

Coastal Batik Motifs Identification Using K-Nearest Neighbor Based On The Grey Level Co-Occurrence Method

Wresti Andriani¹

Informatics Engineering
STMIK YMI Tegal

Email: wresti.andriani@gmail.com

Gunawan^{2*}

Informatics Engineering
STMIK YMI Tegal

Email: gunawan.gayo@gmail.com

Sawaviyya Anandianskha³

Informatics Engineering
STMIK YMI Tegal

Email: sayaviyyaa@gmail.com



Abstrak-Indonesia merupakan negara yang kaya akan sumber daya alam, budaya dan pariwisata. Salah satu warisan budaya manusia yang terkenal di Indonesia adalah batik. Batik memiliki keunikan motif yang sangat beragam sehingga sulit untuk mengenali golongan tertentu terutama generasi muda. Penelitian ini dilakukan untuk mengklasifikasikan batik pesisir, khususnya batik Tegal, batik Pekalongan, dan batik Cirebon sehingga dapat membantu memudahkan pengenalan dan pemahaman batik pesisir jika dibandingkan dengan batik pedalaman, seperti batik Yogyakarta. Metode yang digunakan adalah Gray Level Co-occurrence Matrix (GLCM) untuk mengekstrak fitur tekstur, sedangkan untuk menentukan kedekatan citra uji dengan data latih menggunakan metode K-Nearest Neighbor (KNN), perhitungan jarak yang digunakan adalah Euclidean Distance dan Manhattan Distance berdasarkan karakteristik tekstur dari citra batik yang diperoleh. Hasil yang diperoleh pada penelitian ini dimana skor tertinggi adalah 64% untuk Euclidean Distance dan 66% untuk Manhattan Distance pada $k = 15$.

Kata Kunci— Pesisir, GLCM, KNN, Identifikasi

Abstract-Indonesia is a country rich in natural resources, culture, and tourism. One of the famous human cultural heritage in Indonesia is batik. Batik has unique motifs that are very diverse, making it difficult to recognize certain groups, especially the younger generation. This research was conducted to classify coastal batik, especially Tegal batik, Pekalongan batik, and Cirebon batik so that it can help facilitate the introduction and understanding of coastal batik when compared to inland batiks, such as Yogyakarta batik. The method used is the Grey Level Co-occurrence Matrix (GLCM) to extract texture features, while, to determine the proximity of the test image to the training data using the K-Nearest Neighbor (KNN) method, the calculation of the distance to be used is the Euclidean Distance and Manhattan Distance based on the texture characteristics of the batik image obtained. The result obtained in this study where the highest score of 64% for Euclidean Distance and 66% for Manhattan Distance at $k=15$.

Keywords— Coastal, GLCM, KNN, Identification

I. INTRODUCTION

Indonesia is a country rich in culture and beautiful nature. One of the wealth that is owned is the batik culture. Indonesian batik varies according to the many cultures that are owned in each region in Indonesia. Batik is an art that produces picture cloth which is the original heritage of the Indonesian nation and is one of the world heritages and has been inaugurated by UNESCO, the United Nations world agency in the fields of culture and education. Based on the region, batik can be divided into two types, namely coastal batik and inland batik and based on the manufacturing process, batik is also divided into two, namely written batik, namely batik made by writing use liquid "wax" on cloth and stamp batik, namely batik made by depicting cloth using a stamp that has a pictorial part and is shaped like a relief and inscribed using wax liquid, then processed in a certain way. Each batik from various regions has its own uniqueness and high traditional artistic value. The diversity of traditional batik motifs is due to differences in geography, flora and fauna, differences in lifestyle and livelihoods. In the past, batik work was often done by women and was an exclusive job, because the batik cloth produced was presented to the nobility or distinguished guests.

Coastal batik motives are batik motives produced in coastal areas such as the Tegal coast, Pekalongan coast, and Cirebon coast, while the motives can vary, usually in addition to being influenced by geography and native culture, but also often influenced by culture brought by fishermen or traders from outside the area, making coastal fabric motives more diverse compared to batik motifs from the interior, such as batik motives from Yogyakarta which are more influenced by the sense of nobility of the palace.

The wide variety of traditional batik motifs in Indonesia confuses traders and the younger generation (millennials) in recognizing the origin of existing batik fabrics. Because there is no center or office that specifically handles and provides information about the history of batik itself. In this paper, the researcher hopes to help facilitate the millennial generation and related parties to become more familiar with motifs so that they can love these traditional products more, especially Tegal batik, Pekalongan batik, and Cirebon batik.

This study, will use the Gray Level Co-Occurrent Matric (GLCM) method as its feature extraction and use the K-

Nearest Neighbor (KNN) method. The distance calculation uses Euclidean Distance and Manhattan Distance to determine the proximity of the test image to the training data and is expected to help identify and classify the image of batik motives. The well-identified image of batik will provide clear information and can be used for the preservation of Indonesian batik fabric motifs from extinction.

Similar studies that have been conducted by some researchers using batik objects include Zulfrianto Y. Lamasigi [1] DCT For Extraction Of GLCM-Based Features on Batik Identification Using K-NN. The highest accuracy obtained by DCT-GLCM exists at an angle of 135° with a value of k=3 of 64.88% and a value of 64.88% and at an angle of 0° with values k=7 and 9 are 41.86%. Frisnanda Aditya, etc [2] Pekalongan Batik Identification Using the Grey Level Co-Occurrence Matrix and Probabilistic Neural Network Method. From the results of this classification test, the best accuracy is 61.33%. There is has been no research that examines the identification of the difference between coastal and inland batik using GLCM and KNN method before.

II. METHOD

A. Gray Level Co-occurrence Matrix (GLCM)

Grey Level Co-occurrence Matrix (GLCM) was first submitted by Haralick in 1979 with 28 features to explain spatial patterns. There are several steps that are taken, namely: first, calculating the features of the GLCM by converting an RGB image into a grey scale image. Second, creating a co-occurrence matrix is continued by determining the spatial relationship between reference pixels and neighbouring pixels based on angle θ and distance d . Third, creating a symmetrical matrix by adding a co-occurrence matrix to the transpose matrix. Fourth, the symmetrical matrix is normalized by calculating the probability of each element of the matrix. Fifth, calculate the features of the GLCM. Each feature is calculated by one pixel distance in four directions, i.e. 0°, 45°, 90°, and 135° to detect co-occurrence [1] [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12]. There are 5 GLCM features used in this study, including:

1) Angular Second Moment (ASM)

ASM is a measure of the homogeneity of the image.

$$ASM = \sum_{i=1}^L \sum_{j=1}^L (GLCM(i, j))^2 \quad (1)$$

2) Contrast

Contrast is a measure of the presence of variations in the grey level of images.

$$Contrast = \sum_i^L \sum_j^L |i - j|^2 GLCM(i, j) \quad (2)$$

3) Inverse Different Moment (IDM)

Used to measure homogeneity

$$IDM = \sum_{i=1}^L \sum_{j=1}^L \frac{(GLCM(i, j))^2}{1 + (i - j)^2} \quad (3)$$

4) Entropy

Entropy for represents the measure of the grey level irregularity in the image.

$$Entropi = - \sum_{i=1}^L \sum_{j=1}^L (GLCM(i, j)) \log(GLCM(I, J)) \quad (4)$$

5) Correlation

Correlation is a measure of the dependence between the grey values in the image.

$$Correlation = \frac{\sum_{i=1}^L \sum_{j=1}^L \frac{(i - \mu_i')(j - \mu_j')(GLCM(i, j))}{\sigma_i \sigma_j}}{\sigma_i \sigma_j} \quad (5)$$

This equation is based on the mean value of the grey image intensity and the standard deviation. The standard deviation is obtained from the square root of the variant which shows the distribution of pixel values in the image, with the following formula:

$$mean\ i = \mu_i' = \sum_{i=1}^L \sum_{j=1}^L i * GLCM(i, j)$$

$$mean\ j = \mu_j' = \sum_{i=1}^L \sum_{j=1}^L j * GLCM(i, j)$$

$$varian\ i = \sigma_i^2 = \sum_{i=1}^L \sum_{j=1}^L GLCM(i, j)(i - \mu_i')^2$$

$$varian\ j = \sigma_j^2 = \sum_{i=1}^L \sum_{j=1}^L GLCM(i, j)(j - \mu_j')^2$$

$$standard\ deviation\ i = \sigma_i = \sqrt{\sigma_i^2}$$

$$standard\ deviation\ j = \sigma_j = \sqrt{\sigma_j^2}$$

B. K-Nearest Neighbor

K-Nearest Neighbour (KNN) is a method that uses a supervised algorithm the results of the newly classified query instance based on the majority of the categories on the KNN. This algorithm aims to classify new objects based on attributes and training samples. The KNN algorithm is very simple, based on the shortest distance from the query instance to the training sample to determine its KNN. The training sample is projected into a multi-dimensional space, where each dimension represents a feature of the data. Space is divided into sections-part based on the classification of the training sample. A point in this space is marked class c if class c is the most common classification in the k nearest neighbour of that point.

1) Euclidean Distance

Calculation of the distance of Euclidean Distance which is represented as follows [3] [13]:

$$d = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2}$$

$$d = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (6)$$

Where $d(a, b)$: The Euclidean distance between the vector a and the vector b , a_i : feature vector a , b_i : features of the vector b , n : the number of features in the a and b vectors.

2) Manhattan Distance

Manhattan or City Distance is a similarity measurement that is most suitable for project approvals that represent relevant cases with natural numbers or with quantitative data. Also used to retrieve matched cases from the case base by calculating the absolute weighted sum of the differences between the current case and other cases the case base. To calculate the weight, the following quationis used:



$$d_{ij} = \sum W_k |x_{ik} - C_{jk}| \quad (7)$$

d_{ij} = distance between cases i and j

W = represent the sum of weight

X = newly reduced case with C

C. Confusion Matrix

The confusion matrix is a table consisting of many rows of data test that predicted true and false by the classification model, to determine the performance of a classification model [14] [15].

Table I. Confusion Matrix

Actual Class		Predicted Class	
		Class = 1	Class = 0
Class = 1	F 11	F 10	
Class = 0	F 01	F 00	

Accuracy calculation using confusion matrix as follows:

$$accuracy = \frac{F11+F00}{F11+F10+F01+F00}$$

or

$$Accuracy = \frac{\text{Number Classification true}}{\text{number data}} \times 100\%$$

The flow/stages that will be done in this study can be seen in Figure 1 below:

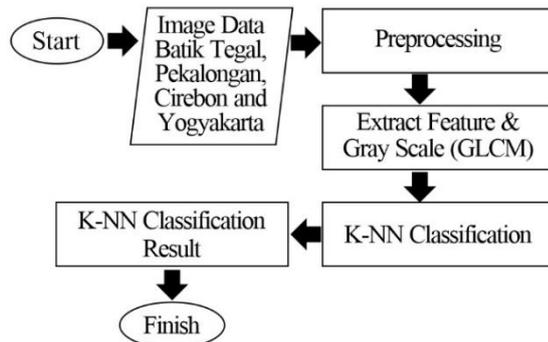


Figure 1. Research Flowchart

III. RESULTS AND DISCUSSION

Data collection (data collection) is obtained from literature study, observation, and data search in the form of batik motive images via the internet. Image data obtained are as many as 90 batik motives, consisting of coastal batik from the city of Tegal as many as 20 batik motifs, 50 kinds of Pekalongan batik motifs, and 20 kinds of Cirebon batik motives, also 10 kind of Yogyakarta batik motives. Batik Yogyakarta as representative of inland batik. All the batik data motives obtained are converted into the same size and extension * jpg. This data image is then grouped into training data. Yogyakarta batik motif images. All of the test data is the same size and also has a jpg extension.

Below are some examples of images of batik motives that have been collected:

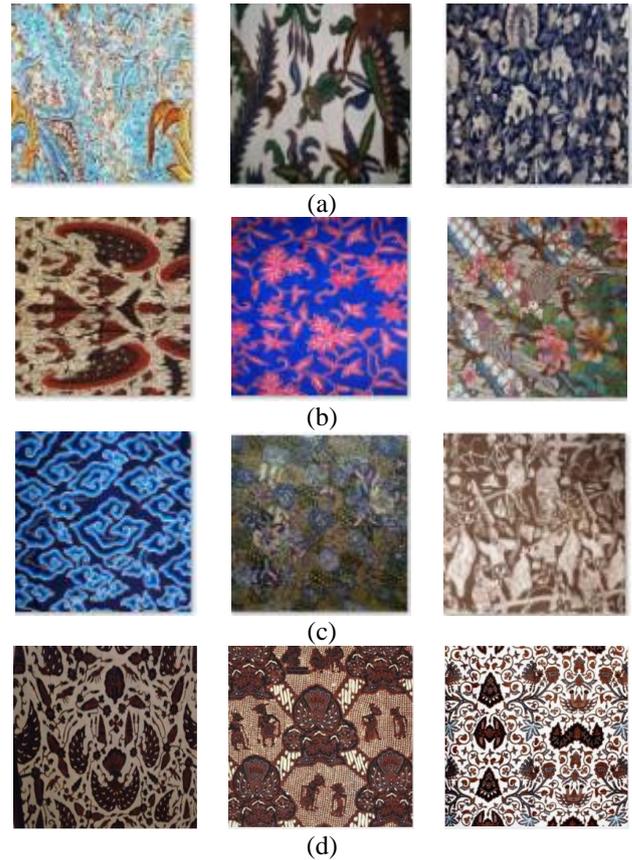


Figure 2. Example Batik (a) From Tegal (b) From Pekalongan (c) From Cirebon (d) From Yogyakarta

Each image is done with the feature extraction process using GLCM using MATLAB 2020. From MATLAB results, then feature extraction is classified using KNN algorithm with 2 kinds of distance measurement, namely Euclidean Distance and Manhattan Distance and using rapid miner tool.

A. Texture Feature Extraction

In the previous process, we obtained a total of 90 data of batik motifs as training data consisting of 20 Tegal batik motifs, 50 kinds of Pekalongan batik motifs and 20 kinds of Cirebon batik motifs along with 10 kinds of batik motifs from Yogyakarta as samples of inland batik motifs as a preprocessing stage. The results of this stage are then continued at the extraction stage of the texture feature. The extraction of this feature is done using MATLAB 2020 which results in Table II below:

Table II. GLCM Results

No	Matrix	Eccentat ion ...	Next ...	Korel ...	Energy	Homo...
1	116,570	68,016	0.617	0.005	5,453	0.066
2	90,403	42,895	0.148	0.006	5,134	0.028
3	132,540	56,192	0.210	0.005	5,338	0.046
4	104,627	72,138	0.814	0.005	5,318	0.074
5	187,068	46,154	1,568	0.008	5,062	0.032
⋮						
90	93,465	46,579	0.464	0.006	5,185	0.032

Once the GLCM feature extraction results using MATLAB 2020 are obtained, the feature extraction data is obtained from GLCM, then the data is entered into the rapid miner tool and



using the Loop Parameter, it can be determined the value $k = 1, 3, 5, 7, 9, 11, 13, 15$ for used, the k which produces the highest accuracy value to be used from the distance method, The method used are Euclidean Distance and Manhattan (City) Distance. The results can be seen in the Table 3 below:

Table III. Comparison between the Manhattan and Euclidean Methods

	Accuracy							
Distance	k=1	k=3	k=5	k=7	k=9	k=11	k=13	k=15
Manhattan	55	60	61	62	59	61	57	66
Euclidean	52	59	60	61	63	62	62	64

The comparison of accuracy in Table III between Euclidean Distance and Manhattan Distance above, a graphic image is obtained as shown in Figure 3.

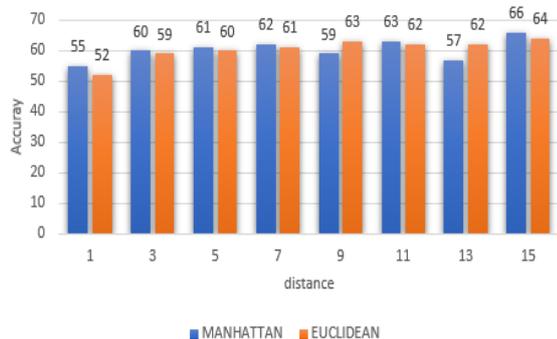


Figure 3. Comparison of Manhattan and Euclidean

In the Table III and Figure 3, it can be seen that the comparison from Euclidean Distance and Manhattan Distance, both of them have the highest accuracy at the value $k = 15$, where the highest accuracy is obtained at 64% for Euclidean Distance and is 66% for Manhattan Distance.

The same occurrence for the smallest accuracy value is obtained at the value $k=1$. Manhattan's accuracy is 55% and Euclidean's is 52%.

B. Confusion Matrix

In this study, in addition to knowing the comparison between Euclidean Distance and Manhattan City Distance, it will also be seen how accurate the coastal batik is with the inland batik by knowing the confusion matrix. The result of the Confucius matrix from the classification of coastal batik with inland batik, in this case, represented by Tegal batik, Pekalongan batik, and Cirebon batik for coastal batik and Yogyakarta batik as inland batik can be seen from the following Table IV.

Table IV. Prediction Motive Batik Performance

	True Cirebon	True Yogyakarta	True Pekalongan	True Tegal	Class Precision
Pred. Cirebon	8	1	3	1	61,54%
Pred. Yogyakarta	0	1	0	4	20,00%
Pred. Pekalongan	8	2	44	4	75,86%
Pred. Tegal	4	6	3	11	45,83%
Class Recall	40%	10%	88%	55%	45,83%

When the prediction results above are tested using the confusion matrix obtained results in Table V below:

Table V. Matrix of the Predictions

No.	Label	Prediction	Result
1	Cirebon	Pekalongan	0
2	Cirebon	Pekalongan	0
3	Yogyakarta	Pekalongan	0
4	Pekalongan	Pekalongan	1
5	Pekalongan	Pekalongan	1
6	Pekalongan	Pekalongan	1
7	Pekalongan	Pekalongan	1
8	Pekalongan	Pekalongan	1
9	Tegal	Pekalongan	0
10	Tegal	Pekalongan	0
11	Cirebon	Pekalongan	0
12	Cirebon	Pekalongan	0
13	Yogyakarta	Pekalongan	0
14	Pekalongan	Pekalongan	1
15	Pekalongan	Pekalongan	1
16	Pekalongan	Pekalongan	1
17	Pekalongan	Pekalongan	1
18	Pekalongan	Pekalongan	1

The number 0 indicates the wrong label prediction, the number 1 indicates the correct prediction. The accuracy of predictions using Manhattan (City) Distance is higher than Euclidean Distance which is 66%. From the batik motif data that has been collected, there is a prediction that the true batik from Pekalongan is higher 75.86%, than the true batik from Cirebon is 61.54%. It can be concluded data from the training data obtained, after being tested was more predicted as Pekalongan batik, then followed by Cirebon batik.

IV. CONCLUSION

Based on the results of the above research, which uses 90 training data consisting of 20 data Tegal batik motifs, 50 data Pekalongan batik motifs, and 20 Cirebon batik motifs, as well as 40 test data, consisting of 10 data of original Tegal batik motifs, 10 Pekalongan batik motifs, can be concluded:

- 1) The accuracy of Cirebon batik motifs and Yogyakarta batik motifs using Manhattan distance method is better than Euclidean Distance, which is 66%.
- 2) The value $k=1$ of Manhattan Distance and Euclidean Distance is the smallest is 54% for Manhattan and 52% for Euclidean Distance.
- 3) Predicted result of batik motif from Pekalongan, which is 75% and followed by Cirebon batik by 61.54%, then the prediction of batik Tegal is 45.85% and then batik Yogyakarta that is equal to 20%.

For more research, researchers suggested that in testing it is recommended to use different object retrieval sizes and use different methods from the research created by current researchers in order to produce even better accuracy.

REFERENCES

- [1] F. A. etc, "Pekalongan Batik Identification Using the Gray Level Co-Occurrence Matrix and Probabilistic Neutral Network Method," *e-Proceeding of Engineering*, vol. 6, p. 10234, 2019.



- [2] H. C. D. a. S. A. A. Halim, "Image Retrieval Application Using a Combination of Color Moment and Gabor Texture Methods," *JSM STMIK Mikrisil*, vol. 14, 2013.
- [3] A. K. a. A. Susanto, *Image Processing Theory and Application*, Yogyakarta: Andi, 2012.
- [4] I. S. a. Y. C. AJ Arriawati, "Classification of Texture Images Using K-Nearest Neighbor Based on the Characteristics Extraction of the Cookbook Matrix Method," Diponegoro University, Semarang.
- [5] N. S. a. A. W. B. Arisandi, "Introduction to Batik Motif Using Rotated Wavelet Filters and Neural Networks," *JUTI*, vol. 9, pp. 13-19, 2011.
- [6] S. D. Cahyo, "Comparative Analysis of Several Edge Detection Methods Using Delphi 7," Gunadarma University, Depok, 2009.
- [7] C. C. a. T. B. o. C. Regency, "Casta and Taruna, Batik Cirebon, World Cultural Heritage from Indonesia, Cirebon," Cirebon, 2007.
- [8] D. P. Pamungkas, "Image Extraction Using GLCM and KNN Methods to Identify Types of Orchids (Orchidaceae)," vol. 1, p. 2, 2019.
- [9] E. Prasetyo, *Data Mining Processes Data Into Information Using Matlab*, Yogyakarta: Andi, 2014.
- [10] Eliyani, *The introduction of ripe papaya fruit levels using RGB color-based image processing with k-means clustering*, Lhokseumawe: Lhokseumawe State Polytechnic, 2013.
- [11] H. Priyanto, *Digital Image Processing Theory and Real Applications*, Bandung: Informatics Bandung, 2017.
- [12] H. Wijayanto, "Klasifikasi Batik Menggunakan Metode K-Nearest Neighbour Berdasarkan Gray Level Co-Occurance Matrixes (GLCM)," 2015.
- [13] J. Ong, "Implementation of the K-Means Clustering Algorithm to Determine President University's Marketing Strategy," *Scientific Journal of Industrial Engineering*, vol. 12, pp. 10-30, 2013.
- [14] I. M. Johan Wahyudi, "Introduction to Traditional Fabric Image Patterns Using GLCM and KNN," *JTIULM*, vol. 4, pp. 43-48, 2019.
- [15] M. S. etc, *Comparison of Texture and Color Feature Extraction for Classification of Lamongan Batik*, Tuban, 2017.
- [16] A. H. a. A. P. H. Rangkuti, "Content-Based Drawing of Batik," *Journal or Computer Science*, vol. 10, pp. 925-934, 2014.
- [17] A. P. D. & W. R. Triprasetyo, "Application of Trenggalek Batik Pattern Recognition Using Sobel Edge Detection and KMeans Algorithm," *Generation Journal*, vol. 2, pp. 25-32, 2018.
- [18] Z. Y. Lamasigi, "DCT for Feature Extraction based on Glcm on Batik Identification Using K-NN," *Jambura Journal of Electrical and Electronics Engineering*, vol. 3, 2021.
- [19] N. L. W. S. R. Ginantra, "Detection Of Batik Parang Using The Co-Occurrence Matrix And Features

