Indonesian Batik Image Classification Using Statistical Texture Feature Extraction Gray Level Co-occurrence Matrix (GLCM) and Learning Vector Quantization (LVQ)

Nafik'ah Yunari[1] , Eko Mulyanto Yuniarno[1,2] and Mauridhi Hery Purnomo[1,2]
[1]*Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia.*
[2]*Department of Computer Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia.*
yunari15@mhs.ee.its.ac.id

***Abstract*—Batik, as a cultural heritage from Indonesia, has two kinds of true batik, batik tulis or handwritten batik and batik cap or stamped batik. However, it is still difficult to discern what is really batik and which are not. So many clothes are claimed as batik whereas they aren't. These pseudo batiks have spread on the market and become deceitful. The objective of this research is to obtain the pattern value of batik to recognize the real batik using feature texture extraction with Gray Level Co-occurrence Matrix (GLCM) as a method for extracting textural features and Artificial Neural Network Learning Vector Quantization (LVQ) as a method to classify. The result of this study shows the average accuracy of data training without normalization of 90.43% and data with normalization of 98.40%. While on data testing the average value of accuracy on the dataset without normalization of 92.79% and after normalization the average value is 98.98%, so the increase in the average value of accuracy of 8%.**

***Index Terms*—Batik; GLCM; Image; Indonesian Batik; LVQ; Texture Feature.**

I. INTRODUCTION

Batik is a technique of decorating textiles Indonesian's culture nuance ornaments and colors through wax resist dyeing, in which designs are created by preventing specific parts of a textile from being exposed to dye by a special tool called Canting [1]. There are two kinds of canting; canting tulis or handwritten canting and canting cap or stamp canting. A cloth made by using this technique is also called "batik". Thus, a cloth may be called batik when it has involved these two basic elements; a wax resist dyeing by Canting and containing Indonesian's culture nuance ornaments.

There are two kinds of real batik; Batik tulis or handwritten batik and batik cap or stamped batik. However, it is still difficult to discern which really batik is and which is not. So many clothes are claimed as batik whereas they aren't. These pseudo batiks have spread on the market and become deceitful. These textiles are called *kain motif batik*.

Several efforts marking real batik to distinguish batik from the kind of pseudo batik are taken by shareholders. But, they are inapplicable because not only are those efforts complicated but they also are not standardized. Batik has specific pattern which is different from other fabrics. It can be recognized. Several studies have been proposed for batik image retrieval, batik image classification and batik pattern recognition. Minarno et al. [2] applied a characteristic extraction method called microstructure descriptor to retrieve Batik images based on their content. On their research, Nurhaida et al. [3] also asserted that Gray Level Co-occurrence Matrices (GLCM) is the best characteristic extraction method to recognize Batik's image, compared to Canny Edge Detection and Gabor filters. Teny [4] applied characteristic extraction of Gray Level Co-occurrence Matrices (GLCM) and VF15 to classify batik Lasem. Many proposed batik image study has been performed using image features such as color, shape and texture [5,7,8,9]. In order to obtain the most influential feature of classifying batik image process is the main object of this research.

As part of the artificial neural network method, Learning Vector Quantization (LVQ) has been widely used to recognize image, sound, and other patterns [10]. The initial pattern recognition process also applied LVQ networks to extract characteristic patterns. LVQ Neural Network Methods including Supervised Learning is determining the weight/learning model that establish result during the process. Many classifications using LVQ has been done on [11]. Simplicity, flexibility and efficiency are some of the reasons why LVQ introduced by Kohonen, is used in various fields [6]. The weakness of this method is highly dependent or sensitive to the selection of reference vectors that are also used as network weights in LVQ training processes. The reference vector is representative of each class representing a class in the LVQ network.

The easiest and most commonly used determination of reference vectors is directly selecting a number of input vectors as representatives of each class in the training process [6]. Initialization of the reference vector (initial weight) in this way is very sensitive to the degree of accuracy so the inaccuracy in its election can result in poor accuracy [12]. Another alternative for the determination of reference vectors is by selecting statistical features using GLCM.

II. BATIK

Traditional batiks, defined as the real original batik, are batik tulis (handwritten batik) and batik cap (stamped batik). The process of making them are handmade. Hot wax is applied by using canting on the cloth. The handwritten applied wax produces a typical pattern motif. This motif will be the main feature object of this research.

Despite various batik motives, their patterns can be recognized because of their different making techniques. According to SNI (Indonesian National Standard) 0239: 2014, Batik - Definition and Terminology [13]; Batik is a craft as staining results in counteraction using night (wax batik) color barrier with the heat as the main tool in the form of adhesive wax batik canting write or stamp to establish a specific motive that has meaning.

The development of industrial batik generates modification process of making batik, either raw materials or equipment used. Many manufacturers had left traditional methods to increase production capacity and reduce the time of making batik. The impact of these technological developments is the emergence of products that resemble the original batik. These products are generated by several printing techniques to produce very similar to batik characteristic.

According to SNI (Indonesian National Standard) 8184: 2015- Tiruan Batik dan Paduan Tiruan Batik dengan Batik - [13]; Artificial Batik kinds are: color printed imitation batik, unplug color printed imitation batik, cold wax printed imitation batik, combination of colors printed and unplug color printed imitation batik, combined product with imitation batik, color printed imitation product combined with batik, printed color imitation product combined with combination batik, unplug color printed imitation product combined with batik, unplug color printed imitation product combined with batik cap, batik counterfeit printed products. All kinds of imitation batik are not batik.

III. METHODOLOGY

There are 4 stages on worldview on this research. They are (1) Image Acquisition with collection and selection of batik image data. (2) The preprocessing image. (3) Feature texture extraction with GLCM (4) Classification with Learning Vector Quantization.

The Batik data process was performed as follows (Fig.1). First, batik image is collected and selected. Next, the selected image will be preprocessed by making all the color image into grayscale image. It is intended that the image will not be affected by the color RGB / HSV while running the texture feature extraction [14]. In the process of texture feature extraction, the data is translated into seven texture features variables on GLCM. Statistical values of this extraction will be calculated in the next identification process. The calculation results of statistical variables will be the input value to the next process.

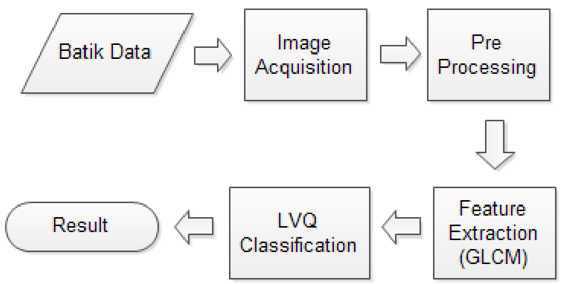

Figure 1: Flowchart

A. Dataset

Batik data used in this research were batik image obtained from image-making batik measuring 2.5 meters x 1 meter using a digital camera Fuji Film mirrorless brands XA2 with specifications total number of pixels: 16.5 million pixels. It is using the standard image capture in accordance with the specification of the camera.

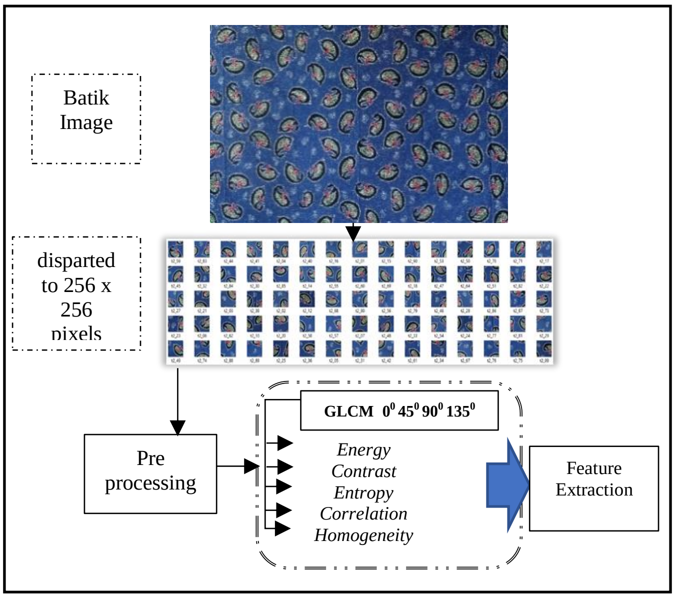

Figure 2: Dataset process

Figure 2, is the complete dataset process. The number of batik data used in the experiment is 30. They consist of 10 batik tulis (handwritten batik), 10 batik cap (stamp batik) 10 textile motif batik. Every batik used in this dataset have a different motif. These collected motives are Tulungagung motives, Solo motives, Yogyakarta motives and Pekalongan motives. Every batik has a various size between 2.3 to 2.5 meters. Every 30 batik image used, is disparted to 256 x 256 pixels that generate 135 pieces. Therefore initial data sets acquired are 4050 images.

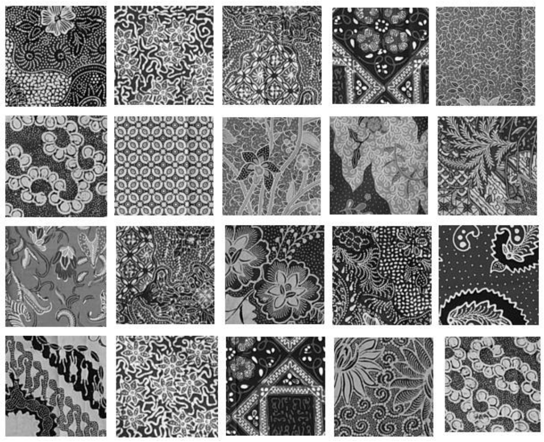

Figure 3: Dataset Image after preprcessing

The data classifying process is done by taking a class based on the type of batik randomly. Each dataset consists of a group class of 20 datasets with the same kind of class. The group of data will be carried out statistical calculations all the features GLCM. The results of grouping the data, as shown in Figure 3, will be analyzed to determine the process for identifying features that affect the image of batik.

B. GLCM Texture Feature

Texture is a basic visual cue, helping the human visual system in segmentation and recognition tasks. According to Haralick et al. [15], Texture is the characteristics possessed by a large enough region that naturally repeated. In this case, The texture is the regularity of certain patterns formed from the arrangement of the pixels in the image. The surface arrangement has a texture if the properties of the expansion result area have a similarity to the original surface after being enlarged without changing the scale. The texture is calculated based on the statistical distribution of the intensity of the pixel to the position between the pixels. The texture is represented in a matrix.

The second-order statistical feature is the Gray Level Co-occurrence Matrix (GLCM). According to this, the second-order Statistics Character is the probability calculation of an adjacency relationship between two pixels at a distance (d) and a specific angle orientation (θ). This approach is applied by forming a Co-occurrence matrix from the image and then determining the characteristics of the matrix. The Co-occurrence matrix is the square matrix with the sum of the elements as much as the square of the sum of the intensity levels. Co-occurrence is a common occurrence, the number of occurrences one level at the neighboring pixel by one level in distance (d) angular angle. Distance in pixels and orientation in degrees. Orientation in four angular directions with a certain distance of 0^0 45^0 90^0 135^0 while the distance between pixels is usually set to 1 pixel

Some textures features for batik motif recognition obtained through the second-level statistics calculation as follows in Table 1.

Table 1
GLCM Statistic Feature

Feature Parameter	Formula
Contrast	$Contrast = \sum_{i,j} \lvert i-j \rvert^2 P(i,j)$
Correlation	$Correlation = \sum_{i,j} \frac{(i-\pi_i)(j-\pi_j)P(i,j)}{\sigma_i \sigma_j}$
Energy	$Energy = \sum_{i,j} P(i,j)^2$
Homogeneity	$Homogeneity = \sum_{i,j} \frac{P(i,j)}{1+\lvert i-j \rvert}$
Maximum Probability	$MP = \max(\mathrm{P}(i,j))$
Inverse Difference Moment	$IDM = \sum_{i,j} \frac{P(i,j)^2}{1+\lvert i-j \rvert}$
Entropy	$e = -\sum_{i,j} P(i,j) \log \mathrm{P(i, j)}$

C. Learning Vector Quantization Network

Classifier used for batik classification in this study is Learning Vector Quantization (LVQ). LVQ is part of Artificial Neural Network (ANN) that uses competitive learning algorithm. Competitive learning algorithms make a set of output neurons on the network to compete into active neurons called winner nueron. The competition process from LVQ is based on Euclidean distance. Only winning neurons are selected (active) only, that is output neurons with minimum Euclidean distance to input data, which is entitled to respond to the network.

LVQ learning process is also done iteratively using the following steps:

- Calculate the value of the learning rate for the t-iteration (αt) based on the decreased learning rate function.
- Select an input data from the training data, then calculate the Euclidean distance from the input data vector to the weighted vector of each output neuron.
- Determine the winning neuron based on the minimum value of Euclidean distance calculated.
- Repair the weighted vector values of the winning neuron.
- Repeat steps until the termination condition of the learning process is reached.

Evaluation process in this research is done to get the best classifier feature and classification in classifying batik obtained by selecting the best parameter values for feature extraction process and classifier development. Classifier performance will also be evaluated classifier performance

IV. EXPERIMENT AND RESULT

We implemented the system on the environment, hardware which is used is a laptop with specification AMD APU A8-7410 quad-core 2,2GHz Turbo 2,4Ghz Processor, Memory RAM 8GB DDR3L, Windows 10 64bit system operation and Matlab 2015a. process statistical methods GLCM using the input dataset in bitmap image. After preprocessing and grouping datasets, the average results obtained in each class.

After calculating the statistical value of each data set, the next process is calculating mean for each class from 5 groups datasets. GLCM statistic calculation results are shown in Table 2.

Table 2
Deviation of mean GLCM feature

Feature	Batik		Non-Batik	
	Min	Max	Min	Max
ASM	0.000691	0.000973	0.000175	0.000529
IDM	0.143136	0.157011	0.069260	0.122417
Entropy	8.738125	8.827925	9.031268	9.353093
Correlation	0.000225	0.000416	0.000370	0.000670

Table 2 shows the results that the features can be used to measure the statistical value of batik are feature of ASM (Angular Second Moment), IDM (Inverse Different Moment), Entropy and correlation. The mean value of written batik class is the highest in ASM (Angular Second Moment) Feature and IDM (Inverse Different Moment) feature. The entropy of written batik is the lowest among the other classes. The correlation value of written batik class is lower than motif textile batik and higher than stamp class batik.

After the feature value of each data set is specified, the next process is classification using the angle parameter GLCM. This parameter measures the test image at every angle used in GLCM feature selection.

Table 3
Testing image in dataset GLCM angles

Classification test	Testing Image			
	GLCM 0	GLCM 45	GLCM 90	GLCM 135
Correct	100%	100%	100%	100%
Incorrect	0%	0%	0%	0%

From Table 3 it is known that batik data at angles 0, 45, 90, and 135 each have the same value as dataset input. So that the input data can be selected at a certain angle or all bias is used.

In the LVQ training process, the value of the Learning Rate is determined first to obtain an optimal value, the smallest error and the shortest period. This value is used for next process. Table 4 shows that learning rate 0.3 has smallest MSE value, the shortest period, and the smallest error in classification.

Table 4
Selection of learning rate

LR	Iteration	HL	MSE	Time	K10 Fold Validation Correct	Incorrect
0.001	1000	40	0.4472	0.08	80%	20%
0.01	1000	40	0.4378	0.08	80.83%	19.16%
0.1	1000	40	0.4082	0.05	83.33%	16.67%
0.3	1000	40	0.3764	0.04	85.83%	14.16%
1	1000	40	0.3764	0.04	85.83%	14.16%

*LR = Learning Rate, HL= Hidden Layer, MSE=Mean Square Error

From Table 4, the learning rate of 0.3 for the smallest MSE, short time and the classification result have little error is applied to determine the value of Learning Rate from the classification of K10 Fold Validation to be done.

Scenarios in the classification test using K10-Fold Validation Method and Hold Out Method (percent split). K-Fold Cross Validation method is one of validation method that aims to find the best rule / model by testing the amount of error in data testing. K-Fold Cross Validation divides the data into k samples of the same size. Then k-1 sample is used as training data, while the remaining 1 sample is used for data testing. For example, if there are 10 datasets, 10-fold Cross Validation will be used, then 10 sets of data will be divided into 2 parts 9 sets of data used as training data, and 1 set of data used as data testing. From the results of these experiments, it can be calculated as the average error. The best Rule / model will have the smallest average error.

The holdout method is one of the validation methods to select the best rule / model, which will be used for classification process in data testing. In this method, data are randomly and independently divided into two sets, one set is used as training data and one set is used as test data. Typically, two-thirds of the data is allocated for training data, and the remaining one-third is allocated for testing data. The training set is used for obtaining the rule / model with the best accuracy, so the rule / model is applied using data testing to obtain the best classification results [16].

Table 5
LVQ K10 fold validation

Scenario	Dataset	Precision (%)	Recall (%)	Accuracy (%)
K fold 10	Dataset 1	92.20	89.83	89.83
	Dataset 2	91.63	91.50	91.50
	Dataset 3	90.34	89.67	89.67
	Dataset 4	89.43	89.00	89.00
	Dataset 5	92.20	92.17	92.17
Average		91.16	90.43	90.43

Table 5 shows the result of the experiment after testing data using K10-Fold Cross-Validation. It shows that the result of testing on the training data, using the original data without any normalization with the original Batik class with an average accuracy of 90.43%. The most optimal combination of LVQ parameters is at Learning Rate 0.03 with epoch 1000 iteration.

Table 6
LVQ percent split 33%

Scenario	Dataset	Precision (%)	Recall (%)	Accuracy (%)
Percent Split	Dataset 1	94.37	94.15	94.17
	Dataset 2	90.28	87.70	87.62
	Dataset 3	92.28	91.68	91.67
	Dataset 4	95.07	94.66	94.67
	Dataset 5	95.90	95.82	95.83
Average		91.16	93.58	92.80

After data training test, the next process is testing the result data that is 33% of the amount data by split percent method. Table 6 shows the results of testing on this process with the original Batik class, using the original data without any normalization with an average accuracy of 92.80 %. The most optimal combination of LVQ parameters at Learning Rate 0.03 with epoch 1000 iteration and has the shortest time of 0.04.

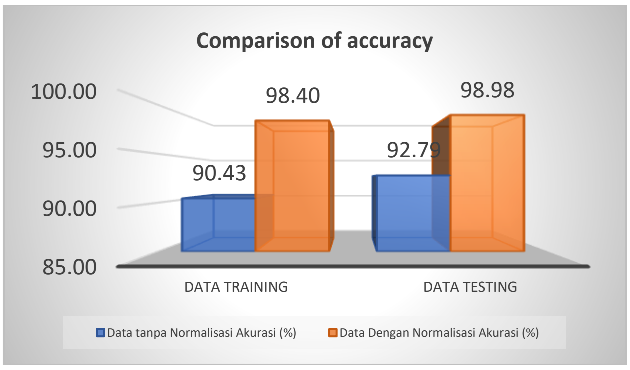

Figure 4: Comparison accuracy training and testing

From the calculation of ROC Table 5 and 6, it can be seen that the results for precision value of LVQ classification and ROC value of 0.95555 which shows satisfying classification.

The next scenario is by comparing the data of the normalization process with the original data without normalization. It aims to measure the difference in the level of accuracy. The results can be seen in figure 4. on the training data using K fold validation data accuracy with normalization of 98.40%. While in data testing data with normalization of the average value of accuracy increased to 98.98% so that the average value of accuracy increase of 8%.

V. Conclusion

The result of this research shows that GLCM features and LVQ can be used to classify batik and non-batik fabrics obtained from batik image processing. GLCM feature selection that can be used to measure statistic value of batik are Entropy feature, ASM (Angular Second Moment) feature, correlation and IDM (Inverse Different Moment) feature. For all angles of GLCM (0, 45, 90, 135) they all share the same weight as the input data for the classification process.

The result of this study shows that the best parameters LVQ by using 1000 epoch, learning rate 0.3 and 600 hidden neurons. The identification stage of 600 samples of batik and non-writing image by applying LVQ classification method is done by 2 stages, the first phase of data training is 70% from the amount of data sample, both stages of testing done at 30%. For LVQ prediction rate reached the average accuracy of data without normalization of 90.43% and data with normalization of 98.40%. While on the test data or data testing the average value of accuracy on the dataset without normalization of 92.79% after normalization of the data the average value of accuracy increased to 98.98%, so the increase in the average value of accuracy of 8%

Hopefully, it can be developed further methods to recognize the type of batik based on its making technique in the next research.

ACKNOWLEDGMENT

The authors would like to thank Central for Handicraft and Batik, Ministry of Industry, Yogyakarta, Indonesia for providing data and their cooperation and to the Ministry of Communication and Information Republic of Indonesia for the master degree scholarship.

REFERENCES

[1] UNESCO, *Indonesian Batik: Inscribed in 20019 on the Representative List of Intagible Cultural Heritage of Humanity*, United Nations. http://www.unesco.org/culture/ich/en/RL/indonesian-batik-00170.

[2] A. E. Munarko, Y. Bimantoro, A. F. Kurniawardhani, and N. Suciati, "Batik image retrieval based on enhanced micro-structure descriptor," in *Computer Aided System Engineering (APCASE), 2014 Asia-Pacific Conference*, 2014, pp. 65-70.

[3] I. Nurhaida, R. Manurung, and A.M. Arymurthy, "Performance Comparison analysis Feature Extraction Methods for Batik Recognition," in *Proc. Int. on Advanced Computer Science and Information System (ICACSIS)*, 2012, pp.207 - 212.

[4] T. Handhayani, "Batik Lasem Images Classification Using Voting Feature Intervals 5 and Statistical Features Selection Approach*". International Seminar on Intelligent Technology and Its Application (ISITIA)*, 2016

[5] A. H. Rangkuti, R. B. Bahaweres and A. Harjoko, (2012, December). "Batik image retrieval based on similarity of shape and texture characteristics," in *Advanced Computer Science and Information Systems (ICACSIS), 2012 International Conference*, 2012, pp. 267-273.

[6] T. Kohonen, R. Schroeder, S. Huang, "Self-Organizing Maps", *Springer Series in Information Sciences*, Volume 3, Berlin Heidelberg Edition: 30, 2001.

[7] R. Pawening, R. Dijaya, T. Brian, and N. Suciati, "Classification of Textile Image using Support Vector Machine with Textural Feature," in *International Conference on Information and Communication Technology and Systems (ICTS),* IEEE, 2015, pp. 163-168.

[8] V. S. Moertini and B. Sitohang, "Algorithms of clustering and classifying batik images based on color, contrast and motif," *Journal of Engineering and Technological Sciences*, 37(2), 2015, pp. 141-160.

[9] A. E. Minarno, Y. Munarko, A. Kurniawardhani, F. Bimantoro and N. Suciati, "Texture Feature Extraction Using Co-Occurance Matrices of Sub-Band Image For Batik image Classification," in *Proc. 2nd Int. Conf. on Information and Communication Technology (ICoICT)*, Bandung, 2014, pp.249-254.

[10] M. Tuceryan, A. K. Jain, "Texture Analysis", *Handbook of Pattern Recognition and Computer Vision (2nd ed.)*, World Scientific Publishing Company, 1998, pp. 207-248.

[11] A. Ghosh, M. Biehl, and B. Hammer, "Performance Analysis of LVQ Algorithm: A Statistical Physics Approach," *Journal Neural Networks – 2006 Special Issue*, Vol 19, 2006, pp. 817-829.

[12] Z. K. Huang, P. W. Li and L. Y. Hou, "Segmentation of textures using PCA fusion based Gray-Level Co-Occurrence Matrix features," in *Test and Measurement, 2009. ICTM'09. International Conference on*, Vol. 1, IEEE, 2009, pp. 103-105.

[13] Standar Nasional Indonesia (SNI), SNI 8184, 2015, Tiruan batik dan paduan tiruan batik dengan batik, pengertian dan istilah. Badan Standardisasi Nasional, Jakarta.

[14] R. Gonzalez and R. Woods, *Digital Image Processing*. New York: Addison-Wesley Publishing Company,Inc, 1992.

[15] R. M. Haralick, K. Shanmugan and I. Dinstein, "Textural Features for Image Classification," in *IEEE Transactions on Systems: Man, and Cybernetics SMC*, Vol. 3, 1973.

[16] R. Remco, F. Eibe, H. Mark, K. Richard, R. Peter, S. Alex, and S. David, *WEKA Manual 3-8-1*, Hamilton, New Zealand: University of Waikato, 2016.