

Classification of Sambas Traditional Fabric “Kain Lunggi” Using Texture Feature

Alda Cendekia Siregar*¹, Barry Ceasar Octariadi²

^{1,2}Department of Informatics, Universitas Muhammadiyah, Pontianak, Indonesia
e-mail: *¹alda.siregar@unmuhpnk.ac.id, ²barry.ceasaro@unmuhpnk.ac.id

Abstrak

Kain tradisional adalah warisan budaya yang harus dilestarikan. Kain Lunggi adalah kain tradisional Sambas yang mengalami penurunan pada perajinnya. Untuk memperkenalkan Kain Lunggi dalam masyarakat nasional dan global yang lebih luas untuk melestarikannya, sistem berbasis pengolahan citra digital untuk melakukan pengenalan pola Kain Lunggi perlu dibangun. Ekstraksi fitur adalah bagian penting dari pemrosesan gambar digital. Fitur visual yang tidak mewakili karakteristik objek akan mempengaruhi keakuratan sistem pengenalan. Tujuan dari penelitian ini adalah untuk melakukan pemilihan fitur pada set fitur untuk menentukan fitur terbaik yang dapat meningkatkan akurasi pengenalan.

Penelitian ini dilakukan dalam beberapa langkah yaitu akuisisi gambar pola Kain Lunggi, preprocessing untuk mengurangi noise gambar, ekstraksi fitur untuk mendapatkan fitur gambar, dan pemilihan fitur. GLCM diimplementasikan sebagai metode ekstraksi fitur. Hasil ekstraksi fitur akan digunakan dalam proses pemilihan fitur menggunakan metode CFS (Correlation based Feature Selection). Fitur yang dipilih dari proses CFS adalah Angular Second Moment, Contrast, dan Correlation. Evaluasi fitur yang dipilih dilakukan dengan menghitung akurasi klasifikasi dengan metode KNN. Akurasi klasifikasi sebelum ekstraksi fitur adalah 85,18% dengan nilai $K = 1$, sedangkan akurasi meningkat menjadi 88,89% setelah pemilihan fitur. Peningkatan akurasi tertinggi 20,74% di KNN terjadi ketika menggunakan nilai $K = 4$.

Kata kunci— Klasifikasi, Fitur Tekstur, GLCM, KNN, Kain Lunggi

Abstract

Traditional fabric is a cultural heritage that has to be preserved. Kain Lunggi is Sambas traditional fabric that saw a decline in its crafter. To introduce Kain Lunggi in a broader national and global society in order to preserve it, a digital image processing based system to perform Kain Lunggi pattern recognition need to be built. Feature extraction is an important part of digital image processing. The visual feature that does not represent the character of an object will affect the accuracy of a recognition system. The purposes of this research are to perform feature selection on sets of feature to determine the best feature that can increase recognition accuracy.

This research conducted in several steps which are image acquisition of Kain Lunggi pattern, preprocessing to reduce image noise, feature extraction to obtain image features, and feature selection. GLCM is implemented as a feature extraction method. Feature extraction result will be used in a feature selection process using CFS (Correlation-based Feature Selection) methods. Selected features from CFS process are Angular Second Moment, Contrast, and Correlation. Selected features evaluation is conducted by calculating classification accuracy with the KNN method. Classification accuracy prior to feature extraction is 85.18% with K values $K=1$; meanwhile, the accuracy increases to 88.89% after feature selection. The highest accuracy improvement of 20.74% in KNN occurred when using K value $K=4$.

Keywords— Classification, Texture Feature, GLCM, KNN, Kain Lunggi

1. INTRODUCTION

Cultural heritage is an important aspect of human civilization that has to be protected and preserved. Traditional fabric is a significant cultural heritage of Indonesia. “Kaing Lunggi” or gold woven fabric from Sambas is one of Indonesia traditional fabric that needs to be preserved and introduced to the global world. Kain Lunggi is regional pride of Sambas and has been inducted to *Unesco Award of Excellence for handicrafts* in 2012 [1].

In current days, there are very few Kain Lunggi weavers that still active, and the numbers keep decreasing. Based on data from Sambas’s Dinas Koperasi UMKM Perindustrian dan Perdagangan, in 1999 there are 808 Kain Lunggi weavers, and it decreases to 365 in 2009 up until 2013 in which only 256 weavers remain. [1].

One of the effort to preserve kain Lunggi and keep it recognized by local and international society is to build a system that able to recognize and identify Kain Lunggi patterns. In return, the importance of research on Kain Lunggi digital image processing is increased. Feature extraction is an important part of image recognition. Researches about traditional fabric pattern recognition based on texture have been conducted by many researchers [2][3][4][5][6].

Utilization of improper feature in object recognition will affect its result. Features that do not represent the character of an object will cause low accuracy; on the other hand, correct features will positively affect the accuracy of a recognition system. Feature selection is a process to determine features that correlate one another without having to use the entire result of feature extraction [7]. CFS (*Correlation-based Feature Selection*) in one of the feature extraction methods that can perform the selection automatically. CFS identifies a relevant feature, which means the features do not require highly dependent on other feature. Features that will be selected in this research are GLCM texture feature.

Based on the explanation above, it is necessary for the researcher to conduct research about feature selection in a pattern of Sambas’s traditional fabric “Kain Lunggi” to increase the accuracy of Kain Lunggi pattern recognition system which in return can help to identify various Kain Lunggi pattern.

2. METHODS

This research will use 5 patterns of Kain Lunggi. Each class contains 30 images which mean the total images count are 150 images. Image acquisition conducted by a camera. Each pattern was photographed 30 times continuously with a fix position and good illumination. 5 types of Kain Lunggi pattern used in this research are shown in Figure 1.



Figure 1 Kain lunggi pattern type (a.bunga kangkung, b.bunga tabur, c.rantai, d.zigzag, e.sapar perangi)

The proposed method in this research started with image acquisition, in which the result will be used as input. Afterwards, the process will be followed by pre-processing, segmentation, feature extraction, feature selection, and classification accuracy evaluation. The complete steps are shown in Figure 2.

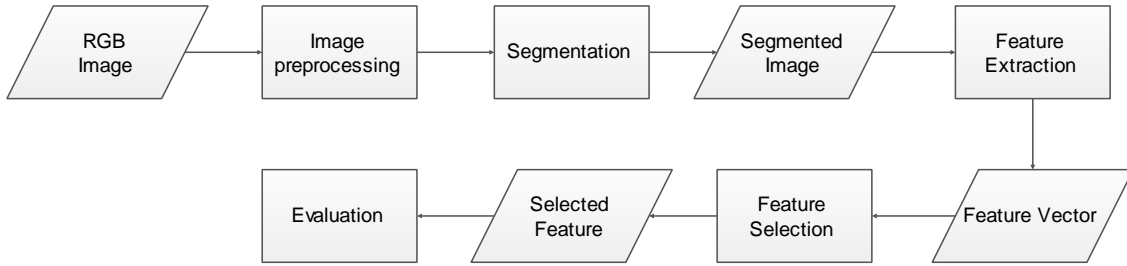


Figure 2 Steps of Kain Lunggi feature extraction research

2.1 Preprocessing

Kain Lunggi image acquired in the image acquisition process is RGB image. Segmentation process with canny edge detection requires grayscale image as its input so it needs to be transformed into Grayscale. Figure 3 shown the result of the grayscale transformation to change an image from RGB colour to grayscale. *Grayscale transformation* conducted using formula (1)[8].

$$R(x,y) = \alpha.R + \beta.G + \gamma B \tag{1}$$

where x and y are the coordinates of the pixels. $R(x,y)$ is the grayscale value resulted from the conversion. In this research, we used $\alpha=0.299$, $\beta=0.587$, $\gamma=0.114$



Figure 3 Grayscale transformation result(a) RGB Image (b) Grayscale image

2.2 Segmentation

Image segmentation is a process to separate pixels that was part of the object and pixel outside of the object [8]. The purpose of image segmentation is to separate the fabric pattern (foreground) with the background. The segmentation was conducted using canny edge detection. Figure 4 shown the flowchart of canny edge detection.

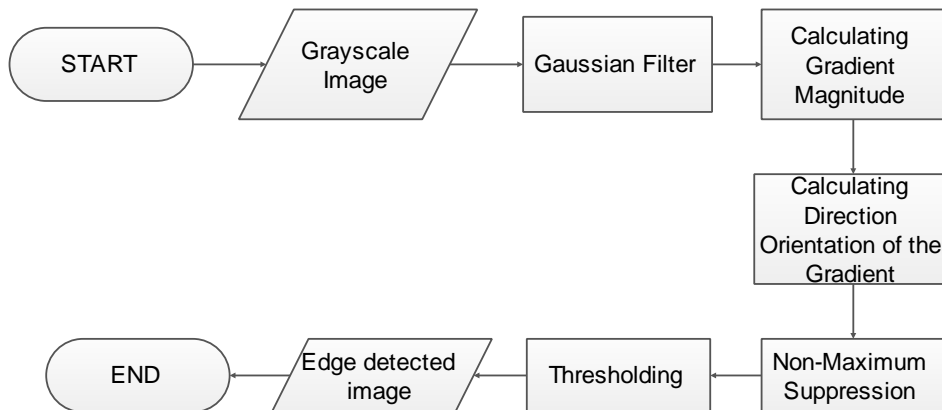


Figure 4 Canny edge detection flowchart

In Figure 4 is shown that canny edge detection started with the input of the grayscale image. The next step is applying the Gaussian filter followed by gradient magnitude calculation using the Sobel operator. In this research we used Gaussian $\sigma = 2$. These steps, followed by edge direction calculation. *Non-maximum suppression* is a process to calculate the non-maximum value in which intensity level that is not maximum will be set to 0. The next step is thresholding that used 2 thresholds, which is T_{min} (low valued *threshold*) = 2.5 and T_{max} (high-value *threshold*) = 7.5. The use of double threshold is done so that edges with values greater than the max threshold are marked as strong points, whereas edges with values smaller than min threshold will be marked as weak edges so that only edges with strong values will be maintained. The threshold value that will be used at the low threshold and high threshold will be done by trying the threshold value that can produce the best edge detection. *Threshold* formula is displayed in formula (2).

$$g(x, y) = \begin{cases} 1, & \text{iff } T_{min} \leq f(x, y) \leq T_{max} \\ 0, & \text{iff else} \end{cases} \quad (2)$$

In which $g(x, y)$ is pixel intensity from edge detection process and $f(x, y)$ is edge pixel resulted in non-maximum suppression. Result of canny edge detection is shown in Figure 5.

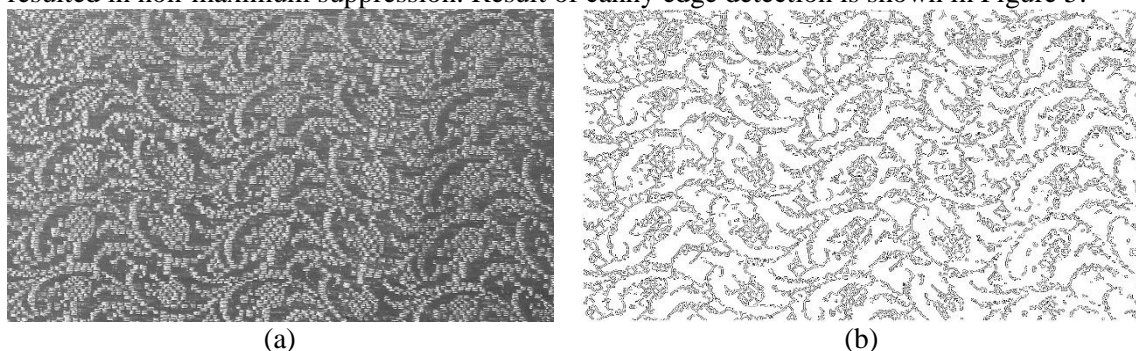


Figure 5 Canny edge detection result(a) *grayscale* image (b) line that represents fabric pattern

2.3 Texture Feature Extraction

GLCM is used to extract texture feature. This research used 4-direction GLCM with 0° , 45° , 90° , 135° angle. 4-direction GLCM will produce 4 features which are the *angular second moment*, *contrast*, *correlation*, *inverse different moment* and *entropy*. Average of all those features will be calculated to produce the feature vector.

GLCM is a matrix with elements is a summation of a paired pixel that has a certain luminance level, in which paired pixel are separated by distance (d) and inclination angle of θ . In conclusion, the co-occurrence matrix is the probability of grey level occurrence i and j between 2 pixels separated by distance d and θ angle. Figure 6 illustrates the spatial relationships of pixels that are defined by this array of offsets, where D represents the distance from the pixel of interest.

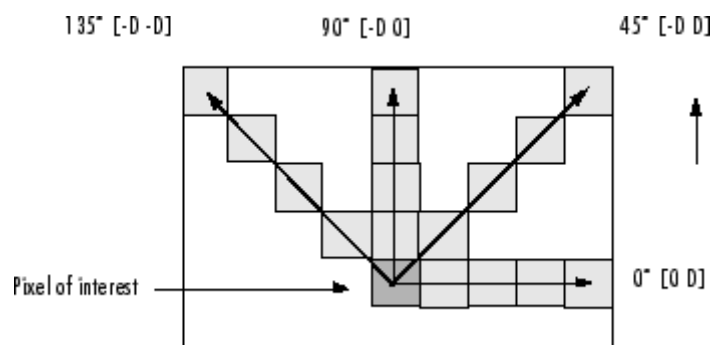


Figure 6 the spatial relationships of pixels [9]

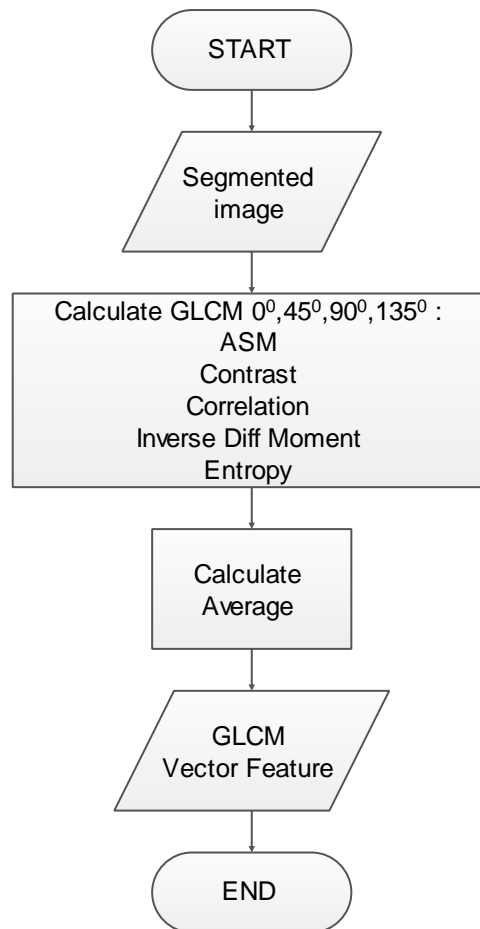


Figure 7 GLCM vector feature extraction process

Figure 7 is a process of extracting GLCM features. The results of the segmentation process are used as input to the GLCM feature extraction process. The extraction process begins with the formation of a co-occurrence matrix by defining the direction θ and distance r . The occurrence matrix is a matrix of size $L \times L$ (L denotes the amount of grey level) with elements $P(x_1, x_2)$ which is a joint probability distribution of pairs of points with gray levels x_1 located at coordinates (j, k) and x_2 located at coordinates (m, n) . The coordinates of the pairs of points are spaced with an angle θ [10]. The second level histogram $P(x_1, x_2)$ is calculated by equation (3) as follows:

$$P(x_1, x_2) \approx \frac{\text{count of paired pixel with luminance level } x_1 \text{ and } x_2}{\text{total count of pixel in an image}} \quad (3)$$

Equations (4), (5), (6), and (7) are the rule for the co-occurrence of paired pixel with angles $0^\circ, 45^\circ, 90^\circ, 135^\circ$ at distance r [10].

$$P_{0^\circ, r}(x_1, x_2) = \left\{ \begin{array}{l} (j, k), (m, n) \in R: \\ j - m = 0, |k - n| = r, \\ f_{j, k} = x_1, f_{m, n} = x_2 \end{array} \right\} \quad (4)$$

$$P_{45^{\circ},r}(x_1, x_2) = \left\{ \left((j, k), (m, n) \right) \in R: \right. \\ \left. (j - m = r, |k - n| = -r) \text{ OR } (j - m = -r, k - n = r), \right. \\ \left. f_{j,k} = x_1, f_{m,n} = x_2 \right\} \quad (5)$$

$$P_{90^{\circ},r}(x_1, x_2) = \left\{ \left((j, k), (m, n) \right) \in R: \right. \\ \left. |j - m| = r, k - n = 0, \right. \\ \left. f_{j,k} = x_1, f_{m,n} = x_2 \right\} \quad (6)$$

$$P_{135^{\circ},r}(x_1, x_2) = \left\{ \left((j, k), (m, n) \right) \in R: \right. \\ \left. (j - m = r, k - n = r) \text{ OR } (j - m = -r, k - n = -r), \right. \\ \left. f_{j,k} = x_1, f_{m,n} = x_2 \right\} \quad (7)$$

After the co-occurrence matrix is formed, the GLCM statistical features can be calculated, which are an angular second moment (ASM), contrast, correlation, inverse different moment (IDM) and entropy for each direction. Feature calculations are performed using equations (8), (9), (10), (11), and (12) [11].

$$ASM = \sum_i \sum_j P_d^2(i, j) \quad (8)$$

$$Contrast = \sum_i \sum_j (i - j)^2 P_d(i, j) \quad (9)$$

$$Correlation = \sum_i \sum_j \frac{ij P_d(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (10)$$

$$IDM = \sum_i \sum_j \frac{1}{1 + (i - j)^2} P_d(i, j) \quad (11)$$

$$Entropy = - \sum_i \sum_j P_d(i, j) \log(P_d(i, j)) \quad (12)$$

After obtaining all the feature values for each direction, the next step is to calculate the average statistical feature values in all directions formed. The result will be a feature vector of the angular second moment (ASM), contrast, correlation, inverse different moment (IDM) and entropy.

2.4 CFS (correlation-based features selection)

CFS is one of the optimization methods for attribute selection. The method used is to calculate and compare the level of correlation between each variable with its class variable and the attribute itself. CFS identifies relevant features, meaning that there is no strong dependence on other features. Feature selection can improve classification accuracy. Heuristic techniques became the basis of the CFS method to determine the value of the feature subset [12]. This technique considers the use of individual features for class label estimates with the level of intercorrelation between features. The features individually test which size is related to the observed variable (as the target class). Formula (13) shown heuristic value normalization:

$$Merit_s = \frac{k\bar{r}_{cf}}{\sqrt{k+k(k-1)r_{ff}}} \quad (13)$$

In which Merits is a heuristic value of feature S subset that contains k-rfc features which is the average feature-class correlation, rff is the average of the inter-correlation from features to features. The numerator of Equation 13 can be thought of as providing an indication of how predictive of the class a set of features are; the denominator of how much redundancy there is among the features. We used following steps to conduct features selection :

1. Compute feature correlation

We used Symmetrical Uncertainty to measure correlations defined by Equation (14)

$$SU = 2.0 \times \frac{gain}{H(Y) + H(X)}$$

in which:

H = Entropy of the attribute

X = feature 1

Y = feature 2

SU compensates information gain's bias toward features with more values and normalizes its value to range of [0,1] with 1 showing that knowledge of either one completely predicts the value of other and 0 shows that X and Y are independent. It considers pair of features symmetrically. Entropy based measures require nominal features, but they can be applied to measure correlations between continuous features as well if they are discretized properly

2. Calculate Merits value using Equation (13)
3. Perform Merits value on every possible combination of chosen feature to find highest Merits Value

2.5 Evaluation

The purpose of this step is to evaluate the features that resulted in the feature selection process with CFS. The classification accuracy of the selected feature will be evaluated with KNN classification algorithm. Evaluation conducted using *k-fold cross-validation*. In this research, *5-fold cross-validation* was used. 150 images were distributed into *5-fold*. Data distribution into 5-fold shown in Table 1 [13]

Table 1 Data distribution in *k-fold cross-validation* (k=5)

Fold	Testing Data	Training Data
1	1	2,3,4,5
2	2	1,3,4,5
3	3	1,2,4,5
4	4	1,2,3,5
5	5	1,2,3,4

There are five groups in 5-fold cross-validation where each class has 30 image data. In Table 1, the division of data for the 1st fold is the 1st group data from each class will be test data, while the 2nd, 3rd to 5th data for each class will be the training data and so on for the 2nd fold up to 5th. In total, there are 120 training data, and there are 30 test data. The training and testing process in 5-fold cross-validation was repeated five times using different test data portions in each iteration.

Evaluation scenarios are performed by comparing the results of KNN accuracy calculations with all features and KNN accuracy calculation results using CFS selected features. The K value that will be tested is K = 1,2,3,4, and 5.

3. RESULTS AND DISCUSSION

This part will discuss the feature evaluation result based on its classification accuracy. For each angle of 4-direction GLCM ($0^0, 45^0, 90^0, 135^0$), feature extraction results in 4 features which are the *angular second moment*, *contrast*, *correlation*, *inverse different moment* and *entropy*. Afterwards, the average was calculated for each feature, resulted in the GLCM feature vector shown in Table 2:

Table 2 GLCM feature average for all class

Angular Second Moment	Contrast	Correlation	Inverse Difference Moment	Entropy	Class
0.2945	76.434.067	0.00005791	0.5999	29.201	BUNGA_KANGKUNG
.....
0.5619	46.295.419	0.00009362	0.7805	17.862	BUNGA_TABUR
.....
0.2576	66.270.285	0.0000647	0.5737	30.249	RANTAI
.....
0.2941	57.542.352	0.00008189	0.6042	2.871	ZIGZAG
.....
0.2245	95.285.629	0.00004317	0.5366	32.842	SAPAR PERANGGI

First testing scenario conducted to observe classification accuracy using all GLCM features without performing feature selection. Evaluation conducted using *k-fold cross-validation*, $K=5$. The training and testing process in 5-fold cross-validation was repeated 5 times using different test data portions in each iteration. The results of the test can be seen in Table 3.

Table 3 Evaluation result before feature selection

KNN	Accuracy	Recall	Incorrectly Classified Instances
K=1	85.18 %	0.852	14.81%
K=2	74.81%	0.748	25.18 %
K=3	54.07%	0.541	45.92 %
K=4	36.29%	0.363	63.70 %
K=5	36.29%	0.363	63.70 %

Table 3 explains the acquisition of feature evaluation values before undergoing the selection process with CFS. The calculation is done using the KNN classification method. The results obtained are accuracy value, recall, and percentage of data misclassification. The highest accuracy value of 85.18% is produced using $K=1$, where the percentage of misclassification of data also results in the smallest value of 14.81%.

Accuracy can be improved by feature selection. Feature selection performed using the CFS method yields a percentage correlation of features used. The results of feature selection with CFS can be seen in Table 4.

Table 4 Feature formed with CFS

Feature	Value	Rank
<i>Angular Second Moment</i>	0.3934	1
<i>Contrast</i>	0.3891	2
<i>Correlation</i>	0.3873	3
<i>Inverse Difference Moment</i>	0.1749	4
<i>Entropy</i>	0.0655	5

From table 4, the results of feature selection using CFS produce the ranking of the selected features that have the most correlation with other features. Features are ranked from the strongest correlation to the least correlated. The next test is using 3 features that have the highest ranking which are Angular Second Moment, Contrast, and Correlation.

The third scenario test is to measure classification accuracy using the Angular Second Moment, Contrast, and Correlation features generated from CSF feature selection. The results of this test are as shown in Table 5:

Table 5 Evaluation result after feature selection

KNN	Accuracy	Recall	Incorrectly Classified Instances
K=1	88.89%	0.889	11.11%
K=2	78.52%	0.785	21.48%
K=3	57.03%	0.570	42.97%
K=4	57.03%	0.570	42.97%
K=5	57.03%	0.570	42.97%

Table 5 shows that the highest accuracy reached by KNN K=1 that has an accuracy value of 88.89. Comparison of classification accuracy for all features before and after CFS shows in Table 6:

Table 6 Classification accuracy before and after feature selection with CFS

KNN	Accuracy before feature selection	Accuracy after feature selection	Classification performance improvement (%)
K=1	85.18 %	88.89%	3.71%
K=2	74.81%	78.52%	3.71%
K=3	54.07%	57.03%	2.96%
K=4	36.29%	57.03%	20.74%
K=5	36.29%	57.03%	20.74%

Table 6 shows the percentage increase in accuracy performance after the feature selection process. Performance improvement occurred in all K values in KNN, but a significant increase occurred at K = 4 and K = 5 at 20.74%. This shows that the features selected from the CFS feature selection results can improve the accuracy of the classification system in Sambas fabric pattern recognition system.

4. CONCLUSIONS

Based on this research, it can be concluded that:

1. The feature extraction method used is the GLCM method. From the feature extraction, a feature selection process will be performed using the CFS (Correlation-based Feature Selection) method. Selected features from the CFS process are Angular Second Moment, Contrast, and Correlation.
2. Evaluation of selected features is done by measuring classification accuracy using the KNN method. The result of classification accuracy before feature selection is 85.18% using the value of K = 1, while after feature selection, the accuracy increases to 88.89%. The highest accuracy improvement of 20.74% resulted from KNN with a value of K = 4.
3. Feature selection with the CFS method increases classification accuracy by 20,74%.

REFERENCES

- [1] I. W. Fajar, "Museum tenun songket sambas," *Jurnal Online Mahasiswa S1 Arsitektur UNTAN*, vol. 4, no. September 2016, pp. 19–32, 2014.
- [2] N. Arwanda., Agani, "Content Based Image Retrieval Batik Tradisional Yogyakarta Dengan Ekstrasi Ciri Berdasarkan Tekstur Filter Gabor Wavelets 2D Skripsi Content Based Image Retrieval Dengan Ekstrasi Ciri Berdasarkan Tekstur Filter Gabor Wavelets 2D," *Ticom*, vol. 1, no. 3, pp. 12–18, 2013.
- [3] A. Haris Rangkuti, A. Harjoko, and A. E. Putro, "Content based batik image retrieval," *J. Comput. Sci.*, vol. 10, no. 6, pp. 925–934, 2014.
- [4] A. A. Kasim, R. Wardoyo, and A. Harjoko, "The selection feature for batik motif classification with information gain value," *Commun. Comput. Inf. Sci.*, vol. 788, pp. 106–115, 2017.
- [5] M. Sholihin, S. Mujilahwati, R. Wardhani, F. Teknik, and U. I. Lamongan, "Classification of Batik Lamongan Based on Features of," vol. 9, no. 1, pp. 25–32, 2017.
- [6] R. T. Wahyuningrum and I. A. Siradjuddin, "An efficient batik image retrieval system based on color and texture features," *J. Theor. Appl. Inf. Technol.*, vol. 81, no. 2, pp. 349–354, 2015.
- [7] Y. A. Sari, R. K. Dewi, and C. Fatichah, "Seleksi Fitur Menggunakan Ekstraksi Fitur Bentuk, Warna, Dan Tekstur Dalam Sistem Temu Kembali Citra Daun," *JUTI J. Ilm. Teknol. Inf.*, vol. 12, no. 1, p. 1, 2014.
- [8] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1992.
- [9] C. Mathworks, "Computer Vision System Toolbox™ User 's Guide R 2015 b," *Mathwork Inc.*, 2015.
- [10] D. Putra, "Sistem Biometrika," *Yogyakarta Andi Offset*, 2009.
- [11] F. Albrechtsen, "Statistical Texture Measures Computed from Gray Level Co-occurrence Matrices," ... *Lab. Dep. Informatics, Univ. ...*, 2008.
- [12] M. Hall and L. Smith, "Feature Selection for Machine Learning : Comparing a Correlation-based Filter Approach to the Wrapper CFS: Correlation-based Feature," *Int. FLAIRS Conf.*, 1999
- [13] Kusumadewi. Sri, *Pengantar Jaringan Syaraf Tiruan*, Yogyakarta :Teknik Informatika FT UII, 2010.