Classification of Troso Fabric Using SVM-RBF Multi-class Method with GLCM and PCA Feature Extraction

Dinda Aulia Gustian [1], Nur Laelatul Rohmah[2], Guruh Fajar Shidik[3], Ahmad Zainul Fanani[4], Ricardus Anggi Pramunendar[5], Pujiono[6]
Faculty of Computer Science
University of Dian Nuswantoro
Semarang, Indonesia
e-mail: [1]daulia810@gmail.com, [2]lailarodar@gmail.com, [3]guruh.shidik@dsn.dinus.ac.id, [4]a.zainul.fanani@dsn.dinus.ac.id, [5]ricardus.anggi@dsn.dinus.ac.id, [6]pujiono@dsn.dinus.ac.id

Abstract— **Image Processing has many benefits that we have seen, such utilization is one of them to help analyze images by classification, image pattern recognition, data security, and copyright protection. Troso fabrics are classified using a combination of two methods. Using the Gray Level Co-occurrence Matrix (GLCM) and Principle Component Analysis (PCA) method for feature extraction and multiclass Support Vector Machine (SVM) used is Ones Against All (OAA) & Ones Against One (OAO) with the type of gaussian kernel or Radial Basis Function (RBF) as a classification method. Testing of the two methods was carried out on 3 types of troso fabric using images measuring 480x480 pixels. From the results of tests that have been carried out, the classification of troso fabric produces an accuracy of 90% for SVM OAA and 86,7% for SVM OAO for the best method feature extraction using GLCM.**

Keywords— torso fabric; classification; feature extraction; svm multi-class; pca; glcm.

I. INTRODUCTION

Troso is the name of a village located in Pecangaan sub-district, Jepara Regency. Starting from a community of torso weaving craftsmen, in this village the beginning of the formation of Troso fabrics. Troso fabric is a technique called gedok weaving which has developed into a long weaving, but the stigma of the people of Jepara Regency & surrounding areas is better known as "Tenun Troso" (R.A Kartini Museum). Weaving Troso has many features and patterns for each type of fabric. Weaving that has a different pattern will produce different traits.

Image processing is the study of algorithms by the process of taking images as input and producing images as output [1]. Image processing and pattern recognition are used for the introduction of object characteristics through image enhancement, characteristic acquisition, image segmentation, grouping analysis, and data compression of images obtained. Therefore, acquisition characteristics are very important in image processing, image pattern recognition, data security, and copyright protection [2]. In the image processing classification can be used to classify an object according to the class specified before, in this study using Weaving Troso.

Blanket (a) Rangrang (b) Baron (c)

Fig 1. Troso Weaving Pattern

In figure 1 shows several patterns from Weaving Troso. Weaving Blanket (1.a) has motifs inspired by animals and humans. Fabric Troso Rangrang (1.b) has a zigzag-shaped woven fabric motif. Whereas the baron torso woven fabric (1.c) has the motif of squid kaliman, the motif of Kalimantan stocking and weaving barons have a tie line.

II. RELATED WORK

Wildan et al [3] this research utilized two patterns of Batik which will being compared with the combination of the median filter using the FCM and then be clustered with the algorithm FCM so that both MSE and PSNR can be achieved. The best result is Batik 2 cluster 2 value MSE 0.2917, PSNR 4.9258 whereas batik 2 cluster 1 value MSE 0.2917, PSNR 5.3508.

Teny Handhayani [4] VFI5 was successful in classifying images of hand painted Batik Lasem that have the same characteristics of the main features. Maximum accuracy

978-1-7281-3832-9/19/$31.00 ©2019 IEEE

without feature selection is only around 62%. The experimental results show that the selection of features based on feature variants increases the average accuracy. After applying feature selection using Mean Absolute Deviation, the maximum accuracy is 99.96%.

Alireza Soleimanipour et al [5], two main steps in object recognition systems involve classification and representation of features. In this study, a combination of PCA, LDA and SVM was used to developed a classification system for cultivars for Anthurium flowers, where PCA, LDA, and SVM were applied for data reduction, feature extraction, classification, and respectively.

Christian Sri Kusuma Aditya et al [6], Batik has many variations of motifs based on the region. the use of Gray Level Co-occurrence Matrix (GLCM) and statistical RGB color features can represent more characteristics in extracting batik images. The vector features extracted are further classified into motives using the Backpropagation Neural Network. Some experiments using a single feature and a combination of GLCM and RGB color statistics feature show that the best results for classifying batik images are a combination of feature extraction with an accuracy rate of 90.66% 94% recall and 94% accuracy.

Ratri Enggar Pawening et al [7], This research focuses on the classification of textile images based on their texture. Each texture in textile images has certain characteristics that can be distinguished from other motifs. The feature extraction method used in this study was Gray Level Co-occurrence Matrix (GLCM), Binary Linear Pattern (LBP), and Moment Invariant (MI). Furthermore, all texture features are then reduced using Principal Component Analysis (PCA). Experiments show that the best results can be achieved by using a combination of GLCM and LBP features with an accuracy of 74.15% using a 94% linear SVM recall kernel and 94% accuracy.

III. METHODOLOGY

In this section, fundamental knowledge of the methods applied in this study was introduced. In this study we separate the dataset into 2 namely training data and testing data, the data is preprocessing to improve the quality of the image. After the data is preprocessed into the feature extraction stage, the feature extraction used is GLCM and PCA. When obtaining feature extraction values, then enter the classification stage by using the Multi-class SVM algorithm method used in OAA and OAO.

3.1 Dataset

System testing is done using the image of 3 types of troso fabric, namely Blanket, Baron, and Rangrang with a total of 120 images including 90 training data images and 30 test images. Troso fabric image taking in torso craftsmen in Jepara alone using a Nikon D80 camera.The testing process is done using MATLAB R2015b on the ASUS Intel (R) Core ™ i5-7200U @ 2.50GHz 2.71 GHz Laptop and 4.0 GB RAM running under Windows 10 Pro 64 bit.

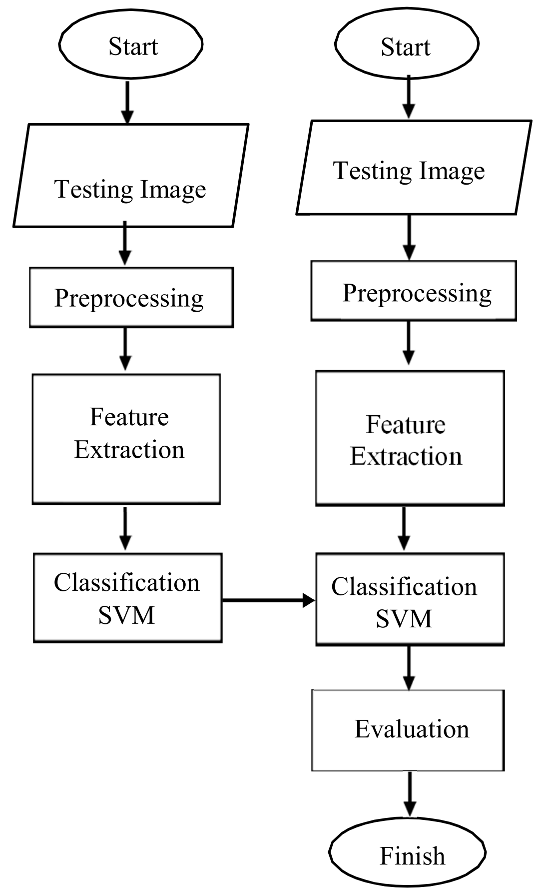


Fig 2. Proposed Method Flow Chart

3.2 Processing Data

In training, the training image is cropped to be equalized to 480x480 pixels using Photoshop CS6.

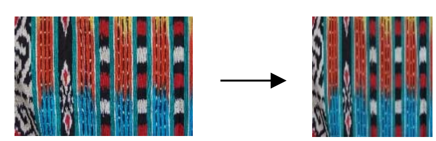


Fig 3. Process of cropping an image

Next, the RGB to Gray image is converted using the rgb2gray function in Matlab.

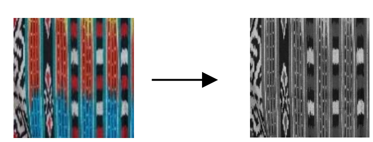
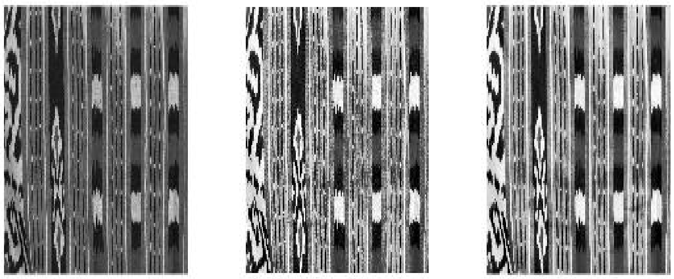


Fig 4. Converting RGB images to grayscale

After the image is converted, followed by image enhauncement to improve the quality of the converted image using the equalization histogram using the histeq function in MATLAB. Then use the Median Filter to reduce image noise.



(a) (b) (c)

Fig 5. greyscale image (a) histeq function (b) median filter (c)

After the preprocessing process is complete, the next process is feature extraction using PCA.

3.3 Feature Extraction

3.3.1 Principle Component Analysis (PCA)

PCA is generally used to reduce the dimensions of the dataset but retains most of the original variability in the data. PCA converts image vectors into subspaces or "feature spaces" and is a way to express data to highlight similarities and differences, which facilitate identification of patterns in data. In fact, it distinguishes input images into several classes [3]. PCA can improve performance in classification, save memory space, improve accuracy, and reduce excessive installation problems. Another advantage of PCA is that PCA is good for use in various types of classifications, especially classifications that use distance measurements [8].The steps of the PCA technique include preparing data, reducing the average, calculating the covariance matrix, calculating the eigenvector and the eigenvalue of the covariance matrix and finally choosing the main component.

3.3.2 Gray Level Co-occurrence Matrix (GLCM)

GLCM is one of feature extraction methods of image in which it can be interpreted in a matrix that can detect special nature of gray-level distribution in image texture [1]. This can be measured by pixels, the probability of occurrence together with the level of ash, and pixels with gray levels, the level of occurrence of the joint level of gray can be represented in the matrix of each element. This function can be read as the probability of pixel co-occurrence with gray level distance from gray-level pixels, towards 0 [6]. GLCM illustrations for each batik motif, the shared appearance matrix is calculated from four directions, namely horizontal, vertical, diagonal 450 and diagonal 135 degree. Some texture features for the introduction of batik motifs are obtained through second level statistical calculations.

3.4 Classification

In this study propose the PCA and GLCM method as the feature extraction and SVM as a method for the classification process.

3.4.1 Support Vector Machine (SVM)

SVM is a supervised learning algorithm that has succeeded in proving itself as an efficient and accurate classification technique for various applications. According to its supervised nature, SVM is implemented in two steps, namely training and classification. In the training step, SVM studies decision boundaries in the input space of training data that have been previously classified. In the classification step, classify input vectors according to the boundaries of the decisions studied [3]. Data for training is a set of eigenspace feature vectors obtained from the PCA algorithm.

Each test image is projected onto the same subspace and compared and classified using different SVM models. In total, two SVM models were developed with RBF kernel functions and different multi-class coding methods. The multi-class SVM coding method used includes OAA and OAO.

In the SVM method, OAA will construct a number of binary SVM k, where k is the number of classes. For each, a binary problem is created where all vectors belonging to the class are seen as positive samples, while others are seen as negative samples. Binary SVM classifications of K are then formed to separate class vectors from the others.

$$min\frac{1}{2}(w^i)^T(w^i) + C\sum_t \xi_t^i \qquad (1)$$

These pairwise probabilities are coupled into a common set of posterior probabilities. OAA decomposition converts multi-class problems into a series of binary C subtasks that can be trained by binary SVM. Let the training set contain the modified hidden status that is defined as :

$$\begin{aligned} \text{s.t } & (w^i)^T\Phi\,(x_t) + b^i \geq 1 - \xi_t^i \rightarrow y_t = i \\ & (w^i)^T\Phi\,(x_t) + b^i \geq -1 + \xi_t^i \rightarrow y_t \neq i \\ & \xi_t^i \geq 0 \end{aligned} \qquad (2)$$

Where the training dataset x is mapped to a higher dimensional plane using the kernel function $\emptyset(x)$, $(w^n)^T$ represents a vector in the feature space of training set, and b is a scalar value. The input training set x is assigned to the class that gives higher decision function value.

This OAO SVM method follows the binary classification with the following equation:

$$\frac{k(k-1)}{2} \qquad (3)$$

Each classification model is trained on data from 2 classes. Therefore, for training data taken from i and j classes, the binary classification problem is solved.

$$min\frac{1}{2}(w^{ij})^T(w^{ij}) + C\sum_t \xi_t^{ij} \quad (4)$$

After construction of classifiers, different methods can be employed for future testing such as the max-win strategy, which uses majority voting for selecting the output class of x and is computed as

$$\text{s.t } (w^{ij})^T\Phi(x_t) + b^{ij} \geq 1 - \xi_t^{ij} \rightarrow y_t = i \quad (5)$$
$$(w^{ij})^T\Phi(x_t) + b^{ij} \geq -1 + \xi_t^{ij} \rightarrow y_t \neq j$$
$$\xi_t^{ij} \geq 0$$

If the data x is entered into the training function $f(x) = (w^{ij})^T\Phi(x_t) + b$ and the result states x is the class of data i then the sound class i is added 1 The class of data x will be determined by the highest number of votes. If there are two classes with the same number of votes, the class whose index is smaller is declared as the class of data. So in this approach there are K(K-1)/2 quadratic programming problems, each of which has a 2n/k variable (n is the amount of training data).

3.5 Evaluation

Evaluation is done by using Confusion Matrix, which is a table consisting of the number of test data that is predicted to be correct and incorrect by the classification system, to determine the performance of a classification system. The training and testing process uses the feature extraction of GLCM-SVM and PCA-SVM that get the results of the level of accuracy or the success of the introduction of classification of troso types in each class.

Accuracy is formulated as follows:

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{FP} + \text{TP} + \text{TN} + \text{FN}} \quad (4)$$

We use a confusion matrix to measure the performance of the model we use by only taking accuracy. This measurement is often used for classification evaluation models. Using the confusion matrix, classifier accuracy can be calculated using equation [9].

IV. RESULTS AND DISCUSSION

This paper uses MATLAB to conduct experiments. All preprocessing, feature extraction, classification and evaluation processes are carried out in MATLAB. In this feature extraction, several PCA values are chosen, namely PC15, PC30, PC45, PC60, PC75, and PC90. These PC (Principal components) affect the dimensions of the new matrix to be compiled by PCA. The classification process starts from SVM training using training characteristic data which is in the form of principal component values which are then tested using SVM OAA & OAO.

GLCM with 4 features that are Contrast, Homogeneity, Energy, and Correlation using distance 1-5 and phase 45-135 degree. The next process after obtaining the Principal Component and GLCM on training image data is the classification using the OAA and OAO SVM with the type of Gaussian Kernel or RBF.

TABLE 1. PCA-SVM

PC	ACCURACY (%)		COMPUTATIONAL TIME (Second)	
	OAA	OAO	OAA	OAO
15	43,3%	43,3%	0.6727	0.9999
30	53,3%	66,7%	0.6499	1.0191
45	70,0%	76,7%	0.6517	1.0397
60	73,3%	80,0%	0.6580	1.0017
75	73,3%	80,0%	0.6725	1.1097
90	73,3%	80,0%	0.7547	1.0149

The results of testing the use of PC values 15 to PC 90 show different accuracy results. It can be seen that testing using SVM OAA on PC 15 and PC 30 values does not provide greater accuracy than PC 60 to 90. Can be seen testing using SVM OAA on PC values 60 to PC 90 produces the highest accuracy which has the same accuracy presentation OAA 73,3% and OAO 80,0%. But it can be taken PC 60 which produces the best accuracy and faster computing time compared to PC 75 and PC 90.

TABLE 2. GLCM-SVM

D	45°		90°		135°	
	OAA	OAO	OAA	OAO	OAA	OAO
1	80,0%	80,0%	73,3%	73,3%	80,0%	80,0%
2	80,0%	80,0%	76,7%	80,0%	80,0%	80,0%
3	76,7%	80,0%	80,0%	80,0%	83,3%	83,3%
4	86,7%	83,3%	73,3%	76,7%	86,7%	83,3%
5	90,0%	86,7%	76,7%	80,0%	86,7%	86,7%

GLCM with 4 features namely Contrast, Homogeneity, Energy, and Correlation using a distance of 1 to 5 and phases of 45, 90 and 135 degrees. As shown in the table for distance 1 and phase 90 degrees is the lowest accuracy compared to the others with an accuracy of OAA and OAO 73.3%. While with a distance of 5 and 45 degrees phase shows an accuracy of 90% OAA and 86.67% OAO. This is evident from testing using GLCM values with a distance of 5, and 45 degree Phase which is able to provide the highest accuracy of 90% in OAA SVM and 86.67% using OAO SVM.

V. CONCLUSION

Troso fabrics are classified using a combination of two methods. Using the GLCM method for feature extraction and multiclass SVM used is OAA & OAO with the type of Gaussian Kernel or RBF as a classification method. Testing of the two methods was carried out on 3 types of troso fabric using images measuring 480x480 pixels. From the results of tests that have been carried out, the classification of troso fabric produces an accuracy of 90% for SVM OAA and 86,67% for SVM OAO for the best method feature extraction. For further improvement, using more training data to get better accuracy, adding prepocessing can improve accuracy for system recognition and using another classification method.

REFERENCES

[1] Y. Sari, M. Alkaff and R. A. Pramunendar, "Classification of Coastal and Inland Batik using GLCM and Canberra Distance," in *AIP Conference Proceedings*, 2018.

[2] WU and Jun-Yang, "The Application of Image Processing and Pattern Recognition to the Analysis of the Microstructure Characteristics of Synthetic Minerals," in *International Conference on Information System and Artificial Intelligence*, 2016.

[3] Wildan, D. N. Santi, A. A. Mahfudh and M. A. Soeleman, "Image Enhancement Segmentation Indonesian's Batik Based On Fuzzy C-Means Clustering Using Median Filter," *International Seminar on Application for Technology of Information and Communication,* pp. 1-4, 2017.

[4] T. Handhayani, "Batik Lasem Images Classification Using Voting Feature Intervals 5 and Statistical Features Selection Approach," *International Seminar on Intelligent Technology and Its Application,* pp. 13-16, 2016.

[5] A. Soleimanipour, G. R. Chegini and J. Massah, "Classification of Anthurium flowers using combination of PCA, LDA and," *Agricultural Engineering International: CIGR Journal,* vol. 20, no. 1, pp. 219-229, June 2018.

[6] Aditya, C. S. Kusuma, Hani'ah, Mamluatul, Bintana, R. Raaiqa, Suciati and Nanik, "Batik Classification using Neural Network with Gray Level Co-occurence Matrix and Statistical Color Feature Extraction," in *International Conference on Information, Communication Technology and System*, Surabaya, Indonesia, 2015.

[7] R. E. Pawening, R. Dijaya, T. Brian and N. Suciati, "Classification of Textile Image using Support Vector Machine with Textural Feature," in *International Conference on Information, Communication Technology and System*, 2015.

[8] Y. Sari, M. Alkaff and R. A. Pramunendar, "Iris recognition based on distance similarity and PCA," in *AIP Conference Proceedings*, 2018.

[9] R. A. Pramunendar, C. Supriyanto, D. H. Novianto, I. N. Yuwono, G. F. Shidik and P. N. Andono, "A Classification Method of Coconut Wood Quality Based on Gray Level Co-occurrence Matrices," in *International Conference on Robotics, Biomimetics, Intelligent Computational Systems (ROBIONETICS)*, Yogyakarta, 2013.