TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol. 13, No. 3, March 2015, pp. 561 ~ 567
DOI: 10.11591/telkomnika.v13i3.7197

Modeling Singular Value Decomposition and K-Means of Core Image in Clasification of Potential Nickel

Agung Prajuhana Putra*[1], Agus Buono[2], Bib Paruhum Silalahi[3]
[1,2]Departement of Computer Science, Faculty of Mathematics and Natural Sciences,
Bogor Agricultural University, 16680 Bogor, Indonesia, Ph/Fax. +62-251-628448/622961
[3]Departement of Mathematic, Faculty of Mathematics and Natural Sciences,
Bogor Agricultural University, 16680 Bogor, Indonesia
*Corresponding author, e-mail: agungpp_mail@yahoo.com[1], pudesha@yahoo.co.id[2],
bibparuhum1@yahoo.com[3]

Abstract

Exploration is a main process in the nickel mining activities. One of the most important steps in exploration is obtain soil samples (cores) to determine the potential of nickel in the soil. Laboratory testing is a way to know how much the nickel content on the core. This research aims to utilize the core image of the statistical characteristics of color and texture, Biplot analysis using SVD, K-Means and identification using SVM method with RBF kernel and polynomial to determine the potential of nickel.

Keywords: *SVD, K-Means, SVM, RBF, polynomial*

1. Introduction

Indonesia is rich in natural resources, one of which is a mineral. One type of sediment which has the potential to be used as mines is sedimentary laterite. Sediment widely used for mining laterite nickel ore and / or iron-nickel (Fe-Ni) [1].

To obtain the nickel ore should be done by mining activity exploration. Exploration is a survey activity of the investigation and assessment area that estimated contains valuable minerals. These activities generate information of the soil that is usually performed by a geologist.

According to Hazria [2], the research using the data from the satellite fractal just take a sample surface, so it cannot be used as an indicator to obtain information on the percentage of nickel content below the surface. One of the methods done is the analysis of the nickel content of the bedding to see the stratigraphic. Therefore, this study aimed to determine the stratigraphic laterite of nickel content in the sediment.

Another research is conducted to determine the potential of nickel in soil by using XRF (X-Ray Fluorescence). XRF is a tool that uses spectrometry method to analyze the content of particular material elements. XRF spectrometry utilizes the X-rays emitted by the material that is subsequently captured by the detector to analyze the content of the element. So far, the use of XRF instrument is to analyze metal alloys, copper, aluminum alloy, rocks, minerals, and crust. Materials that can be analyzed are in the form of a massive solid, plate or powder. Elemental analysis performed both qualitatively and quantitatively. Quantitative analysis is to determine the number of elements contained in the material [3].

Until now, the method to determine the potential of nickel in the core should be done through laboratory testing using X-ray beam with a relatively long process that is more than 10 hours as well as the considerable cost. Based on this, it is necessary to investigate the characteristics of the model extraction and classification methods appropriate to the classification of nickel by using the image classification process nickel cores that can be done more quickly.

Several studies based on the identification of such imagery has been developed for face recognition with a 84 % accuracy rate [4], and then developed reached 97.3 % [5]. In this research, the classification of nickel in soil sample results of exploration (core) based on the characteristics of color and texture with the singular value decomposition (SVD) and K-Means

Received October 2, 2014; Revised December 16, 2014; Accepted January 5, 2015

as a reduction of the characteristics, as well as the use of Support Vector Machine (SVM) as a method for classification.

2. Research Method

2.1. Preprocessing

This research used 160 image data which are divided into four classes, where each class consists of 40 images. This image comes from the results of the exploration performed 10 times at different places in 2013. The exploration results image have different dimensions, then do cropping with dimensions of 1200 x 120 pixels.

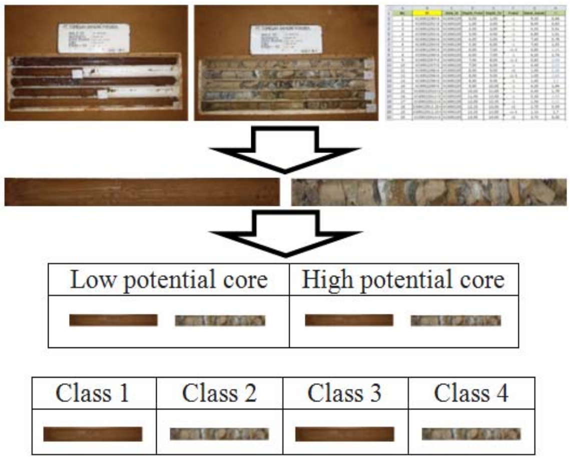

Figure 1. Preprocessing

2.2. Grouping with K - Fold Cross Validation

At this stage, the image of each class will be divided or scrambled into subgroups, from subgroup k, taken one subgroup to be used as data validation test, and the rest is used for training data, the process is repeated so that all subgroups can be used as testing data. Fold used in this research consists of 4 subgroups.

2.3. Feature Extraction

This research used a characteristic extraction based on statistical color and texture. Feature color can be obtained through statistical calculations such as mean (1), standard deviation (2), skewness (3) and kurtosis (4) [6]. For example, enable this feature can be used to identify the interests of ornamental plants [7]. Calculation imposed on each component of R, G and B.

$$\boldsymbol{\mu} = \frac{\mathbf{i}}{\mathbf{MN}}\sum_{\mathbf{i=1}}^{\mathbf{M}}\sum_{\mathbf{j=1}}^{\mathbf{N}}\mathbf{P_{ij}} \tag{1}$$

$$\boldsymbol{\sigma} = \sqrt{\frac{\mathbf{1}}{\mathbf{MN}}\sum_{\mathbf{i=1}}^{\mathbf{M}}\sum_{\mathbf{j=1}}^{\mathbf{N}}\left(\mathbf{P_{ij}} - \boldsymbol{\mu}\right)^{\mathbf{2}}} \tag{2}$$

$$\boldsymbol{\theta} = \frac{\sum_{\mathbf{i=1}}^{\mathbf{M}}\sum_{\mathbf{j=1}}^{\mathbf{N}}\left(\mathbf{P_{ij}}-\boldsymbol{\mu}\right)^{\mathbf{3}}}{\mathbf{MN}\boldsymbol{\sigma}^{\mathbf{3}}} \tag{3}$$

$$\boldsymbol{\gamma} = \frac{\sum_{\mathbf{i=1}}^{\mathbf{M}}\sum_{\mathbf{j=1}}^{\mathbf{N}}\left(\mathbf{P_{ij}}-\boldsymbol{\mu}\right)^{\mathbf{4}}}{\mathbf{MN}\boldsymbol{\sigma}^{\mathbf{4}}} \tag{4}$$

While the texture characteristics obtained from statistical Grey Level Co-occurrence Matrices (GLCM). GLCM first proposed [8] with 28 Feature to explain the spatial patterns [9]. GLCM uses textures in second-order calculation. In the second order, the pairwise relationship between two original image pixels is taken into account [10]. To get Feature GLCM, only some scale that proposed by Haralick. For example [11] only uses five scale for GLCM, in the form of angular second moment (ASM) (5), Contrast (6), inverse different moment (IDM) (7), entropy (8) and correlation (9).

$$\sum_{i=1}^{L}\sum_{j=1}^{L}(\mathbf{GLCM}(\mathbf{i},\mathbf{j}))^2 \quad (5)$$

$$\sum_{n=1}^{L}\mathbf{n}^2\left(\sum_{|i-j|}\mathbf{GLCM}(\mathbf{i}-\mathbf{j})\right) \quad (6)$$

$$\sum_{i=1}^{L}\sum_{j=1}^{L}\frac{(\mathbf{GLCM}(\mathbf{i},\mathbf{j}))^2}{1+(\mathbf{i}-\mathbf{j})^2} \quad (7)$$

$$\sum_{i=1}^{L}\sum_{j=1}^{L}\left(\mathbf{GLCM}(\mathbf{i},\mathbf{j})\mathbf{log}(\mathbf{GLCM}(\mathbf{i},\mathbf{j})\right) \quad (8)$$

$$\frac{\sum_{i=1}^{L}\sum_{j=1}^{L}(\mathbf{ij})\left(\mathbf{GLCM}(\mathbf{i},\mathbf{j})-\mu_i{}'\pi_j{}'\right)}{\sigma_i{}'\sigma_j{}'} \quad (9)$$

Sample feature extraction presented in Table 1.

Table 1. Units for Magnetic Properties

Stat. Warna	Stat. Tekstur (GLCM) 0°	45°	90°	135°
mean_r: 117.0014 mean_g: 104.6245 mean_b: 86.0524 dev_r: 31.8589 dev_g: 32.8633 dev_b: 33.0866 skew_r: 0.1130 skew_g: 0.3337 skew_b: 0.6607 cur_r: -0.3372 cur_g: -0.1220 cur_b: 0.3154	asm: 1.5540e-004 Contrast: 478.2468 idm: 0.0936 entropy: 9.1237 correlation: 6.5999e-004	asm: 1.5270e-004 Contrast: 493.1846 idm: 0.0926 entropy: 9.1377 correlation: 6.5191e-004	asm: 3.6057e-004 Contrast: 83.9058 idm: 0.1966 entropy: 8.3030 correlation: 7.8858e-004	asm: 1.5187e-004 Contrast: 499.1584 idm: 0.0918 entropy: 9.1411 correlation: 6.5043e-004

2.4. SVD and Biplot Analysis

Biplot was derived from the SVD decomposition of the data matric and the characteristic extraction results will be displayed in the form of 2 dimension image Biplot visualization. From the picture, we can see the relationship between the variables and the comparison between classes that characterizes the potential of nickel in the core. If the variables were overlapping or contiguous then these variables have in common / linkages that can be reduced into one new variable.

2.5. K-Means

K-Means is used to optimize the combination of variables that most represents the core of nickel characteristics, so that when tested with SVM models will generate value Optimum accuracy. Combination value (K) generated from the previous stages in the biplot analysis phase.

2.6. Modeling With Support Vector Machine

Modeling by using Support Vector Machine is multi class; SVM has several methods in comparing objects, they are one-against-one and one-against-all. In this research will be used one-against-all with RBF kernel functions and polynomial.

2.7. Stages of Testing

The test data used in the research possess of 4 groups, each group contains 40 images, and each of these groups will be used as test data alternately, so that the whole image will be tested.

Model testing will be used the Support Vector Machine modeling. Data tested is the data reduction that resulted from training data to yield the correct classification of reduces test images. The method used is "one - against - all" where the classification is trained by the entire data to compare the accuracy of each kernel function and feature extraction.

2.8. Result Analysis

Training and testing process used the SVM modeling to yield the accuracy or successfulness of identifying the classification of potential nickel and image recognition error rate in the core of each class. Accuracy is calculated based on test data on 4 fold validation process, in order to determine the fault distribution by using the confusion matrix.

$$\text{Accuracy} = \frac{\text{number of true classification}}{\text{number of all data}} \text{x}100\%$$

3. Results and Analysis

Data core was obtained from the data of 10 times exploration preprocessing by cropping with dimension of 1200x120 pixels, and classified based on lab test results data into 4 classes with 40 images for each class. The four classes are: Class 1: Low Potential of Smooth Texture, Class 2: Low Potential of Rough Texture, Class 3: High Potential of Smooth Texture, Class 4: High Potential of Rough Texture

Low potential category if the percentage of ≤ 1.5 % of nickel and high potential if levels ≥ 1.5 % of nickel. While the texture can be seen from the shape, if it is rocky then the category is rough texture and if it is not then the category is smooth texture.

3.1. Training and Testing Data

Each class image of K-Fold method will be divided into subgroups, from the subgroup k will be taken a subgroup as data validation test, and the rest is used for training data. The process is repeated so that all subgroups can be used as test data. Fold used in this study consists of 4 subgroups. The K-Fold distribution of the testing and training data is shown below.

Class	Number of Data			
1	Test Data	Training data	Training data	Training data
2	Training data	Test Data	Training data	Training data
3	Training data	Training data	Test Data	Training data
4	Training data	Training data	Training data	Test Data

Table 2. Training and Testing Data with K - Fold

Pattern	Training Data	Testing Data
1	11,12,13,14,15,16,17,18,19,20, 21,22,23,24,25,26,27,28,29,30, 31,32,33,34,35,36,37,38,39,40	1,2,3,4,5,6,7,8,9,10
2	1,2,3,4,5,6,7,8,9,10,21,22,23,24,25, 26,27,28,29,30, 31,32,33,34,35, 36,37,38,39,40	11,12,13,14,15,16,17,18,19,20
3	1,2,3,4,5,6,7,8,9,10, 11,12,13,14,15,16,17,18,19,20, 31,32,33,34,35,36,37,38,39,40	21,22,23,24,25,26,27,28,29,30
4	1,2,3,4,5,6,7,8,9,10, 11,12,13,14,15,16,17,18,19,20, 21,22,23,24,25,26,27,28,29,30	31,32,33,34,35,36,37,38,39,40

3.2. Color and Texture Feature Extraction

K-Fold training data subgroups extracted based on the color and texture characteristics. Color extraction provided by calculating the mean value of the component R (red), G (green), and B (blue). While the texture extraction provided by GLCM value of the image. The overall value extraction dimension of 160x32 is shown in Table 3.

Table 3. Color and Texture Feature Extraction

	1	2	3	..	..	..	32
1	96.6	57.6	27	13.8	12.2	8.85	-0.2
2	91.8	51.5	19.5	14.8	12.4	6.96	-0.44
3	75.9	42.4	21	13.3	12.6	10.1	-0.14
4	108	94	56.2	22.7	22.4	22.3	-0.38
.	88.7	77.6	46.7	25.7	30.4	32	-0.2
.	80.2	67.6	35.5	29.1	30.1	28.3	0.73
.	98.7	88	55.1	25.3	25.9	26.9	-1
.	88.9	77.9	45.5	27.2	27.5	26.4	-0.43
.	104	83.1	44.2	25.2	26	24.8	0.28
160	114	97.8	58.8	22.4	22	20.4	-0.18

3.3. Biplot Analysis

From the data matrix extraction of 160x32 pixels with singular value decomposition (SVD) technique can be described in a biplot, as shown in Figure 2.

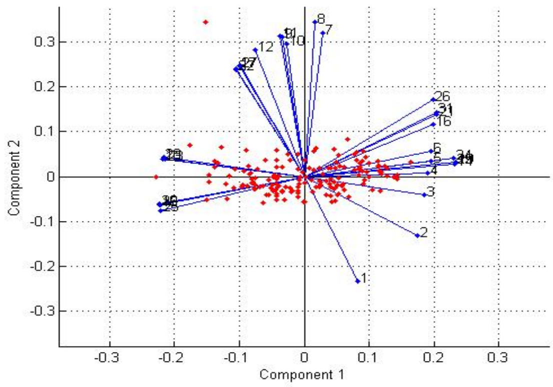

Figure 2. Biplot relationship of variable with the object.

Biplot analysis is to reduce the variable by combining variables that have relevance or nearly equal into one new variable. The biplot variable reduction analysis described in Figure 3 and Table 4 below.

Table 4. Reduction with Biplot Analysis

No	New Variable	Variable Reduction
1	V1	1
2	V2	2
3	R1	3,4,5,6,14,19,24,29
4	R2	16,21,26,31
5	R3	7,8,9,10,11,12,17,22,27,32
6	R4	13,18,23,28
7	R5	15,20,25,30

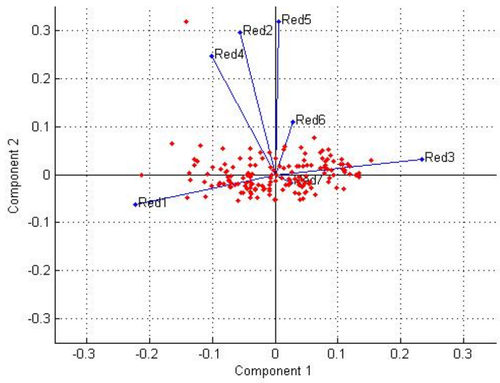

Figure 3. Variable reductions with biplot analysis

Reduction variable with biplot analysis resulted in seven variables that represent the characteristics of the nickel core image. The value of 7 variable is then became constant values in the K-Means.

3.4. Variable Reduction with K –Means

The number of biplot analysis variables became a constant (K) in the reduction stage by using the K-Means method. This method is to optimize the combination that represents the nickel core on the image automatically. The pair wise combination of these variables was formed from the iteration process during the training/testing models.

3.5. Nickel Potential Classification Using SVM

In SVM model building process, the determination value of parameter in the kernel function is very affected to the output. The more optimal the value of parameter then the better the resulting model. The value of parameter is needed to produce the precision on a model. In this research, the kernel parameters obtained by using grid search method within a certain interval and with RBF kernel and polynomial. The classification process using SVM with one against all method functions of the kernel, image size, and the reduction of different variables. Some of them such as:

a) Experiments using the RBF kernel function parameter $\sigma = 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50$ with a 0-90 % image reduction and with the constant K-Means = 4, 5, 6, 7, 8, 9, 10.
b) Experiments using a polynomial kernel function parameter d = 1, 2, 3, 4, 5 with a 0-90 % image reduction and with a combination Means= 4, 5, 6, 7, 8, 9, 10. After each process is performed, resulting in optimum accuracy in parameter RBF values σ= 75, image reduction= 0% (1200x120) and k-means variable reduction with k=7. The accuracy of each process is illustrated in Figure 4.

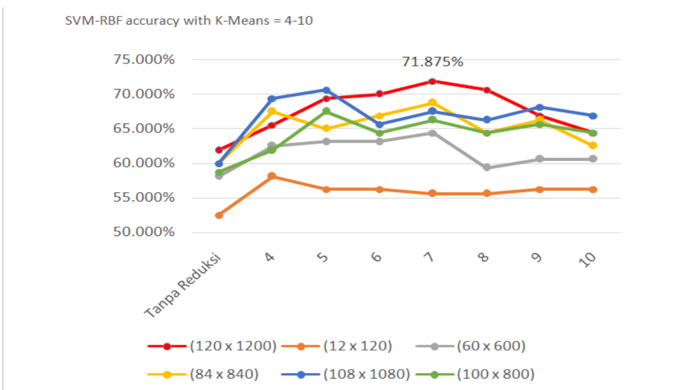

Figure 4. SVM-RBF accuracy results using K-Means

4. Conclusion

Potential nickel classification in the image core is using SVD Biplot reduction and K-Means and SVM as modeling. Reductions of the imagery used in this research experiments are between 0% until 90%, while the SVM model using RBF kernel function and kernel polynomial with the method of one–against-all.

The research concluded as follows:

1) Biplot analysis results can be used as the value of K in K-Means.
2) The results of the potential nickel classification using SVM method with RBF kernel function has an accuracy rate of 71.875 % on K-Means = 7.
3) The results of the potential nickel classification using SVM with polynomial kernel function has an accuracy rate of 62.5 % in the K-Means = 7.
4) The results of image classification of nickel after reduced variable with K-Means have higher accuracy compare to before.

This research can be developed with fitted the selection of methods for obtaining optimal value parameter in SVM kernel compare with other classification methods such as K-NN method.

Acknowledgements

I would like to express my special thanks of gratitude supervisor commission who gave me golden guidance until I am able to complete this research, as well as the Directorate General of Education in Higher Education (DIKTI), who respectively have contributed funds to study and suggestions for me.

References

[1] Simanjuntak. Determination of Nickel Content on Sediment stratigraphy laterite. 1994.
[2] Hazria. Laterite nickel in sediments. 2007.
[3] Rudi Suryadi. Determination of Nickel Content on Sediment stratigraphy laterite in Kendari. Thesis. Surabaya: University Haluoleo; 2011.
[4] Yang Jiang and Zhang David. A New Approach to Appearance-Based Face Representation and Recognition. *IEEE Transaction on Pattern Analysis and Machine on Intelligence.* 2004; 26(1): 1-9.
[5] Le Thai Hoang, Bui Len. Face Recognition Based on SVM and 2DPCA*. International Journal of Signal Procesing, Image and Pattern Recognition Procesing.* 2011; 4(3): 85-93.
[6] Martinez, WL and Martinez, AR. Computational Statistics Handbook with Matlab. CRC Press LLC Florida. 2002.
[7] Abdul Kadir and AdhiSusanto. Theory and Application of Image Processing. Andi Yogyakarta. 2013.
[8] Haralic RM, K Shanmugam, ItshakDinstein. Image texture classification*. IEEE Transactions on Systems, Man and Cybernetics.* 1973; 3(6).
[9] Kulkarni, AD. Artificial Neural Network for Image Understanding. Van Nostrand Reinhold, New York. 1994.
[10] Hall-Bayer, M. HIS Co-representation of circular and non-circular variables using harmonic analysis parameter. *Canadian Journal of Remote Sensing*. 2007; 33(5): 416-421. (in this case Vol.33, Issues 4, and page 416-421).
[11] Newsam S, Kammath C. *Comparing Shape and Texture Fitures for Pattern Recognition in the Simulation Data*. On the IS&T/SPIE 'S Annual Symposium on Electronic Imaging. San Jose, USA. 2005.