Fuzzy C Means for Image Batik Clustering based on Spatial Features

Anita Ahmad Kasim
Doctoral Program Department Of Computer Science and Electronics Gadjah Mada University Yogyakarta, Indonesia

Retantyo Wardoyo, Ph.D
Department Of Computer Science and Electronics Gadjah Mada University Yogyakarta, Indonesia

Agus Harjoko., Ph.D
Department Of Computer Science and Electronics Gadjah Mada University Yogyakarta, Indonesia

ABSTRACT
Patterns of batik produce geometric shapes unique, the number and name of the batik patterns make it difficult to recognize each motif. Spatial information is an important aspect of image processing such as computer vision and recognition structure / pattern in the context of modeling and resolution of the uncertainty caused by the ambiguity in the low-level features. Batik motifs are very diverse produce values that are sometimes not obvious feature class. There is a feature that is included in the value of two different classes. This causes an error in the process of cluster motif. The problems described above can be formulated problem in this research is how to build a cluster of batik images using fuzzy C means cluster based on spatial features. From the results of experiments that have been done obtained Fuzzy C Means able to build a cluster of batik based on the spatial characteristics of contrast, energy, correlation and homogeneity. Cluster obtained seen that the use of FCM in the cluster image Batik produces batik image data distribution in cluster grouping characteristic obvious visible difference.

General Terms
Fuzzy C Means, Image, Feature

Keywords
Cluster, Batik, Spatial Feature

1. INTRODUCTION
Patterns of batik produce geometric shapes unique. The number and name of the batik patterns make it difficult to recognize each motif. Cluster data is needed to identify the characteristics of the objects contained in a database and categorized into different groups [1]. The purpose of batik cluster is split image into classes according to the pattern motifs. This cluster is expected to make easy the recognition of each batik pattern. Batik can be clustered based on geometric motifs form of motive, non-geometric motifs and specific motives.

Spatial information is an important aspect of image processing such as computer vision and recognition structure / pattern in the context of modeling and resolution of the uncertainty caused by the ambiguity in the low-level features. The traditional method for image retrieval based on text indexing is not able to meet the needs of the massive image search. The mapping between the image's vision and image's semantic feature information has trouble, ambiguity (vagueness), flexible and has a weakness. The value attribute vague image (vague) can't be described accurately and clearly, while the extraction of image attributes are ambiguous [2]. Batik motifs are very diverse produce values that are sometimes not obvious feature class. There is a feature that is included in the value of two different classes. This causes an error in the process of cluster motif. The problems described above can be formulated problem in this research is how to build a cluster of batik images using fuzzy C means cluster based on spatial features.

2. LITERATURE REVIEW
Research on batik pattern recognition have been done by using the Rotate Wavelet Filter and combine it with artificial neural networks [3]. The purpose of these studies to inventory and identify the data batik batik based motif. Rotate Wavelet Filter is used for feature extraction batik will be recognized and artificial neural network is used to create a cluster image based on the motive. Wavelet is used to define the multi-resolution produces energy features and standard deviation with the image size of 128 x 128 pixels. The results of image recognition of batik using neural networks have batik image recognition accuracy of 79%. Research on batik by adding a discrete Wavelet Transform features and features of Grey Level Co Occurrence Matrix on Wavelet Filter Rotate features and clusters using Fuzzy C Means (FCM). The experiment was carried out several times on different images to get an idea of the quality of the use of Fuzzy C Means in recognition motif multi label. From these experiments conclusion that the Fuzzy C Means method could reasonably be used to identify multi-label batik [4]. Batik research by utilizing the features of Grey Level Co Occurrence Matrix also been carried out by using artificial neural network to cluster batik image. Classification by using the backpropagation algorithm on neural network obtained results for the ability of neural networks in clusters of 2 pieces of batik image in the non geometry and geometry of 95.7% on the training data. Non-geometric clusters 37 (91.9%) characteristics batik image there are 3 (8.1%) characteristics that can't cluster into class. As for the image characteristic geometry of all the features in the cluster image can be correctly (100%). For testing process there are two features (18.2%) that can't cluster and on the validity test all the features can form a good cluster [5].In clusters needed clear rules in determining the image of each class. Rules cluster batik image can be obtained with the fuzzy decision tree [6]. Fuzzy cluster rules are applied to the fuzzy cluster using fuzzy decision tree can build a cluster batik image. The result is a cluster of rules that can form a cluster batik pattern. There were 20 percent of the data that can't be clustered with either, although the type of fabric has the same pattern class. Research on batik pattern recognition has been done to make clusters and cluster batik image based on color, contrast and motif [7]. The method used to perform cluster and the cluster is the cluster system HVS color and wavelet systems in the cluster contrast, texture and shape. The results obtained in clusters rather good color and contrast. Three clusters and cluster process based motif pretty good. The level of accuracy obtained is still to be repaired. To improve the accuracy of pattern recognition, research that is

capable of forming clusters classical batik automatically into sub-classes based on the ornament shape and texture [8]. Cluster-based image content using Daubechies wavelet transform with type 2 level 2 to process the characteristic texture consisting of a standard deviation, mean, and variable energy input, by using Fuzzy Neural Network (FNN). All input values will be processed using the fuzzification with 5 categories: Very Low (VL), Low (L), Medium (M), high (H) and very High (VH). The results will feed into the process of fuzzy cluster neural network method. Batik image will be processed using the 7 (seven) types of batik motifs fried, kawung, Slope, machetes, megamendung, patched and nitik. The result of the cluster using the FNN is Rule generation, so for a new image of batik motifs can be determined immediately after cluster FNN is completed. For this level of precision, this method is between 90-92%, including use the rules to determine the precise level of generation between 90-92% [9].

3. RESEARCH METHOD

The methods used in this study include batik image acquisition, preprocessing, feature extraction and spatial cluster process images using Fuzzy C Means. The following are the stages in this study:

3.1 Image Acquisition Batik

This stage Collect batik image of some motifs and different types. Motif used in this study is Grompol, kawung, nitik, machetes, sidomukti, slobog, revelation tumurun, Tirta teja, Sidoasih and klitik. Each Batik image kept on file with .jpg image format.

3.2 Preprocessing

The preprocessing performed sampling on certain parts of the object (cropping) with a size of 300 x 300 pixels. Furthermore, the change of color images to gray scale. Changes into a gray scale image can be seen in Figure 1.

3.3 Feature Extraction

Characteristics that are used are obtained from the characteristic spatial relationships between the pixels that form of Grey Level Co-occurrence Matrices (GLCM). Co-occurrence of significant events together, ie the number of occurrences of the level of neighboring pixel values with one another pixel value level in the distance (d) and the orientation angle (θ) specific. Haralick proposed 14 characteristics that can be used to build clusters texture. Features that will be used in this study include contrast, energy, homogeneity and correlation obtained through Equation (1) to the equation (4) [10].

$$contrast = \sum_{a,b} P_{a,b}(a-b)^2 \tag{1}$$

$$Energy = \sum_{a,b} P_{\theta,d}^2\,(a,b) \tag{2}$$

$$homogeneity = \sum_a \sum_b \frac{1}{1+(a-b)^2} P_{\theta,d}(a,b) \tag{3}$$

$$correlation = \frac{\sum_{a,b}[(ab)P_{\theta,d}(ab)]-\mu_x\mu_y}{\sigma_x\sigma_y} \tag{4}$$

Dengan

$$\mu_x = \sum_a a \sum_b P_{\theta,d}(a,b),\ \mu_y = \sum_b b \sum_a P_{\theta,d}(a,b)$$

$$\sigma_x = \sum_a (a-\mu_x)^2 \sum_b P_{\theta,d}(a,b),$$

$$\sigma_y = \sum_b \left(b-\mu_y\right)^2 \sum_a P_{\theta,d}(a,b)$$

3.4 Cluster with FCM

Fuzzy C-Means (FCM) is a technique of grouping (clustering) of data by means of any data in a cluster is determined by the value of membership. This technique was first introduced by Jim Bezdek in 1981. The basic concept of the FCM, the first is to determine the center of the cluster that will mark the average location for each cluster. At the initial condition, the center of this cluster is still not accurate. Each of the data has a degree of membership for each cluster. By improving the center of the cluster and the membership value of each data repeatedly, it will be seen that the center of the cluster will move to the right location. This iteration is based on minimizing the objective function. Data clustering is the process of dividing the data elements into clusters so that the elements that exist in a cluster are an element that is very similar to each other. Elements in different clusters may be at similar. Clustering method to be used is Fuzzy C Means (FCM) which has been modified. FCM algorithm will divide the data set into a n number with c clusters. The results of the algorithm will produce a list of cluster centers C = {c_1, …, c_c} and matrix partition U = μ_{ij}, i = 1, …, n, j = 1, …, c. where each element μ_{ij} shows the level of closeness with i-th element of the j-th cluster. The FCM algorithm has one major limitation, namely the total membership of the data to all cluster must be equal to one. The calculation of the new cluster centers c_j based on the membership value of each data. In general, the FCM algorithm can be expressed as follows [11]:

1. Determine the value of membership of a cluster of data with all the random
2. Repeat these steps until there is convergence.
   i. Calculate the center of each cluster c_j by Equation (5)

$$c_j = \frac{\sum_{i=1}^{n} \mu_{ij}{}^{m} \cdot x_i}{\sum_{i=1}^{n} \mu_{ij}{}^{m}} \tag{5}$$

   ii. Calculate the distance between the two data used methods Canberra Distance. This method is used because the values in the feature vector has a different scale so that the necessary normalization that has been taken into account in Canberra Distance. Equation (6) is used to compute the two data using methods Canberra Distance is as follows:

$$d_{xy} = \sum_{a=1}^{p} \frac{|x_a - y_a|}{|x_a| + |y_a|} \tag{6}$$

   iii. Update the membership of each data for each cluster μij based equation to calculate the value of μ_{ij} using equation (7):

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{2}{(m-1)}}} \tag{7}$$

   where

   μ_{ij} : Membership values to the data-i in j cluster

   d_{ij} : Distance data to the i-th to the j-th cluster center.

   m : The parameter that indicates the level of fuzziness, m> 1

   c : number of clusters

   d_{ik} : Distance data to the i-th to the k-th cluster center.

   iv. If convergence has not occurred back to 2.i.

4. RESULTS AND DISCUSSION

Processed image data batik consists of 50 images that have been processed from 10 classes a motive to 4 pieces of spatial characteristic each image of batik. Spatial characteristic obtained by Grey Level Co Occurrence Matrix includes contrast, energy, homogeneity and correlation.

Figure1 Batik image conversion results to Grey Scale

Some results on the spatial feature extraction batik image data can be seen in Table 1

Table 1. Spatial Feature Extraction Results Image Batik

NO	BATIK	contrast	Correlation	energy	homogeneity
1	grompol_2	1.47E+04	0.00280	1.36E-05	0.0318
2	Kawung picis	1.50E+04	-0.00700	1.48E-05	0.0315
3	Klitik_3	1.48E+04	0.00038	1.48E-05	0.0319
4	Parang barong_4	1.49E+04	-0.00410	1.37E-05	0.032
5	sidoasih_5	1.50E+04	0.00420	1.39E-05	0.032
6	sidomukti	1.47E+04	-0.00200	1.26E-05	0.0319
7	slobog	1.55E+04	0.00097	2.14E-05	0.0319
8	tirtateja_3	1.49E+04	-0.00110	1.27E-05	0.032
9	Wahyu tumurun_2	1.50E+04	-0.00024	1.50E-05	0.0321
10	nitik_5	1.51E+04	0.00820	2.17E-05	0.0322

The results obtained spatial feature extraction computed from the cluster center of each cluster based on Equation (5). Cluster centers obtained are shown in Table 2. Further calculated the distance between the two data used methods Canberra Distance. This method is used because the values in the feature vector has a different scale so that the necessary normalization that has been taken into account in Canberra Distance accordance Equation (6). This is done until there is convergence.

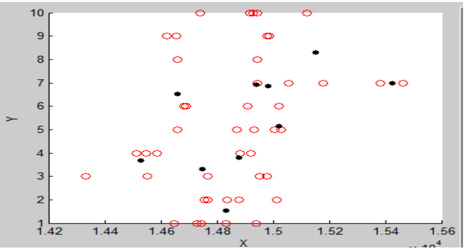
Table 2. Results of Cluster Data Center Batik image based spatial features

Iteration Count	Cluster Data Center
1	285126.119852

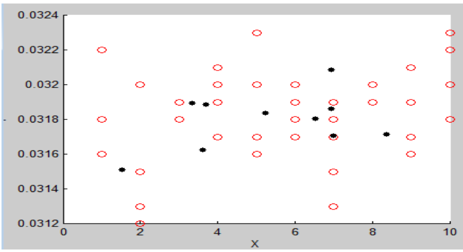
2	198826.735071
3	172516.605450
:	
35	27233.476723
36	27233.476713

Results cluster batik image based on spatial characteristics can be seen from the plot clusters formed in Figure 2.

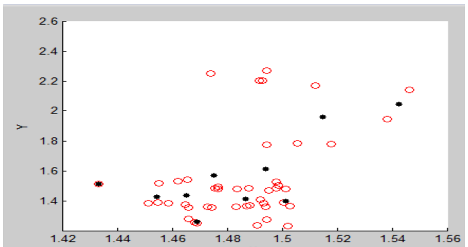
The results of experiments have been conducted obtained Fuzzy C Means able to form clusters based on spatial features batik image contrast, energy, correlation and homogeneity. Cluster obtained seen that the use of FCM in the cluster image Batik yield a range of image data cluster geometry batik spread so that the visible difference grouping clusters clear.



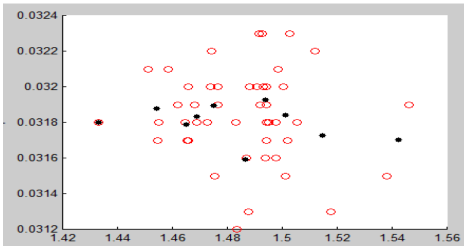
Cluster results based on contrast feature



Cluster results based on energy feature



Cluster results based on correlation feature



Cluster results based on homogeneity feature

Figure 2 Results of analysis of cluster centers Image Batik

5. CONCLUSION

This study did Fuzzy C Means clustering with the motif batik image obtained the following conclusions:

1. Fuzzy C Means able to build a cluster based on the characteristic spatial batik image contrast, energy, correlation and homogeneity.
2. Cluster obtained seen that the use of FCM in the cluster image Batik produces clusters corresponding to the number of defined classes. Cluster obtained seen that the use of FCM in the cluster image Batik produces batik image data distribution in cluster grouping characteristic obvious visible difference.

For further research can be developed to acquire the image of non-geometric gain so they can know the level of his cluster capabilities make use of Fuzzy C Means.

6. REFERENCES

[1] J. Yu, 2010. Texture Image Segmentation Based on Gaussian Mixture Models and Gray Level Co-occurrence Matrix. Third Int. Symp. Inf. Sci. Eng., no. 1, pp. 149–152

[2] Z. Juanjuan, L. U. Huijun, L. I. Yue, and C. Junjie, 2012. A kind of Fuzzy Decision Tree Based on the Image Emotion Classification. 2012 Int. Conf. Comput. Meas. Control Sens. Netw., no. 60970059, pp. 167–170

[3] B. Arisandi, N. Suciati, and A. Yudhi. 20121. Introduction to Batik With Rotated Wavelet Filter and Neural Network. Paper Session Final, Institut Teknologi Surabaya.

[4] A. A. Pratama, N. Suciati, and D. Purwitasari. 2012. Implementation of Fuzzy C-Means clustering for image Batik Motif Based on Texture Feature J. Tek. POMITS, vol. 1, no. 1, pp. 1–4.

[5] A. A. Kasim and A. Harjoko, 2014. Clustering Image Batik Using Neural Network Based on Gray Level Co-Occurrence Matrices (GLCM) in Seminar Nasional Teknologi Informasi.

[6] A. A. Kasim and R. Wardoyo. 2013. Batik Image Classification Rule Extraction Using Fuzzy Decision Trees in Information System International Conference (ISICO) Bali.

[7] V. S. Moertini and B. Sitohang. 2005. Algorithms of Clustering and Classifying Batik Images Based on Color, Contrast and Motif," ITB J. Eng. Sci., vol. 37, no. 2, pp. 141–160

[8] V. S. Moertini. 2005. Towards classifying classical batik images

[9] A. H. Rangkuti. 2014. Content Based Batik Image Classification Using Wavelet Transform And Fuzzy Neural Network J. Comput. Sci., vol. 10, no. 4, pp. 604–613

[10] A. Gebejes, E. M. Master, and A. Samples. 2013. Texture Characterization based on Grey-Level Co-occurrence Matrix in Conference Of Informatics and MAnagement ICTIC 2013, 2013, pp. 375–378.

[11] J. C. Bezdek, R. Ehrlich, and W. Full.1984.FCM : The Fuzzy C-Means Clustering Algorithm Comput. Geosci., vol. 10, no. 2, pp. 191–203