Clustering Batik Images using Fuzzy C-Means Algorithm Based on Log-Average Luminance

Ahmad Sanmorino

Faculty of Computer Science, Universitas Indonesia
ahmad.sanmorino@ui.ac.id

ABSTRAKSI

Batik adalah kain atau pakaian yang dibuat dengan teknik pewarnaan khusus yang disebut pencelupan anti lilin dan merupakan salah satu warisan budaya yang memiliki nilai seni tinggi. Dalam rangka meningkatkan efisiensi dan memberikan yang lebih baik semantik pada gambar, beberapa peneliti menerapkan algoritma pengelompokan untuk mengelola gambar sebelum gambar tersebut dapat diambil. Pengelompokan citra merupakan suatu proses pengelompokan gambar berdasarkan kesamaan mereka. Dalam makalah ini, memberikan alternatif metode pengelompokan gambar batik menggunakan algoritma fuzzy c-means (FCM) berdasarkan log-rata pencahayaan dari batik. Algoritma pengelompokan FCM merupakan algoritma yang bekerja dengan menggunakan model fuzzy yang memungkinkan semua data dari semua anggota kelompok terbentuk dengan derajat yang berbeda keanggotaan antara 0 dan 1. Log rata-rata pencahayaan (LAL) adalah nilai rata-rata dari pencahayaan dalam foto. Pencahayaan gambar yang berbeda dari satu gambar yang lain menggunakan LAL telah dibandingkan. Dari percobaan yang telah dilakukan, dapat disimpulkan bahwa algoritma FCM dapat digunakan untuk pengelompokan citra batik berdasarkan log rata-rata pencahayaan setiap gambar yang dimiliki.

Kata Kunci: pengelompokan, gambar batik, fuzzy c-means, log rata-rata pencahayaan

ABSTRACT

Batik is a fabric or clothes that are made with a special staining technique called wax-resist dyeing and is one of the cultural heritage which has high artistic value. In order to improve the efficiency and give better semantic to the image, some researchers apply clustering algorithm for managing images before they can be retrieved. Image clustering is a process of grouping images based on their similarity. In this paper, provide an alternative method of grouping batik image using fuzzy c-means (FCM) algorithm based on log-average luminance of the batik. FCM clustering algorithm is an algorithm that works using fuzzy models that allow all data from all cluster members are formed with different degrees of membership between 0 and 1. Log-average luminance (LAL) is the average value of the lighting in an image. The different image lighting from one image to another using LAL is compared. From the experiments that have been made, it can be concluded that FCM algorithm can be used for batik image clustering based on log-average luminance of each image possessed.

Keywords: clustering, batik images, fuzzy c-means, log-average luminance

1. INTRODUCTION

Batik is a fabric or clothes that are made with a special staining technique called wax-resist dyeing and is one of Indonesia's cultural heritage which has high artistic value. The learning process continues to be done to preserve batik, this is evidenced by various studies that have been made. Comprehensive study not only focus on the representation of batik in the real world, but how batik is represented in cyberspace, the Internet. It becomes something very important, because the internet is currently the principal means to deliver information. People from around the world can obtain information about batik, by simply clicking the search button on the search engines. And now, the most widely used representation of batik in the internet is a digital image. In order to improve the efficiency and give better semantic to the image, some researchers such as Chen [1], Liu [2], Guan [3], Kim [4], Park [5], Liu [6], Fakouri [7] apply clustering algorithm for managing images before they can be retrieved. Image clustering is a process of grouping images based on their similarity. By clustering image, the retrieval process does not need to examine images one by one to match with the user query. The system just needs to compare user query with the centroid of the clusters, then returns all images belong to the matched cluster [8]. In this paper we attempt to provide an alternative method of clustering batik images using the fuzzy c-means (FCM) algorithm [9] [10] [11] based on log-average luminance of the batik.

2. METHODOLOGY

2.1. Data

For the data used, we took 16 samples of batik image with the following details:

Table 1
Samples of Batik Image

ID	Type of Batik Image	ID	Type of Batik Image
B1	Batik Banten	B9	Batik Palembang
B2	Batik Ciamis	B10	Batik Pekalongan
B3	Batik Garut	B11	Batik Riau
B4	Batik Indramayu	B12	Batik Solo
B5	Batik Jambi	B13	Batik Tasikmalaya
B6	Batik Kuningan	B14	Batik Tuban
B7	Batik Madura	B15	Batik Cirebon
B8	Batik Minang	B16	Batik Yogya

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

2.2. Fuzzy C-Means

Fuzzy C-Means (FCM) is an algorithm that uses fuzzy models in the clustering data, allowing all data to be a member of any cluster is formed with different degrees of membership between 0 and 1. Figure 1 shows the clustering results are shown in chart [12].

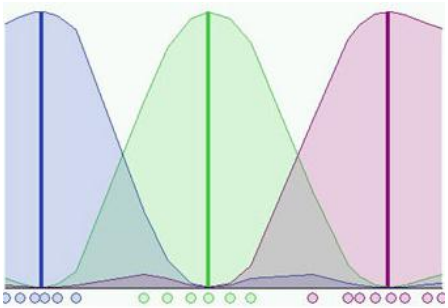

Figure 1. Fuzzy C-Means Clustering

The first step to cluster datas using FCM is to determine the cluster center. Next, specify the degree of membership of each data for each cluster. Then, perform value iteration updates the cluster centers and the degree of membership of each data. So that steps FCM clustering algorithm can be described as follows:

1. The first step specify:
   a. Matrix X, n x m, n = the number of datas; and m = the number of attributes
   b. The number of clusters (C ≥ 2)
   c. Weight (w > 1)
   d. The number of iterations allowed
   e. Criteria for termination of iterations (ε)
2. Create an initial partition matrix U (with the degree of membership of each data in the cluster)

$$U = \begin{bmatrix} \mu_{11}(x_1) & \mu_{12}(x_2) & \cdots & \mu_{1n}(x_n) \\ \mu_{21}(x_1) & \mu_{22}(x_2) & \cdots & \mu_{2n}(x_n) \\ \vdots & & & \vdots \\ \mu_{c1}(x_1) & \mu_{c2}(x_2) & \cdots & \mu_{cn}(x_n) \end{bmatrix}$$

3. Calculate the V cluster center for each cluster

$$V_{ij} = \frac{\sum_{k=1}^{n} (\mu_{ik})^{w} \cdot x_{kj}}{\sum_{k=1}^{n} (\mu_{ik})^{w}} \quad (1)$$

4. Update the degree of membership of each data on all cluster

$$\mu_{ik} = \left[\sum_{j=1}^{C} \left(\frac{d_{ik}}{d_{jk}} \right)^{2/(w-1)} \right]^{-1} \quad (2)$$

$$d_{ik} = d(x_k - v_i) = \left[\sum_{j=1}^{m} (x_{kj} - v_{ij}) \right]^{1/2}$$

5. Specify the iteration termination criteria based on changes in the partition matrix on the current iteration and previous iterations

$$\Delta = \left\| U^t - U^{t-1} \right\| \quad \Delta < \varepsilon \quad (3)$$

2.3. Log-Average Luminance

We can compare the luminance of different images using so-called log-average luminance of an image. The log-average luminance is calculated by finding the geometric mean of the luminance values of all pixels. In a gray scale image, the luminance value is the pixel value. In a color image, the luminance value is found by a weighted sum of RGB value [13].

RGB color space represents each pixel by using red (R), green (G) and blue (B) coordinate, each color component is represented by one or more bytes [8]. Standard formula that can be used to calculate the log-average luminance value is:

L = (0.2126*R) + (0.7152*G) + (0.0722*B)

L = (0.299*R) + (0.587*G) + (0.114*B) (4)

There is another formula with a better accuracy rate, but takes more time. That is:

L = sqrt (0.241*R^2 + 0.691*G^2 + 0.068*B^2) (5)

3. EXPERIMENT

Clustering batik image is determined based on log-average luminance value (LAL) of every image, so the first step clustering parameter is determined as follows:

1. Matrix n x m, which n = 16 batik image data, and m = 1, the number of attributes
2. The number of clusters C = 3
3. Weight (bobot) W = 2
4. Max Iterations = 100
5. Criteria for termination of the iteration $\varepsilon = 10^{-6}$

The next step is to create initial partition matrix, in determining the degree of membership of each cluster data, we determined it out randomly. For experiment we used Mathlab**.**

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

4. RESULT AND DISCUSSION

Calculation of the center cluster and the degree of membership of each data is stopped at iteration 9, with the value of Δ = 2.8712e-06. The final results are V = [C1 = 19.3005; C2 = 60.8180; C3 =118.5947]

Table 2
The Final Results at Iteration 9

ID	Type of Batik Image	LAL	Trend of Clusters		
			C1	C2	C3
B1	Batik Banten	62.315	0.027	0.955	0.018
B2	Batik Ciamis	55.291	0.058	0.911	0.031
B3	Batik Garut	120.620	0.021	0.034	0.945
B4	Batik Indramayu	20.112	0.961	0.028	0.011
B5	Batik Jambi	71.491	0.064	0.867	0.069
B6	Batik Kuningan	115.554	0.031	0.046	0.923
B7	Batik Madura	67.301	0.054	0.895	0.051
B8	Batik Minang	24.114	0.906	0.066	0.028
B9	Batik Palembang	19.132	0.983	0.012	0.005
B10	Batik Pekalongan	128.607	0.046	0.072	0.882
B11	Batik Riau	109.571	0.039	0.068	0.893
B12	Batik Solo	60.284	0.014	0.978	0.008
B13	Batik Tasikmalaya	11.093	0.852	0.105	0.043
B14	Batik Tuban	58.332	0.029	0.952	0.019
B15	Batik Cirebon	49.351	0.087	0.872	0.041
B16	Batik Yogya	21.140	0.935	0.044	0.021

Cluster C1 obtained for the input matrix (LAL) between 11.093 to 24.114. Cluster C2 obtained for the input matrix (LAL) between 49.351 to 71.491. While, cluster C3 obtained for the input matrix (LAL) between 109.571 to 128.607. This experiment obtained 3 clusters:

1. The first cluster, with the members B4, B8, B9, B13, and B16. With the center of cluster 19.30.
2. The second cluster, with the members B1, B2, B5, B7, B12, B14, B15. Center of cluster 60.82.
3. The third cluster, with the members B3, B6, B10, and B11. With the center of cluster 118.59.

For evaluation we use k-means clustering algorithm as a comparison. The results of the comparison can be seen in Table 3 below.

Table 3
Comparison FCM and K-Means

	FCM	K-Means
Center of Cluster	C1 = 19.30; C2 = 60.81; C3 =118.59	C1 = 19.12; C2 = 60.62; C3 =118.59
Final Cluster	At Iteration 9	At Iteration 5
Member of Cluster	C1 = B4, B8, B9, B13, B16 C2 = B1, B2, B5, B7, B12, B14, B15 C3 = B3, B6, B10, B11	C1 = B4, B8, B9, B13, B16 C2 = B1, B2, B5, B7, B12, B14, B15 C3 = B3, B6, B10, B11

FCM algorithm requires more iterations than k-means algorithm. This happens because FCM algorithm must update the cluster centers and the degree of membership of each data, and calculate the distance at each iteration. Unlike k-means clustering algorithm, where each iteration only perform for distance calculations and update the center of cluster.

5. CONCLUSION

The value of membership degrees that has been determined at the beginning and the number of attributes used greatly affect the final result of clustering. In this experiment we only use one attribute that is the log-average luminance. At the future work can be added several attributes such as color composition, contrast or motifs of the batiks. From the experiments that have been made it can be concluded that the fuzzy c-means algorithm can be used for batik image clustering based on log-average luminance of each image possessed. The process of learning and more detailed analysis can be done in the future so that the resulting clustering becomes more optimal.

REFERENCES

[1]. Chen, Y., J.Z. Wang, R. Krovetz, "Content-Based Image Retrieval by Clustering", Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval, 2003, pp. 193-200.

[2]. Liu, R., Y. Wang, T. Baba, Y. Uehara, D. Masumoto and S. Nagata, "SVM-Based Active Feedback in Image Retrieval Using Clustering and Unlabeled Data. LNCS, Computer Analysis of Images and Patterns", Springer Berlin / Heidelberg, Volume 4673/2007, August 2007, pp. 954-961.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

[3]. Guan, J., G. Qiu, "Spectral images and features co-clustering with application to content-based image retrieval", In Proc. of IEEE Workshop on Multimedia Signal Processing, 2005.
[4]. Kim, D., "Qcluster: Relevance Feedback Using Adaptive Clustering for Content-Based Image Retrieval", In Proc. of the ACM SIGMOD Int. Conf. on Management of Data, 2003.
[5]. Park, S., K. Seo, D. Jang, "Fuzzy Art-Based Image Clustering Method for Content-Based Image Retrieval", International Journal of Information Technology and Decision Making, 06(02), 2007.
[6]. Liu, Y., X. Chen, C. Zhang, A. Sprague, "An Interactive Region-Based Image Clustering and Retrieval Platform", In Proc. of the IEEE International Conference on Multimedia and Expo, 2006, pp. 929-932.
[7]. Fakouri, R., B. Zamani, M. Fathy, and B. Minaei, "Region-Based Image Clustering and Retrieval Using Fuzzy Similarity and Relevance Feedback", In Proc. Of the International Conference on Computer and Electrical Engineering, 2008.
[8]. Hidayanto, A.Nizar, Koeanan, E.M., "Journey on Image Clustering Based on Color Composition", Faculty of Computer Science, Universitas Indonesia, 2009.
[9]. Dunn, J. C., "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", Journal of Cybernetics 3: 32-57, 1973.
[10]. Bezdek, James C., Robert Ehrlich, William Full, "FCM: The Fuzzy C-Means Clustering Algorithm". Computer & Geosciences, Vol.10, No.2-3, pp.191-203, 1984.
[11]. Nock, Richard. and Frank Nielsen, "On Weighting Clustering, IEEE Transactions on Pattern Analysis and Machine Intelligence", Vol.28, No.8, 2006.
[12]. Anonymous, "A Tutorial on Clustering Algorithms: Fuzzy C-Means Clustering", http://home.dei.polimi.it/ matteucc/Clustering/tutorial_html/cmeans.html. Accessed on April, 4, 2012.
[13]. Chu, Henry, "Luminance of Images", http://www.cacs.louisiana.edu/~cice/lal/index.html. Accessed on April, 4, 2012.