Land cover classification using Grey Level Co-occurrence Matrix and Naive Bayes

Sofia Sa’idah¹, Nor Kumalasari Caecar Pratiwi², Bandiyah Sri Aprilia³, Rita Magdalena⁴, Yunendah Nur Fu’adah⁵

¹,²,³,⁴,⁵ Electrical Engineering Department, Telkom University, Bandung, Indonesia

E-mail: sofiasaidahsf@telkomuniversity.ac.id

Abstract. Land cover data is important information to describe how much of a region is covered by plantation, forest, residential, rice field and river. In many applications the required information relates to the coverage of land cover class in a region, which is generally derived from a count of the pixels allocated to the class of interest in a classification. The design of the system in this study conducted for detecting land cover using Grey Level Co Occurrence (GLCM) method is used as the extraction in process of taking main image and Naive Bayes as a classification of grouping the images based on the types of land cover. Based on the testing data which is consist of 150 images we obtained the best accuracy is 85% with 206.6715 seconds computation time.

1. Introduction

Information related to land cover is very important and significant to know as a consideration in making a decision continuously. Information about land cover can be obtained by using digital image processing. Previous research has used GLCM method to conduct texture analysis [5]. The present study will classify the type of land cover using Grey Level Co-Occurrence Matrix (GLCM) as extraction method because this method is able to analyze an image based on the grey level between the pixel and its neighbor. This method can also produce 22 second order statistical feature although in this study only 5 parameters are considered as an important parameter. These five statistical features will be used for the classification process with the Naïve Bayes method. According to the Bayes Theorem, Naïve Bayes method has a high accuracy and high speed when it applies to small training and testing data to determine the best parameter which is needed in the classification process [6].

The GLCM, which is a square matrix can reveal certain properties about the spatial distribution of the gray-levels in the texture image [1]. Gray Level Co-Occurrence Matrix (GLCM) is a statistical feature extraction method with histogram and probability calculations. GLCM is used for texture analysis which is formed from an image on the pixels in pairs with a certain intensity in searching for traces of direction. The texture of the image is a regularly arranged pattern that is formed from the arrangement of pixels. The pixels are related to pixel distance (d) and certain orientation angles (θ). The orientation angle (θ) on the GLCM is 0°, 45°, 90° and 135°. Interpretation in imagery has three basic pattern elements: texture, spectral and contextual features [2]. Figure 1 illustrates the details process of generate symmetrical co-occurrence matrix considering a 4 × 4 image represented with four grey-tone values (from 0 to 3). The illustration using one neighboring pixel (d = 1) with four possible directions as [0,1] for 0°, [−1,1] for 45°, [−1,0] for 90° and [−1,1] for 135° [1].
As an illustration, the neighboring pixel can be chosen to the east (right). One way to represent this relationship is in the form of \([1,0]\) which states that the relationship of two pixels that are horizontally lined with pixels 1 followed by a pixel 0. Based on the composition, the number of groups of pixels that meet that relationship is calculated. This is illustrated in Figure 2 below [2].

The matrix shown in Figure 2 (c) is called the matrix framework. This matrix needs to be processed into a symmetrical matrix by adding the transposition before normalization, values of the GLCM elements need to be normalized so that the amount is 1.

Naive Bayes is a classification with probability and statistical methods proposed by British scientist, Thomas Bayes [3] then perfected by Laplace [4]. The Bayes theorem is a statistical calculation by calculating the probability of similarity of an old case with a new case. Naive Bayes is the simplest Bayes theorem calculation that has a high level of accuracy and low computation time when applied to large databases [3]. Naive Bayes is a classification technique whose attributes are mutually independent or there is no connection between attributes [3]. The functions of Naive Bayes can reduce computational complexity into simple multiplications of probabilities and can handle data sets that have many attributes. Naive Bayes can be made to make a decision if it has the initial data in the form of training data in the classification stage which is calculated the highest probability value to be labeled the input data class [3]. The flow of the Naive Bayes algorithm is as follows [5]:
The Naïve Bayes probability equation can be calculated using the following equation [3]:

$$P(X|Y) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$  \hspace{1cm} (1)

Where $P(X|Y)$ is probability value, $\pi = 3.14$ , $\sigma$ is deviation standard, $\mu$ is mean parameter, $e$ is exponential number (2.71828 ...), and $x$ is parameter value.

2. Design System
The system designed in this study shown in the following flowchart:
Data retrieval or image acquisition is obtained from samples of land cover images. Sampling of land cover images has a size of 3000x4496 pixels with the total image of 60 images divided into two parts, namely 30 training data and 30 test data. Image retrieval using a digital camera and saved in *.jpg format. Pre-processing is the stage for equalizing the image size (resize) and covert from three dimension Red Green Blue (RGB) images into grayscale to improve image quality.

Feature extraction uses the GLCM method with second order parameters. Feature extraction is done by taking images that have gone through the pre-processing stage to obtain important information or features obtained from the image for further analysis. Feature extraction will be stored in the database then entered into the classification process by Naive Bayes in the testing process.

![Feature Extraction using GLCM](image)

**Figure 5.** Feature Extraction using GLCM

### 3. Experiments and Result

The results of testing 150 test images consisting of 30 plantation images, 30 forest images, 30 residential images, 30 rice field images and 30 river images.

| Image Size | Accuracy (%) |
|------------|--------------|
| 256 x 256  | 73.3         |
| 512 x 512  | 58.6         |

Table 1 shown that the highest accuracy result is 73.3%, when the image size used is 256 x 256. The smaller the image size, the higher the accuracy. This is caused when the image size is small then the characteristics generated by each image are clearer.
Table 2. The result of 1 statistical feature testing

| Feature   | 0°  | 45° | 90° | 135° |
|-----------|-----|-----|-----|------|
| Contrast  | 80,66667 | 68,66667 | 64 | 62 |
| Correlation | 71,33333 | 67,33333 | 70 | 72 |
| Energy    | 77,33333 | 72,66667 | 74 | 66 |
| Homogeneity | 66 | 78,66667 | 74 | 68,66667 |
| Entropy   | 78 | 84 | 69,33333 | 64 |

Based on Table 2, it can be concluded that the average accuracy is generated when the entropy feature is used. This indicates that the characteristics belonging to each type of data are heterogeneous so that the entropy value can be used to differentiate each class.

Table 3. The result of 2 statistical feature testing

| Feature, Feature | 0°  | 45° | 90° | 135° | Average |
|------------------|-----|-----|-----|------|---------|
| Contrast, energy | 81,33333 | 74,66667 | 74,66667 | 78,66667 | 77,33333 |
| Contrast, homogeneity | 83,33333 | 76 | 75,33333 | 75,33333 | 77,5 |
| Correlation, entropy | 90 | 81,33333 | 74,66667 | 79,33333 | 81,33333 |
| Contrast, correlation | 65,33333 | 80 | 76,66667 | 74 | 74 |
| Contrast, energy | 84,66667 | 80 | 84,66667 | 75,33333 | 81,16667 |
| Correlation, homogeneity | 70,66667 | 60,66667 | 81,33333 | 77,33333 | 72,5 |
| Correlation, entropy | 64 | 81,33333 | 78 | 81,33333 | 76,16667 |
| Energy, homogeneity | 65,33333 | 70 | 77,33333 | 81,33333 | 73,5 |
| Energy, entropy | 70 | 81,33333 | 83,33333 | 72 | 76,66667 |
| Homogeneity, entropy | 81,33333 | 76 | 72 | 60 | 72,33333 |

Meanwhile the correlation and entropy values can significantly distinguish these types of classes as shown in Table 3.

Table 4. The result of quantization level testing

| Quantization Level | Accuration (%) |
|--------------------|----------------|
| 8                  | 75,33333333    |
| 16                 | 78             |
| 32                 | 90,666666667   |

As shown in Table 4, the highest accuracy is generated when the quantization level used is 32. The greater quantization level the smoother the image is processed so that the resulting features are sharpen.

4. Conclusion
The Grey Level Co-Occurrence Matrix (GLCM) method can be used as a feature extraction method in terms of land cover classification. The best classification results are obtained when two statistical features are used, namely correlation and entropy. This indicates that the characteristics belonging to each type of data are heterogeneous so that the correlation and entropy values can significantly distinguish these types of classes. Based on the testing data which is consist of 150 images we obtained the best accuracy is 85% with 206.6715 seconds computation time.
References

[1] B. Pathak and D. Barooah, 2013, “Texture Analysis Based on Thegray-Level Co-Occurrencematrix Considering Possibleorientations,” Int. J. Adv. Res. Electr. Electron. Instrum. Eng., vol. 2, no. 9, pp. 4206–4212.

[2] M. Herry and A. M. Purnomo, 2010, Konsep Pengolahan Citra Digital Dan Ekstraksi Fitur. Yogyakarta: Graha Ilmu.

[3] P. N. Andono, T. Sutojo, and Muljono, 2010, Pengolahan Citra Digital. Yogyakarta: Andi, 2017.

[4] D. Putra, Pengolahan Citra Digital. Yogyakarta: Andi.

[5] A Gebejes, E. M. Master, and a Samples, 2013, “Texture Characterization based on Grey-Level Co-occurrence Matrix,” Conf. Informatics Manag. Sci., pp. 375–378.

[6] Sesnsuse,D.S, 2017,"Perbandingan Algoritma Klasifikasi Naïve Bayes, Nearest Neighbor dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian.