done, in addition to doing point counting geologists still have to identify these minerals one by one so there is a possibility of error due to bias caused inconsistencies descriptors, so as to implement it needed innovation by making use of image processing with the identification of characteristic texture to improve the accuracy in determining the mineral content and efficiency calculate the percentage content of each mineral on sandstone

Image processing is not a new science, the science is widely used in various fields such as industry, education, medicine etc. Image processing into knowledge that is essential for basic image identification based classes that can mimic the human ability to classify image.

Statistical Methods Order One has several characteristics or parameters that can be used to determine the type of mineral based color of gray or

grayscale and Order Statistical Methods Two had several traits or parameters that can be used to determine the type of mineral based texture. In a previous study, entitled "Leaning On Wood Fiber Classification Using Statistical Feature Extraction Based on Image Processing" this study using the statistical feature extraction ie order one and two, the research generate 92.5791% accuracy when using the value k = 3,5,7 Euclidean distance models.

Image classification using the K-Nearest Neighbor, this method is used for image classification test to characterize database feature extraction results, the number of K nearest neighbor method is used because this method has a resilience training data that has a lot of noise and effective when training data is huge.

Based on the issues described in this research will be conducted research using Order Statistics Extraction Method One and Two and K-Nearest Neighbor To Identify Citra Mineral At Sedimentary rocks.

1.1 Problem Formulation

Based on preliminary described, this research is to formulate the problem to be discussed is:

a. How do I know the content of minerals in the sandstone by the texture by using image processing.

b. How do I know the percentage of the mineral content of each mineral in the microscope image of the photo.

c. How to implement the Statistical Methods for the Extraction of characteristics based on texture, K-Nearest Neighbor method for classification of mineral types of sandstone.

1.2 Purpose and Objectives

The purpose of this study was to implement a statistical algorithm of order one and two with KNN methods.

The purpose of this study is:

a. Knowing the type of the mineral content in the image of the fotomikroskop

b. Knowing the percentage of mineral content of each microscope image of the photo.

2. CONTENTS OF RESEARCH

2.1 Image processing

Utilization of digital imagery is widely used in various fields such as education, medicine, industry etc. Digital image processing is one of the disciplines that study the matters relating to the improvement of the image quality (contrast enhancement, transformation, color, image restoration), image transformation (rotation, translation, scaling, transformation, geometric), perform image recovery characteristics (feature images) is optimal for the purpose of analysis, the process of withdrawal of information or description of the object or the introduction of objects contained in the image, compression or data reduction for the purpose of data storage, transmission and processing time datam data dai input image is an image processing, and the output is the result of image processing.

2.2 Textur

Texture is an intuitive concept that describes the nature of smoothness, roughness and regularity in the region / area (region). In digital image processing, texture is defined as the spatial distribution of degrees of gray in a set of neighboring pixels [7]. Generally texture refers to the repetition of elements of the base texture called primitive or tekstel (texture element-texel) conditions for the formation of a texture, among :

1. The existence of primitive patterns consisting of a pixel or more. forms of these primitive patterns may be the point, straight line, curved line, area, and others which are the basic elements of a texture.
2. The primitive patterns recur at intervals and specific direction so that it can be predicted or found characteristic repetition.
3. An image of an interpretation different texture when viewed at different distances and angles. Humans look at the texture based on the description that is random, such as smooth, rough, regular, irregular and so forth. This is an incorrect description and non-quantitative, so it needed a description kuantititif (mathematically) to facilitate analysis [7].

An image of an interpretation different texture when viewed at different distances and angles. Humans look at the texture based on the description that is random, such as smooth, rough, regular, irregular, and so forth. This is an incorrect description and non-quantitative, so it needed a quantitative description (mathematical) to facilitate analysis (U. Ahmad, 2005).

2.3 Texture analysis

Texture analysis is the basis of a wide range of applications, application of texture analysis, among others: remote sensing, medical imaging, identification of quality of a material (wood, leather, textiles, etc.), and also a wide range of other applications. In the image analysis, texture measurement is categorized into five main categories: statistical, structure, geometry, basic model, and signal processing. Statistical approach considering that the intensity generated by a random two-dimensional field, the method is based on the frequency space. Examples of statistical methods is the autocorrelation function, Co-ocurence Matrix,

Fourier transformation, frequency edge, run length and more. Structural engineering related to the preparation of the smallest portions of an image. Examples of structural methods are fractal models. Methods based on the geometry of the existing geometry on the element of texture. Examples of the method is the basic model of a random field. While the method is a method of signal processing based on the analysis of the frequency of such transformation and the Gabor wavelet transformation [6].

2.3.1 Statistical methods of order one and two

Is a methods whose elements represent 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. (Ahmad U., 2005)

A neighboring pixel which has a distance and between them, can be located in eight different directions, as shown Figure

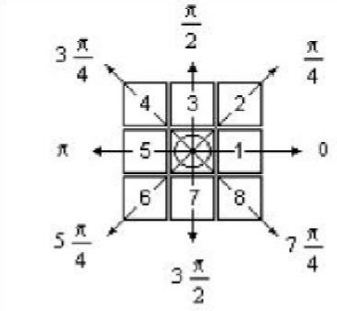


Figure 2 Relationship between pixel adjacency

Dalam metode ini, haralick et al propose various types of statistical characteristics of texture that can be extracted Some of these include among others are: Contrast (contrast), homogeneity (homogeneity), Entropy (Entropy), Energy (Energy) and Dissimilariti (dissimilarity).
The equation for these features are as follows:

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. The results contrast calculation related to the amount of gray in the image intensity diversity.

$$Kontras = \sum_{i,j=0}^{n-1} (Pi, j).(i-j)^2$$

Where:
i and j are the properties of gray pixels of resolution 2
nearby
p (i,j) adalah *Probabilitas kolom(i,j)*

Homogenity

homogeneity determine degree of homogeneity image of a kind of gray. Homogeneous image will have a great homogeneity prices.

$$\sum_{i,j=0}^{n-1} P(i, j) / [1 + (i - j)^2]$$

Where:
i and j are the properties of gray pixels of resolution 2
nearby
p (i, j) is Probability column (i, j).

Entropy

Entropy can show irregularity, size, shape, if a large entropy value for the image with uneven degrees of gray transitions and image of little value if the structure is irregular (variable).

$$Entropi = -\sum_{i,j=0}^{n-1} P(i, j) \ \log(P(i, j))$$

Where:
i and j are grayish nature of two adjacent pixels resolution
p (i, j) is the normalized Symmetric Matrix Cooccurence

Energy

Energy stated measure of the concentration of gray pair with a particular intensity in the matrix, where (i, j) declared value on row i and column j in the matrix kookurensi.

$$Energi = \sum_{i,j=0}^{n-1} P(i, j)^2$$

Where:
i and j are grayish nature of two adjacent pixels resolution
p (i, j) is the normalized Symmetric Matrix Cooccurence

Dissimilarity

Stated dissimilarity measure inequality of gray image so as to provide an indication of the structure in the image.

$$Dissimilarity = \sum_{i,j=0}^{n-1} P(i, j).|\, i - j \ \ |$$

Where:
i and j are grayish nature of two adjacent pixels resolution
p (i, j) is the relative frequency matrix of two adjacent pixels resolution.

2.4 Classification

Classification is a job that assessment of an object's data to fit in a certain class of the number of classes available. (Prasad, 2012).

There are two main jobs:

1. Development of a model as a prototype to be stored as memory
2. Using the model to perform recognition / classification / projection on an object other data entered on which class

2.4.1 K-Nearest Neighbor Classification

K-Nearest Neighbor is a method of classification of a set of data based on the learning data that has been previously classifiable. KNN included in the supervised group, in which instance the new query results are classified by the majority of the proximity of the existing categories in the KNN. Later a new class of data will be selected based on a group class that is close to the distance vector. [8]

Method of k-nearest neighbor (KNN) is simple, works on the shortest distance from the query instance to the training sample to find its KNN. Training projected onto the sample-space lot, where each dimension represents the features of the data. The room is divided into sections based on the classification trainning sample. a point in this space marked C class if the class c is a classification of the most commonly found on the k nearest neighbors of the point. Near or far neighbors are usually calculated based on those Euclidean Distance.

Euclidean 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 is represented as follows:

$$d(x-y) = \sqrt{\sum_{j=1}^{i}(x_j - y_j)^2}$$

Information:

d = distance learning data to the test data.

x_j = j-th test data, with j = 1, 2, . . . n.

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

The precision of the method KNN is strongly influenced by the presence or absence of features that are not relevant or if the weight of such features is not equivalent to its relevance to the classification. Research into this method mostly discusses how to select and weight the feature for performance klasfikasi be better. Steps to calculate k-nearest neighbor:

1. Determine the parameters K (the number of nearest neighbors)
2. Calculate the Euclidean distance squared (query instance) of each object on the data samples provided.
3. Then sort these objects into groups with the smallest euclidean distance.
4. Gather category II (Classification nearest neighbor).
5. Using the nearest neighbor category, the majority of the most predictable queries instance values that have been calculated.

3.5 **Method of testing accuracy**

In weka machine learning, testing accuracy can be done with two types of testing, ie :

1. Training set test
2. Supplied set test Training set test

Testing method using the data that has been in training, in other words, the training data and test data is the same data

1. Training data test

Methods of testing using different data, in other words, the training data different from the data that will be tested

2. Supplied set test

One technique for assessing / validating the accuracy of a model built on a particular dataset. Making models usually aim to predict and classify a new data that may have never appeared in the dataset. The data used in the model development process called training data / training, while the data will be used to validate the models referred to as test data.

2.6 Process Analysis

Analysis of the process to be done in this study is an analysis within the classification image based on texture. stages of work processes within the classification ranging from data input to output data. Here are the stages of the process analysis to be performed can be seen in figure

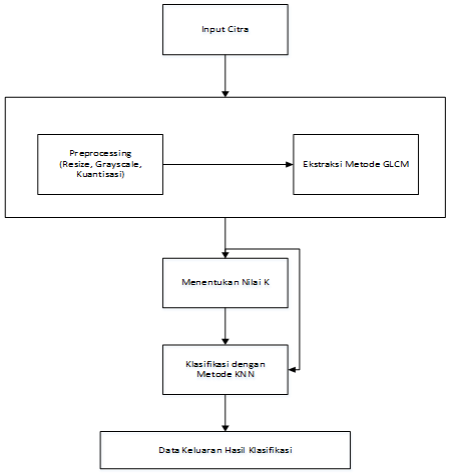


Figure 3 Flowchart analysis process

2.6.1 Analysis Data input

In this study, the first to be done is the analysis of input data. Analysis of the data input is done to obtain an input value that can later be used for the classification process in the method KNN. In This study, the input data is an image, which will

searchable content value by using a feature extraction method of order one and two, the value of output that would be obtained is the contrast, homogeneity, entropy, energy, and dissimilarity, etc. Those values will be used as baseline data that will be used as input in the K-Nearest Neighbor method. Stages that will be performed on the input data analysis is preprocessing is to perform segmentation, cropping, grayscale. After performing preprocessing, feature extraction is carried out by statistical methods and to get the first order nilai fitur dari citra tersebut.

2.6.1.1 Preprocessing

In this study, preprocessing is done to make it easier to get the value of feature extraction. Preprocessing will do is resize, grayscale and image quantization. The following process flow of preprocessing:

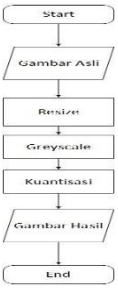


Figure 4. Flow preprocessing

2.6.1.1.2 Grayscale

Grayscale is a process to change the color to greyish. By changing the RGB value of each pixel of the image into the same value so that each pixel has the same value for all three elements of color and grayscale matrix values obtained. Following the flow of the process grayscale:



Figure 6 grayscale process flow

2.6.1.2 Feature extraction

Feature extraction is the process to get the main characteristics contained in the image, which has been in a grayscale image will produce a matrix grayscale who had already in quantization matrix is exactly what will be used at this stage. This stage will calculate the statistical value of kookurensi 5 ie contrast, energy, entropy, homogeneity and dissimilarity with symmetry angles of 0, 45, 90 and 135. Having obtained all grades from the corner will averaged. Following the flow of feature extraction process:



Figure 8 feature extraction process flow

2.6.3 Analysis of testing

Testing is the stages in the process of image classification based texture, this process can be generated imagery that included the probability for more details can be seen in the following process flow :

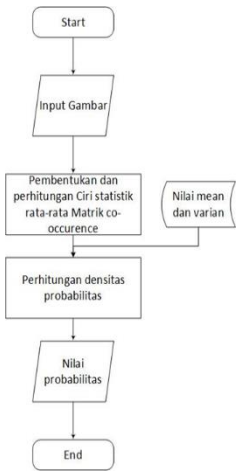


Figure 10 flow testing process

2.6.4 Analysis of output data

Analysis of output data is the final stage that will be done. which is looking for the greatest value on the probability value obtained at the time of testing analysis. The following is the process flow looking output data.

2.7 examination

In this study, the test was done by using two methods. Here is the test to be performed:

1. Test the image included in the database (training set test).
2. Test the image that are not included in the data base (supplied set test).

Testing with method 1 Test image included in the database

Testing method 1 was conducted by examining the image included in the database, this test aims to determine the level of recognition of the image that has been trained, the image data that is used there are 120 pieces of imagery which consists of three classes, with each class there are 40 images.

Training data is used, and the test data in Appendix C1

Results Testing method 1.

Table 1 Level of Accuracy of each class on the testing method 1

Kelas		Prediction		Count Citra	Akurati
		Kuarsa	Feldspar		
Target	Kuarsa	14	2	16	87.5%
	Feldspar	10	4	16	25%
average					56.25%

Testing with Test Method 2 images that are not included in the database

Testing method 2 was conducted by examining the image that are not included in the database, this test aims to determine the level of the test image recognition outside the database on the image of the train in the database. image data used to train there were 100 pieces of imagery which consists of five classes, with each class there are 20 images. And also used the test image data there are 100 pieces of imagery which consists of five classes, with each class there are 20 images.

Training data and test data exist on the attachment.

Assay results using the test method 2.

Table 2 Level of Accuracy of each class on the testing method 1

Kelas		**Prediksi**		**Jumlah Citra**	**Akurasi**
		Kuarsa	**Feldspar**		
Target	Kuarsa	12	4	16	75%
	Feldspar	13	3	16	18%
average					46.5%

2.8 conclusion Testing

Based on the results of one test scenario that is testing the same test data with training data, it can be concluded that the K-Nearest Neighbor method can classify with an accuracy of 70%. Based on the test scenario 2 is test of test data that are not in training data, KNN can classify with an accuracy of 46%.

From the test results, the accuracy has a very good level, because the data generated by the feature extraction feature extraction method of order one and two have a large degree of inequality, so the recognition process can be run well.

3. CLOSING

3.1 Conclusion

From the test results and analysis, it can be concluded the following matters :

1. In this study, the highest level of accuracy of the image obtained in class mineral quartz with an average value of accuracy 90%
2. For the class Feldspar mineral image accuracy results below 50%, it is because the type of image that is homogeneous so happens similarity value extraction characteristics of these images.

3.2 Suggestion

In the making of this final project, there are still many deficiencies that can be corrected for the next development. Some advice that can be given is:

1. Adding some other feature extraction, feature extraction such as color, shape, etc..
2. To be able to compare the performance of a statistical method of order one and two as this feature extraction, texture analysis can be made by different methods, such as the autocorrelation method, Run-Length, Sum and Difference Histogram and more.
3. Conduct research using the image classification other classification algorithms, such as using the Decision Tree, SVM, neural networks and other-other in order to compare the results of the accuracy and speed of the process.

BIBLIOGRAPHY

R. Munir, *Pengolahan Citra Digital, Bandung*: Penerbit Informatika Bandung, 2002.

S. R. Pressman, *Software Engineering: A Practitioner's Approach*, 4th ed, New York: McGraw-Hill Companies, 2010.

Godfrey-Smimth, D. (2005). Beta Dosimetry of Potassium Feldspars in sediment Exstract Using Imaging Microphobe Analysis and Beta Counting. Geochronometria, 7-12.

Indah Ratih, "Pengenalan Motif Batik Menggunakan Metode Transformer Paket *Wavelet*," *Tugas Akhir Teknik Informatika*, no. Universitas Telkom, Bandung, 2013

King, Hobart. "Sandstone, A clastic sedimentary rock composed of sand-sized grains of mineral, rock or organic material". 25 Desember 2015.

http://geology.com/ rocks/sandstone.shtml.

U. Ahmad, *Pengolahan Citra Digital dan Teknik Pemogramannya*, Yogyakarta: Graha Ilmu, 2005.

Y. Ganis, I. Santoso,Dll "Klasifikasi Citra dengan Matriks Kookurensi Aras Keabuan Pada Lima Kelas Biji-Bijian," *Tugas Akhir Teknik Elektro*, no. Universitas Diponegoro, Semarang, 2009.

Widya Eka Wardani, "Pengenalan Motif Batik Menggunakan Metode Transformer Paket *Wavelet*," *Tugas Akhir Teknik Informatika*, no. Universitas Widyatama, Bandung, 2013.

Ariantoko Kusmian, "Implementasi Algoritma Naïve Bayes Untuk Klasifikasi Citra Berdasarkan Ekstraksi Ciri Tekstur dengan Metode Matriks Kookurensi," *Tugas Akhir Teknik Informatika*, Universitas Komputer Indonesia, Bandung, 2014

Nugroho A., *Pemodelan Berorientasi Objek*, Bandung: Informatika, 2005.

Insanudin Andi, "Ekstraksi Informasi Kemacetan pada Media Digital," *Tugas Akhir Teknik Informatika*, no. Universitas Komputer Indonesia, Bandung, 2013.

Nithya, dan Santi, B.,2011 "Comparative Study on Feature Extraction Method For Breast Cancer Clasification, *J Theoritical and Applied Information Technology*", Vol 33 No.2.

Listia Refta, Harjoko Agus "Klasifikasi Massa pada Citra Mammogram Berdasarkan Gray Level Cooccurence Matrix (GLCM)," *Jurnal*, . Universitas Gadjah Mada, Yogyakarta, 2014.

E. Budiyono Permadi, Nikentari Nerfita,. "Analasis Klasifikasi Kadar Karat Emas Menggunakan Metode K-Nearest Neighbor," *Jurnal Teknik Informatika*, Universitas Maritim Raja Ali Haji , Tanjung Pinang,.

L. Farsiah, T. Fuadin A dll "Klasifikasi Gambar Berwarna Menggunakan K-Nearest Neighbor dan Support Vector Machine," *Tugas Akhir Teknik Informatika*, no. Universitas Komputer Indonesia, Bandung, 2013.

Wijaya,. "Pengaruh Kadar Gabah Terhadap Mutu Fisik beras Giling," *Jurnal Fakultas Pertanian*, Universitas Unswagati , Cirebon,.