The Selection Feature for Batik Motif Classification with Information Gain Value

Anita Ahmad Kasim[1,2(✉)], Retantyo Wardoyo[1], and Agus Harjoko[1]

[1] Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Yogyakarta, Indonesia
nita.kasim@gmail.com, {rw,aharjoko}@ugm.ac.id
[2] Department of Information Technology, Faculty of Engineering, Universitas Tadulako, Palu, Indonesia

Abstract. Features in the classification process have an important role. Classification of batik motif has been done by using various features such as texture features, shape features, and color features. Features can be utilized separately or combined between features. The problem in this research is how to get the potential feature to classified the motif of batik. The feature combination causes enhancement in the number of features causing dataset size changes in the classification process. In this research will be done the selection of features to the combination feature of texture and feature of shape from batik motifs. The feature selection process uses the information gain value approach. The feature selection is done by calculating the value of the information gain of each feature of texture and feature of shape. The value of the information gain will be sorted from the highest information gain value. Ten features with the highest information value will be the selected feature to be processed in the process of batik image motif classification. The classification method using in this research is an artificial neural network. The neural network consists of three layers, that is input layer, hidden layer, and the output layer. The data from selection feature processed in the artificial neural network. The result of this study shows that the accuracy of the process of batik motif classification with a combination feature of texture and feature of shape is 75%. The addition of feature selection process to batik motif classification process gives an increase of 12.5% to the yield an accuracy of 87.5%.

Keywords: Feature selection · Batik · Information gain

1 Introduction

The number of batik motifs that are owned and spread throughout Indonesia makes Batik become one of the cultural heritage of Indonesia that must be preserved. The designation of Batik into the List of Culture of Human Heritage by the United Nations (UN) make batik as an icon of Indonesian culture [1].

Batik has a variety of motif designs that are influenced by various cultures and mythology in Indonesia [2]. Classification of classical batik motif cannot be done by using the features of the shape and texture features separately so that the two methods are combined in a method of classification based on texture and shape [3]. The feature

A. Mohamed et al. (Eds.): SCDS 2017, CCIS 788, pp. 106–115, 2017.
https://doi.org/10.1007/978-981-10-7242-0_9

of texture and feature of shape can be selected for use in the classification of batik motifs to obtain potential features that affect the accuracy of the classification of each type of batik. To solve the existing problems in the batik motif recognition model, we need a model that can identify the type of batik motif using the artificial neural network with combined features of texture and features of batik ornament shape.

This research is expected to improve the accuracy of the motif classification of batik so that various batik motifs can be more easily recognized. This paper will describe the feature selection process on batik motif classification includes the first part is the introduction, the second part is related works, the third part is the proposed method and the last is the conclusion.

2 Related Works

Batik is one of the oldest Indonesian art and culture. There are many batiks as a local product in all parts of Indonesia from Sabang to Merauke. Batik is a cultural heritage of Indonesia that has been recognized as a UNESCO international cultural heritage on October 2, 2009. Pattern recognition is one way to preserve the culture of batik. The research of batik pattern recognition has been done by some previous researchers [1–4].

The features used in batik pattern recognition research include texture features, shape features and color features [5–8]. The batik texture features used were obtained from sub-band image of the Co-Occurrence Matrices [9]. The image used was obtained by downloading randomly from the internet. Co-occurrence matrices method with sub-band image is done by combining Gray Level Co-occurrence Matrices and Discrete Wavelet Transform method. The results obtained are good enough to classify the image of batik. Maximum accuracy that can be achieved is 72%. The research of batik classification is also done by combining two batik features namely color (3D-Vector Quantization) and shape feature (Hu Moment) [10]. HMTSeg is used by dividing the image of batik into several areas based on similarity of features including levels of gray, texture, color, and movement [11]. The accuracy obtained by 80% can recognize the image of batik based on texture features. In addition to the use of texture features, batik imagery can also be classified by combining texture features and color stats features. Texture features extracted from Gray Level Co-Occurrence Matrices (GLCM) consisting of contrast, correlation, energy, and homogeneity. The color features are extracted with the Statistical Color Channel RGB which consists of mean, skewness, and kurtosis [12, 13]. The feature combined with the Bag of Features (BOF) method uses Scale-Invariant Feature Transform (SIFT) and Support Vector Machine (SVM) to classify batik images The results obtained are 90.6% precision, 94% recall and 94% overall accuracy [14, 15]. The acquisition of image data in one class comes from one fabric image divided into six sub-images.

The combination of texture features and shape features are also done for batik image classification. The shape features obtained feature values include compactness, eccentricity, rectangularity, and solidity [16]. The combination of features causes the number of features to be more numerous, so in this study proposed a method for selecting features of texture and features of batik shape that potential to improve the accuracy of the batik classification.

3 Proposed Method

The methods we proposed in this study consisted of the acquisition process, the pre-process, the feature extraction process, the selection feature process, and the image classification process. The classification process using the artificial neural network. The role of an artificial neural network as one of the classification method which is quite reliable due to its ability to classify the data. The artificial neural network has a high tolerance for noise-sensitive data. The neural network is able to learn from the data trained. Therefore this research analysis data classification by using ANN Backpropagation to get an accurate result.

3.1 Acquisition and Preprocess of Batik Image

The acquisition process is the process of taking the image of batik using the camera and download some images of batik through the internet. The pre-processing will process the acquired image before it is processed to obtain texture feature values and feature shapes values.

At the pre-processing process, we formed a sub-image with the image size of 100×100 pixels. Next, an image transformation to the gray scale is performed. The grayscale transformation will further facilitate the computing in obtaining the values of batik texture features.

3.2 Feature Extraction Process

Grayscale batik image produced in the process of acquisition and preprocessing into an input on feature extraction process. The texture feature extraction method used begins with the formation of a GLCM matrix from a grayscale image. The image of the normalized result is Gray Level Co-occurrence Matrices which will be used to obtain texture features values. Values obtained in GLCM are processed into textural features consist of the value of features of Angular Second Moment (ASM) or otherwise known as Energy, Contrast, Correlation, and Homogeneity [17].

The transformation from the grayscale image into a binary image by using Otsu Method. It will eliminate the noise in a binary image by using the binary morphology [18–20]. The shape features consist of the compactness, eccentricity, rectangularity, and solidity.

3.3 Feature Selection Process

The feature selection model developed in this study uses the information gain's value approach. The steps to obtain information gain are as follows:

1. Convert continuous values into discrete values by partitioning continuous values into discrete value intervals. The process of converting the continuous values into discrete values as described [21] consists of three steps as follows:
   (a) Sort the continuous values that will be made in discrete values

(b) A cut-point evaluation for dividing continuous values. The definition of cut-point refers to the real value of continuous values that divide the interval for instance into two intervals. In this researches the number of cut-points on the feature there are two cut-points, and there are three cut-points.

(c) Split the continuous values in their respective intervals into discrete values. The process of converting the continuous values to discrete values is shown in Fig. 1.

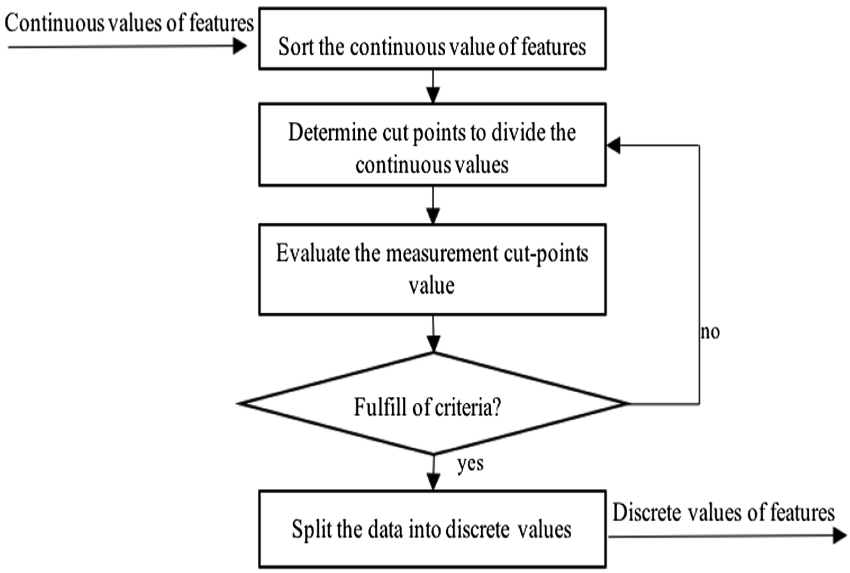

Fig. 1. Discretize process from continous data of features

2. Calculate the value of the expected information or entropy of the whole class by representing it as in Eq. (1)

$$J = (U, C \cup D) \tag{1}$$

where U denotes a set of data consisting of $\{u_1, u_2, \ldots, u_s\}$, C is a feature to be ranked consisting of $\{C_1, C_2, \ldots, C_n\}$ and D is the expected target class consisting of $\{d_1, d_2, \ldots, d_m\}$. If the expected target class has different values of m to distinguish each target d_i for i = 1, 2, …, m then the amount of data included in the class. To calculate the expected value of the information or the required entropy we used Eq. (2)

$$I(s_1, \ldots, s_m) = -\sum_{i=1}^{m} P_i log_2 P_i \tag{2}$$

where $I(s_1,\ldots,s_m)$ denotes the value of the expected information for the entire data. Pi is denoted the probability of sample entering in class and m is number of labels each feature on class.

3. Calculate the entropy value of each feature
Suppose c_i is a feature to be calculated the value of the information and it has a different value v to distinguish each value in the feature, then the feature c_i can be grouped into a set S_j that consisting $\{S_1,S_2,\ldots,S_v\}$. If S_j is the amount of data d_i in a class on a subset of sets S_j, then the feature entropy value can be obtained by Eq. (3).

$$E(c_i) = \sum_{j=1}^{v} \frac{s_{ij} + \ldots + s_{mj}}{s} I(s) \tag{3}$$

where $E(c_i)$ denotes the value of feature entropy and v value in feature c_i. S_{ij} is a number of sample classes in class S_j.

4. Calculate the information gain value of each texture feature and shape feature with Eq. (4)

$$Gain(c_i) = I(s_1,\ldots,s_m) - E(c_i) \tag{4}$$

5. The value of the information gain will be sorted and will be selected for feature data by taking only 10 data with the highest information gain value.

The texture-shape feature selection is done by calculating the entropy value to generate the information gain value of the texture feature and the shape features. The value of the information gain will reflect the feature's quality.

3.4 Classification Process

Motif classification model on batik image based on the feature of texture and feature of shape using neural network method using backpropagation learning method. Backpropagation is a gradient decrease method to minimize the square of output errors or algorithms that use weight adjustment patterns to achieve minimum error values [22]. Artificial neural network in the training process is the number of neurons in the input layer, the hidden layer, the output layer. The number of neurons in the input layer is determined according to the number of inputs on the artificial neural network architecture. At the beginning of the trial will be training by combining 20 features of texture and the feature of shape. For the selection, results will be selected ten features tested. The determination of the number of neurons in the hidden layer is determined using the equation described [23]. For the number of neurons in the output, the layer is tested with two outputs, i.e., three output neurons for eight classes of batik imagery and eight output neurons for eight classes of batik images. Artificial neural network architecture for the motif classification of batik image can be seen in Figs. 2 and 3.

The number of neurons in each layer used in the artificial neural network architecture in this study is shown in Table 1.

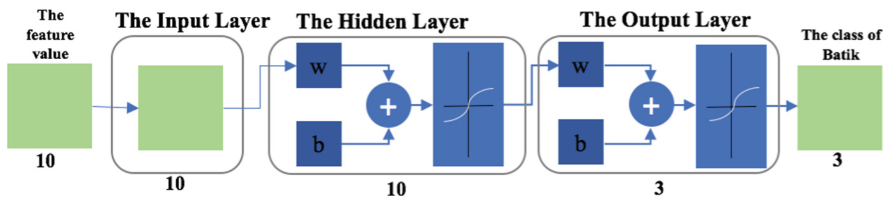

Fig. 2. Artificial neural network architecture for the motif classification of batik image with 3 outputs

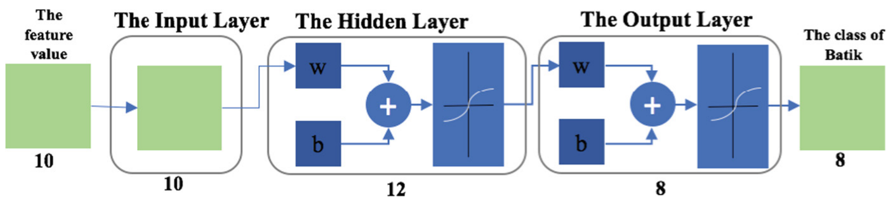

Fig. 3. Artificial neural network architecture for the motif classification of batik image with 8 outputs

Table 1. Architectural number of neurons in batik motif recognition model

Features	Input layer	Hidden layer	Output layer	NN architecture
Texture-shape	20	14	3	20:14:3
Texture-shape	20	17	8	20:17:8
Texture-shape based on feature selection	10	10	3	10:10:3
Texture-shape based on feature selection	10	12	8	10:12:8

4 Results and Discussions

The acquisition process generates images data batik with eight motif class. The motif class consists of *Ceplok* class, *Kawung* class, *Megamendung* class, parang class, *Semen* class, *Batik solo* class, *Sido asih* class and *Tambal* class. The texture's feature extraction process generates the values of ASM/Energy, the contrast value, the correlation value, and the homogeneity values of batik images. Furthermore, the values of ASM/Energy, the contrast value, the correlation value, and the homogeneity values are normalized. The minimum values and the maximum values of the normalization result from batik images can be seen in Fig. 4.

The feature extraction from the shapes generates the compactness value, the eccentricity value, the rectangularity value and the solidity value of batik images. The minimum value of the compactness value is 15.55, and the maximum value is 7296.19.

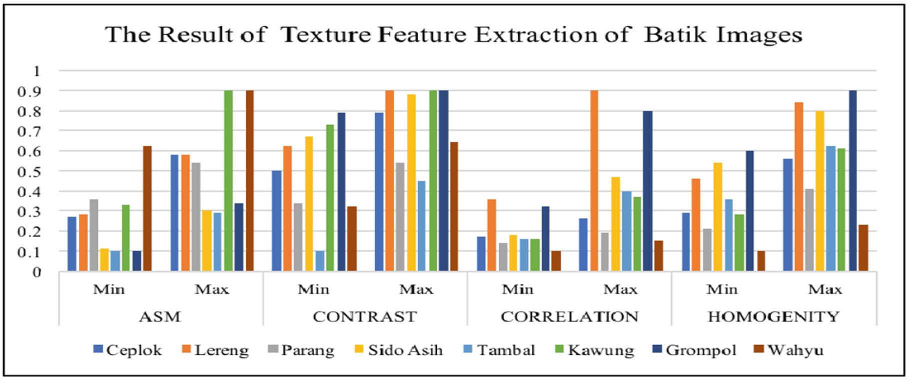

Fig. 4. Minimum value and maximum value of batik texture features

The value of eccentricity feature in the range of 1.22 to 4.08. The rectangularity values in the range of 7.62 to 1001.37 and features of solidity in the range of 0.35 to 0.86.

Figure 5 illustrates the minimum value and the maximum value of the Batik image shape feature.

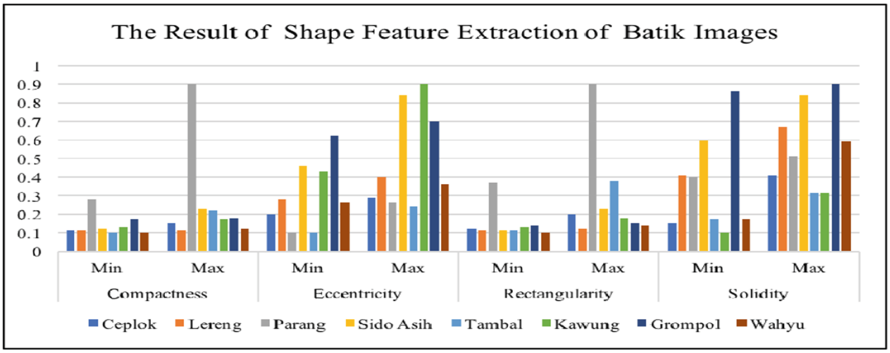

Fig. 5. Minimum value and maximum value of batik shape features

The processing of feature selection on the batik motif classification model is done by first making changes to the continuous values of the extraction result into discrete values. For each feature categorized the grade as shown in Table 2. Based on the selection model proposed in this study obtained the gain value of the texture features and the shape feature.

The resulting feature value generated ranks include the correlation 90°, eccentricity, rectangularity, correlation 135°, compactness, contrast 135°, ASM 90°, ASM 0°, contrast 90°, ASM 45°. The result of gain value ranking is shown in Table 3.

Table 2. The conversion results of continuous values into discrete values

No	Features	Discrete value		
1	Angular second moment	High	Middle	Low
2	Contrast	High	Middle	Low
3	Correlation	High	Middle	Low
4	Invers different moment	Uniform	Not uniform	
5	Compactness	Compact	Not compact	
6	Eccentricity	High	Middle	Low
7	Rectangularity	High	Low	
8	Solidity	Solid	Not Solid	

Table 3. Gain value ranking results on texture features and shape features of motif batik image

Number	Selected features	Gain value
1	correlation 90°	5.254894387
2	Eccentricity	5.157481437
3	Rectangularity	5.138537962
4	correlation 135°	4.769167528
5	Compactness	4.614079875
6	contrast 135°	3.901570398
7	ASM 90°	3.606168067
8	ASM 0°	3.581783311
9	contrast 90°	3.457986592
10	ASM 45°	3.183023418

The top ten of the rated features will be used as a combined feature of the texture feature and shape feature. Ten features will be used in the process of batik image classification using the artificial neural network.

In this study, we used two datasets. The first data set uses twenty features from the combination of texture features and shapes feature. The second data set uses the ten features of the recommended feature selection in this study. The training was conducted with two different neural network architectures. The training was performed on 64 data of batik image features. Forty data were used as training data, and twenty-four data were used as test data. The data can be classified correctly on the first test of a total of 15 data with an accuracy of 62.5%. The use of different neural network architectures yields the correct amount of data classified to 75%. The results of training and testing motif batik classification can be seen in Table 4.

The tests performed on the data using the selected selection feature with ten input neuron layer architecture, twelve hidden layer neurons and eight neurons on the output layer resulted in an accuracy of 87.5%.

There is an increase in accuracy of 12.5% of the classification results by using features from the selection feature process. The accuracy is lower than previous research. Data in the previous research was different. The amount of data in each class

Table 4. The results of training and testing motif batik classification

Dataset	Features	NN architecture	Training data	Testing data	True data	Accuracy (%)
I	Texture-shape	20:14:3	40	24	15	62.50
I	Texture-shape	20:17:8	40	24	18	75.00
II	Texture-shape based on feature selection	10:10:3	40	24	11	45.83
II	Texture-shape based on feature selection	10:12:8	40	24	21	87.50

is less, too. It is expected that in subsequent research can be used larger data so that the accuracy of classification can be increased.

5 Conclusion

The number of neurons in the neural network architecture and the number of features in the process of batik motif classification will affect the results of accuracy. There is an improvement in the accuracy of batik motif classification model with the artificial neural network from 75% (without feature selection process) to 87.5% (with feature selection process). The addition of the feature selection process on the batik motif classification model gives an increase the accuracy of 12.5%. The future works of this research develop the feature selection method with the other method. The other methods are expected to yield potential feature which will increase the accuracy of batik motif classification.

References

1. Imanudin: “Batik Identification Based On Batik Pattern And Characteristics Using Fabric Pattern Feature Extraction” (2010)
2. Kitipong, A., Rueangsirasak, W., Chaisricharoen, R.: Classification System for Traditional Textile : Case Study of the Batik. In: 13th International Symposium on Communications and Information Technologies (ISCIT) Classification, pp. 767–771 (2013)
3. Kasim, A.A., Wardoyo, R.: Batik image classification rule extraction using fuzzy decision tree. In: Information Systems International Conference (ISICO), pp. 2–4, December 2013
4. Nurhaida, I., Noviyanto, A., Manurung, R., Arymurthy, A.M.: Automatic Indonesian’s batik pattern recognition using sift approach. Procedia Comput. Sci. **59**, 567–576 (2015)
5. Kasim, A.A., Wardoyo, R., Harjoko, A.: Feature extraction methods for batik pattern recognition: a review. In: AIP Conference Proceedings, vol. 70008, pp. 1–8 (2016)
6. Suciati, N., Pratomo, W.A., Purwitasari, D.: Batik motif classification using color-texture-based feature extraction and backpropagation neural network. In: IIAI 3rd International Conference on Advanced Applied Informatics, pp. 517–521 (2014)
7. Kasim, A.A., Harjoko, A.: Klasifikasi citra batik menggunakan jaringan syaraf tiruan berdasarkan gray level co-occurrence matrices (GLCM). In: Seminar Nasional Aplikasi Teknologi Informasi (SNATI), pp. 7–13 (2014)

8. Kasim, A.A., Wardoyo, R., Harjoko, A.: Fuzzy C means for image batik clustering based on spatial features. Int. J. Comput. Appl. **117**(2), 1–4 (2015)
9. Minarno, A.E., Munarko, Y., Kurniawardhani, A., Bimantoro, F., Suciati, N.: Texture feature extraction using co-occurrence matrices of sub-band image for batik image classification. In: 2nd International Conference on Information and Communication Technology (ICoICT) Texture, pp. 249–254 (2014)
10. Nugrowati, A.D., Barakbah, A.R., Ramadijanti, N., Setiowati, Y.: Batik image search system with extracted combination of color and shape features. In: International Conference on Imaging and Printing Technologies (2014)
11. Murinto, Ariwibowo, E.: Image segmentation using hidden markov tree methods in recognizing motif of batik. J. Theor. Appl. Inf. Technol. **85**(1), 27–33 (2016)
12. Aditya, C.S.K., Hani'ah, M., Bintana, R.R., Suciati, N.: Batik classification using neural network with gray level co-occurence matrix and statistical color feature extraction. In: 2015 International Conference on International, Communication Technology and System (ICTS), pp. 163–168 (2015)
13. Rao, C.N., Sastry, S.S., Mallika, K., Tiong, H.S., Mahalakshmi, K.B.: Co-occurrence matrix and its statistical features as an approach for identification of phase transitions of mesogens. Int. J. Innov. Res. Sci. Eng. Technol. **2**(9), 4531–4538 (2013)
14. Azhar, R., Tuwohingide, D., Kamudi, D., Sarimuddin, Suciati, N.: Batik image classification using sift feature extraction, bag of features and support vector machine. Procedia Comput. Sci. **72**, 24–30 (2015)
15. Setyawan, I., Timotius, I.K., Kalvin, M.: Automatic batik motifs classification using various combinations of sift features moments and k-nearest neighbor. In: 7th International Conference on Information Technology and Electrical Engineering (ICITEE), Chiang Mai, Thailand, vol. 3, pp. 269–274 (2015)
16. Kasim, A.A., Wardoyo, R., Harjoko, A.: Batik classification with artificial neural network based on texture-shape feature of main ornament. Int. J. Intell. Syst. Appl. **9**, 55–65 (2017)
17. Haralick, R., Shanmugan, K., Dinstein, I.: Textural features for image classification. IEEE Trans. Syst. Man Cybern. **3**, 610–621 (1973)
18. Bhagava, N., Kumawat, A., Bhargava, R.: Threshold and binarization for document image analysis using otsu' s Algorithm. Int. J. Comput. Trends Technol. **17**(5), 272–275 (2014)
19. Otsu, N.: A threshold selection method from gray-level histogram. IEEE Trans. Syst. Man Cybern. **20**(1), 62–66 (1979)
20. Dong, L., Yu, G., Ogunbona, P., Li, W.: An efficient iterative algorithm for image thresholding. Pattern Recognit. Lett. **29**(9), 1311–1316 (2008)
21. Liu, H., Hussain, F.,Tan, C.L., Dash, M.: Discretization: An Enabling Technique (2001)
22. Fausett, L.: Fundamental of Neural Network. Prentice Hall, New Jersey (1993)
23. Stathakis, D.: How many hidden layers and nodes? Int. J. Remote Sens. **30**(8), 2133–2147 (2009)