EMITTER International Journal of Engineering Technology
Vol. 5, No. 1, June 2017
ISSN: 2443-1168

Javanese Character Feature Extraction Based on Shape Energy

Galih Hendra Wibowo, Riyanto Sigit, Aliridho Barakbah

Graduate School of Informatics and Computer Engineering
Politeknik Elektronika Negeri Surabaya
Jl. Raya ITS Sukolilo Surabaya 60111, Indonesia
Telp: 6231 5947280 Fax : 6231 5946114
E-mail: gabo@pasca.student.pens.a.id, {riyanto, ridho}@pens.ac.id

Abstract

Javanese characters is part of Javanese culture, which is one of Indonesia's noble culture. However, the number of Javanese people who are able to read the letters has decreased so that there needs some effort in the form of a system that is able to recognize the characters. One solution to this problem lies in Optical Character Recognition (OCR) studies, where one of its heaviest points lies in feature extraction (which is) to distinguish each character. Shape Energy is one of feature extraction methods with the basic idea of how the character can be distinguished simply through its skeleton. Based on the basic idea, then the development of feature extraction is done based on its components to produce an angular histogram with various variations of multiples angle. Furthermore, the performance test of this method and its basic method is conducted in Javanese character dataset, which has been obtained from various images, is 387 data with 19 labels by using K-Nearest Neighbors as its classification method. Performance values were obtained based on the accuracy which is generated through the Cross-Validation process of 81.90% in the angular histogram with an angle of 20 degrees, 27.32% better performance than using Shape Energy. In addition, other test results show that this method is able to recognize rotated character with the lowest performance value of 87.73% at 180-degree rotation and the highest performance value of 96.21% at 90-degree rotation. It can be concluded that this method is able to improve the performance of Shape Energy in the form of recognition of Javanese characters as well as robust to the rotation.

Keywords: Optical Character Recognition, Javanese Character, Feature Extraction.

1. INTRODUCTION

Pattern recognition is a discipline which learns various methods and theories so that able to recognise some patterns from noisy data by extracting important features. [1]. One of the research sub-areas is character

recognition or known as OCR. In general, commonly characters are English and utilization of OCR is widely used in some areas such as customer billing, digital barcode reading, health insurance data acquisition, and information entry [2]. The world of OCR science continues to flourish, as do the characters which have been studied, such as German, Japanese, Chinese and Indian. These characters are used based on their daily use. This research continues to evolve in rarely used character, such as ancient character of culture in some region like Indonesia.

Indonesia has a very wide variety of culture. Based on Statistics Indonesia Agency (BPS) in 2010, Indonesia consists of 1340 ethnic groups and each group has at least one culture or even more [3] . One of the culture that still exists is the Javanese culture. Javanese script is a specific character used by this culture. In Yogyakarta, one of the cities in Indonesia, there are writings with these characters and can be found on street signboards, wall carvings, and more particularly historical relics. But there are not many Java people able to read Javanese letters [3]. This becomes urgent if that problem is not taken care so that Javanese character will look like a symbol or decoration without any meaning on it. One of effort to resolve this problem is using OCR implementation to recognize / read Javanese character. This solution can inform people, like tourist, that Javanese script is not only a symbol but it can be read.

In general, OCR has 3 main stages, they are preprocessing, feature extraction, and recognition. The preprocessing step is to prepare the image to be ready to extract its features and recognition stage is the process of recognizing image based on the features result. A good preprocessing process and the use of good classification methods are not sufficient to produce a good level of recognition, therefore feature extraction stages play a vital role. Govindan [2] groups two types of features at that stage, they are statistical and structural features. The statistical feature approach is the acquisition of a feature by calculating the distribution of the points that make up the object, such as Zoning, Momments, and Projection Histogram [2][4][5]. While the structural feature approach is the acquisition of features from the edge or frame of the object [6] in the form of intersections, end points, and the number of horizontal and vertical lines [5].

In this research, we try new alternatives in feature extraction process by combining both that are structurally based and continued by statistical. The form of this feature extraction is to calculate the energy of each character (explained more in system design). In the final stage, performance calculations are performed through validation sampling using the classification algorithm of K-Nearest Neighbors. Furthermore, this paper will provide exploration of the facts and knowledge of the proposed method.

2. RELATED WORKS

B. A. M. Pertiwi, et al [7] research is about the development of a system to detect Javanese character which entered by the user and converted into

alphabet letters. The Javanese script used by the researchers is the *Ngalegena* script, consisting of 20 characters. Variants on each of the Javanese script as much as 10, so the total data of learning amounted to 200 data. Furthermore, the data is done by thinning process in order to obtain the skeleton. Each letter is segmented based on clustering by using a shape independent clustering through the utilization of Single Linkage Hierarchical Method (SLHM).

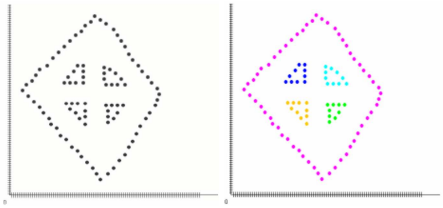


Figure 1. Cell layout in the system architecture

This method is illustrated in Figure 1. In this study, feature extraction is contributed by dividing Javanese character into small boxes with size 10x20. From this process, 200 small boxes are obtained and then the number of points that exist in each small box is counted so that there are 200 values. Those values are converted into binary form with zero thresholds. From each character, 10 variants, the value of its centroid based on binary results is calculated. The value of this centroid is matched to the test data by finding the closest distance between the centroid and test data. The form of testing in this study is to enter a new Javanese character on the device. Then the letters are performed feature extraction based on the type of point determination and the number of clusters. The number of clusters used in this study is 1, 2, 3, 4, 5.1, 5.2, 6, 7, 9, 10, and 11. The results of the experiment show the highest accuracy of 76.88% on 11 features with 10 times trials.

R. Adipranata, et al [8] research is about development of a system for the recognition of Javanese characters by using 2 methods of Evolutionary Neural Network and the combination of Chi2 with Backpropagation Neural Network. The type of Javanese script used in this study is the script Ngalegena, Sandhangan, and numbers with 15 variants for training data and 15 other variants for data testing. The input data in this research is JPEG, PNG or bitmap image. The image is normalized and resized to 40x40 pixels. The feature extraction process in this study used the ICZ-ZCZ method (Image Centroid and Zone - Zone Centroid and Zone). Each image is divided into 20 areas (4 * 5) to obtain 20 output values for ICZ and 20 output values for ZCZ. The total output value is used neuron input on Neural Network. While for combination method, data from ICZ-ZCZ method Chi2 process is done first and obtain an output value 60. So this combination method consists of 60 input value and 40 input value.

The target attributes of these two methods are 31 types corresponding to the total number of Javanese character types used in this study. There are two types of test phase in this research. The first is to test all Javanese script

and the second is to test between each type of Javanese characters, such as *Ngalegena*, *Sandhangan*, and number, on training data and data testing. In the evolutionary neural network method, this study uses percentage of crossover probability parameter by 100%, the percentage of mutation possibility 50%, maximum population 50, maximum epoch 10 million and error limit 0.1. While the combination of Chi2 and Backpropagation Neural Network uses parameters such as 1 hidden layer, 60 units hidden layer, maximum 1000 epoch, learning value of 0.1, 100 input value with target attribute of 31. The experiment is performed and by using the first method obtained 93% accuracy value using training data as test data and 60% using data testing as test data. While the second method obtained accuracy of 95% by using training data as test data and 70% by using data testing as test data. This shows that the combination of Chi2 with Backpropagation Neural Network get better result than the first method of evolutionary neural network.

G. S. Budhi, et al [9] research is to compare some methods of recognition to certain Javanese characters. These methods include Bidirectional Associative Memory, Counterpropagation Network, Evolutionary Neural Network, Backpropagation, and Chi2. While the selected types of Javanese characters are *Ngalegena*, *Sandhangan*, and numbers. The characters detected by the researcher are single or independent characters. The next step in this study is to compare the results of the detection of these characters with each chosen method. The results of experiments conducted by researchers show that bidirectional associative memory and counter propagation methods can not be used in recognizing or detecting Javanese character because the accuracy of the results are very low. The combination of Chi2 and Backpropagation produce better results than evolutionary Neural Network with 1 or 2 layers. The accuracy obtained from this method is 73% of the new data.

3. ORIGINALITY

Culture, in this case Javanese characters, is important for a State to be maintained. Today the number of people who is able to read the character is decreasing. So indirectly, it only leaves the form of its form alone or known as a symbol [3]. The purpose of this research is to utilize the field of OCR science to read Javanese characters by recognizing first. Utilization of OCR as an effort on this problem has been done by several researchers by using various types of feature extraction. In this study, we try to use new feature extraction alternatives ie Shape Energy. This method consists of three main components they are elasticity, curvature, and shape. Based on those components, furthermore, the development is done to get an approximation by multiple angles. This paper proposes a variety of different research paths from previous related works on Javanese issues, through merging ways: (1) starting with Javanese data collection (2) preprocessing in certain parts of the image, Javanese character, to be translated, (4) the formation of frames, skeleton, in each letter and extracting features using the Energy Forms

method and its development/improvement, (5) using the K-Nearest Neighbors algorithm as the method of recognition, (6) and performed Cross Validation method, to calculate the dataset performance value that has been formed. As for the result, we will get an accuracy of translation on each character. Furthermore, the results will be analyzed to find out information about the Javanese character data from the proposed method in order to develop better.

4. SYSTEM DESIGN

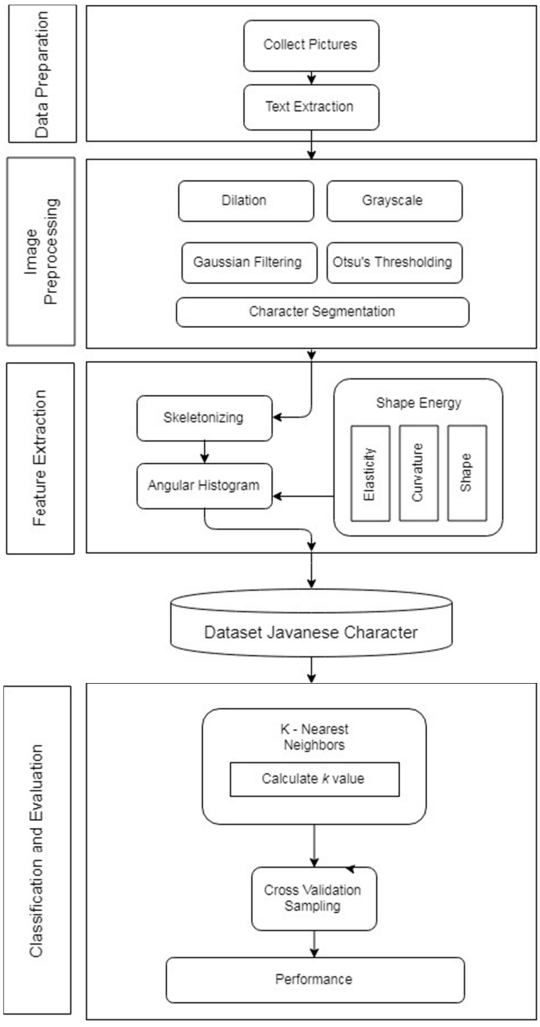


Figure 2. System design of Javanese character recognition

The proposed system consists of several steps, namely (1) Data preparation, (2) Image preprocessing, (3) Character segmentation, (4) Feature Extraction, and (5) Classification and Evaluation. The overall design of the system in this study is shown in Figure 2. Each stage of the system will be described in Section 4.1 - 4.5.

4.1 Data Preparation

4.1.1 Collect Pictures and Text Extraction

The theme of this research is Optical Character Recognition (OCR). In general, the character that recognized in OCR is mostly alphabet letters. Variants of these characters are numerous and can be found in various computer fonts and converted into images. But it is different with the character which is used in this research, Javanese character. The availability of these characters are very limited, so an effort is needed to create a Javanese character image dataset. In this case, one of the criteria in collecting data to become dataset is a standard form character, like a computer font.

In this research, the collection of images/data is conducted by collecting images containing Javanese character like a photo on an object that contains Java writing. It can be found on street signs, decorations, or the name of a place in the area of Yogyakarta. A digital camera, with 16 megaPixel resolution, is used for photo capture. The dimension of image is 1824 x 1368 pixels.

Figure 3. Example of Image Capture Result

In this research, there is a criterion for capturing the image. The image is taken in the normal or natural situation, which means the position of the camera is in front of the object or in the right/left. The distance of the camera to the object is conducted dynamically. It means the Javanese letters can be seen clearly without having to do image magnification. Examples of these images are presented in Figure 3.

Furthermore, the segmentation process is performed for letters area in a whole picture or known as the Region of Interest (ROI). In this research, the ROI segmentation of collected image was done manually. Results at this stage can be seen in Figure 4**.**

Figure 4. ROI segmentation of Javanese letter

4.2 Image Preprocessing

Preprocessing is one of the important stages in the pattern recognition of the image [10]. It aims at processing many data sets into more suitable results, by eliminating unwanted data deviations, which will be used in the classification or recognition process rather than the original data [10][11]. In addition, this stage also prepares images or objects to be ready for the next process which is feature extraction. In this research, feature extraction is based on a skeleton, a form that represents the letter itself. Therefore, we do several techniques of preprocessing in to achieve it.

The preprocessing stage begins with a dilation technique. This technique uses to strengthening the form of Javanese character in the image. The next process is to convert the RGB image format to grayscale and then the filtering process. The process of filtering is done on purpose to eliminate some information that is distorted or known as noise. This filtering process is performed by using Gaussian filter because it is a good technique in handling noise in smoothing image but still maintains the quality of object or character [11].

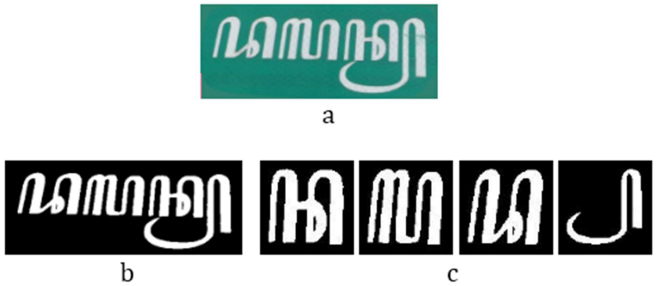


Figure 5. a. Original image, b. Binary image, and c. Segmentation result

The next process is known as binarization. The main point in this technique is the determination of the threshold value so that the image becomes black and white. We use Otsu's thresholding because this technique is very simple and it is also able to determine the threshold value automatically [12]. This stage ends with a segmentation process. We use a simple technique that is by using edge detection on the letters to be recognized. Resize process is performed on each character into 85x85 pixel. At this stage, we get 387 characters dataset with 19 class number, namely *nga*, *q*, *cecek*, *da*, *ga*, *wulu*, *ka*, *la*, *layar*, *ma*, *na*, *pangkon*, *pepet*, *ra*, *sa*, *suku*, *ta*, *taling*, and *tarung*. The result of this step can be seen in Figure 5. The variants of each character are generated from images on different objects. One example of the variant results is presented in Figure 6.

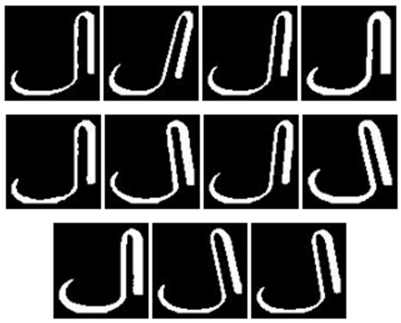


Figure 6. Sample variance images of *Pangkon*

4.4 Feature Extraction

This stage begins with the framing of the main form of character or known as thinning. In this study, we used the Zhang-Suen thinning method [13] because this method has good efficiency and is fast in obtaining the skeleton. The next process is to extract features or characteristics that are able to distinguish between letters, one of which is shape energy. In this study, we tried to improve this method by combining the basic idea of the two main components of the method, curvature and shape resulting in an angular histogram.

4.4.1 Shape Energy Concept

Shape Energy is one of the methods of feature extraction on the basis of the idea that objects can be distinguished in a simple way, that is based on the main frame or skeleton [14]. This method has 3 main components namely elasticity, curvature, and shape.

Elasticity, *el*, is a structural feature that can be obtained from the object of the forming length. The value of this feature can be obtained through the first derivative calculation on each pixel or in other words, this value can be obtained through the following formula

$$el = \sum_{i=0}^{n-1} \left[(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 \right] \quad (1)$$

where *i* is the order of pixels that make up the object, as well as *x* and *y* are the coordinates of each pixel. Thus, the first thing to do is to determine the starting point of the pixel and this rule applies to all objects to be extracted, eg the smallest *x* position and the largest *y*.

Curvatur, ec, is one feature that is capable of distinguishing objects. In this method, the calculation of curvature is done by calculating the second derivative value of each pixel and formulated as follows:

$$ec = \sum_{i=2}^{n-2} \left[(x_{i+1} - 2x_i + x_{i-1})^2 + (y_{i+1} - 2y_i + y_{i-1})^2 \right] \quad (2)$$

In this formula *i* is the order of pixels that make up the object, and *x* and *y* are the coordinates of each pixel.

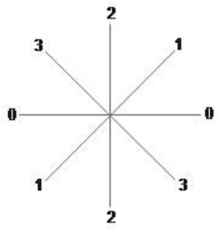


Figure 7. 4-directional chain code

The shape is the last component of the Shape Energy. This feature can be obtained by identifying the relationship between points on the skeleton. This feature uses the chain-code method. This method shows the relationship between two pixels that make up the wind direction and then formed the histogram. The type of chain-code used in the shape energy is shown in Figure 7.

4.4.2 Shape Energy Improvement Concept

In this research, we try to develop the shape energy method to produce an angular histogram. This method is derived from the basic idea of shape energy, which is the form of chain-code and curvature. Shape Energy uses 4-connectivity where each type of connectivity forms an angle of 45 degrees. Through this approach, the final feature of Shape Energy can be similar into an angle histogram of 0, 45, 90, and 135 degrees so that it updates the second major component of curvature, while a degree is formed from three points. In this case, the first point is the dots coming from the skeleton character and uses to measure an angle value towards another points, which is a static reference.

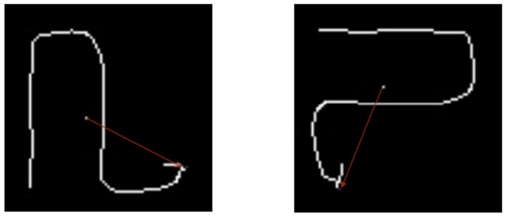
We used centroid values from the main framework as the second point on the formation of this new feature. Calculation of the value of centroid can be obtained by using the theory of moment in digital images, i.e. in the form of matrix into a two-dimensional array. The moment of order (p + q) is given by

$$m_{pq} = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} x^p y^q f(x,y), \qquad p,q = 0,1,2,\ldots,\infty, \tag{3}$$

where M and N represent the horizontal and vertical dimensions of the image, and *f (x, y)* is the intensity at a point *(x. y)* in the image. Then, the centroid coordinates can be obtained by

$$\bar{x} = \frac{m_{10}}{m_{00}}, \qquad \bar{y} = \frac{m_{01}}{m_{00}} \tag{4}$$

In the formation of this new feature, the third point of angle formation is the reference point which will not change despite the rotation of the object. The reference point is obtained from the main framing point that has the furthest distance to the centroid. The example of determining the third point is shown in Figure 8. Thus, this new feature has resistance to rotation, as one of criteria to be called as good feature extraction in OCR, compared to Shape Energy.



(a) (b)

Figure 8. Farthest point from centroid

After a three-point angular shape is defined, the last step is to determine the angle calculation. We use a vector approach that is formulated as follows:

$$\cos\theta = \frac{(\vec{u}.\vec{v})}{(\|\vec{u}\|.\|\vec{v}\|)} \tag{5}$$

In this formula (u) is the point of the skeleton against the centroid and (v) is the centroid point against the reference point. The angle values obtained from the above formula are then rounded to the nearest corner value at a multiple of 45 degrees, based on angle multiplier which is used. Based on the idea, the best angle multiples are made such as 10, 20, 30, and 40 degrees. The reason we use those values is that if the angle multiplier is too small, the difference between original values with rounding value will be too small. Conversely, if the angle multiplier is too large, then the difference will be too large. Table 1 shows the difference values in a sample angle to rounding off a multiple of angles.

Table 1. Sample of difference between original values with rounding value

Angle Values	Rounding Angle Multiplier - Difference (Diff)											
	10	Diff	20	Diff	30	Diff	40	Diff	50	Diff	90	Diff
13	10	3	20	7	0	13	0	13	0	13	0	13
33	30	3	40	7	30	3	40	7	50	17	0	33
57	60	3	60	3	60	3	40	17	50	7	90	33
73	70	3	80	7	60	13	80	7	50	23	90	17
112	110	2	120	8	120	8	120	8	100	12	90	22
176	180	4	180	4	180	4	160	16	200	24	180	4
Average		3.0		6.0		7.3		11.3		16.0		20.3

The rounded angle value is the smallest angle value, so the value is in the range 0 - 180 degrees. At that range, if rounding is done to the angle multiplier value, it produces a number of different features, this can be seen in Figure 9. The greater the multiple angles used, the smaller the number of features generated. After the rounding value is obtained, then calculate the histogram value of each feature.

Angle Multiplier	Feature																			Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
10	0	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	19
20	0	20	40	60	80	100	120	140	160	180										10
30	0	30	60	90	120	150	180													7
40	0	40	80	120	160	180														6

Figure 9. Number of feature based on each approach

The next process is to extract the feature on each image of Javanese characters using this method. This process generates 4 types of datasets with each feature type is numeric and independent with other features. The characteristics of each dataset are shown in the mean and standard deviation values of each feature, presented in

Table ***2***. Once the dataset is obtained from each of these approaches, the next step is the testing process. This process is performed to determine the performance of the model or dataset that has been formed. The testing process is conducted by using cross validation because the features on the dataset are independent.

Table 2. Characteristics of each dataset

Feature	Angle Multiplier							
	10		20		30		40	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
F1	4.22	2.56	8.28	3.91	11.64	4.72	14.65	5.14
F2	7.42	3.08	12.06	4.65	15.78	5.86	19.60	7.13
F3	5.95	2.81	9.59	4.78	13.97	6.18	21.53	11.63
F4	5.14	2.61	8.91	4.52	16.80	10.00	17.51	9.85
F5	4.68	2.92	11.71	7.05	13.10	7.74	22.73	8.14
F6	4.66	2.82	9.95	5.69	13.73	7.10	0.79	0.72
F7	4.46	2.57	8.82	5.58	11.77	4.05		
F8	4.84	2.88	8.28	5.24				
F9	6.18	4.00	10.78	5.06				
F10	5.70	3.61	8.42	3.28				
F11	4.92	2.95						
F12	4.57	2.78						
F13	4.51	3.15						
F14	4.02	2.99						
F15	4.08	3.00						
F16	4.48	3.04						
F17	5.18	2.68						
F18	7.19	3.00						
F19	4.58	2.39						

4.5 Classification and Model Evaluation

In this research, we used k-Nearest Neighbor (k-NN) as its classification method. We use this method because it is a non-parametric classification, which is straightforward and also useful in many cases [15]. This method classifies the object/data test based on the voting, known as *k*, from similarity value to sample/training data known as instance-based learning.

There are several methods can be used to measure the similarity in numerical data, like Cosine similarity [16], Canberra distance, or Euclidean distance, we choose this because its simple measurement way. The voting value or *k* determines the result of classification in this method, so it is necessary to test different *k* values, such as 1, 3, 5, and 7. In addition, we also do an evaluation of the dataset that has been formed by dividing the dataset into 2 parts of training and testing. In this research, we used cross-validation sampling in which data are divided into 10 sections. This process will produce a performance value by the accuracy of the dataset that has been formed based on chosen classification algorithm.

5. EXPERIMENT AND ANALYSIS

The experiment and analysis consist of (1) classification performance evaluation of dataset and (2) rotated image, and (3) interesting facts of feature extraction method.

5.1 Classification Performance Evaluation

We do the performance test of each feature extraction, including Shape Energy, Improvement Method of Shape Energy by using angle multiplier of 10, 20, 30, and 40, and also k-NN classification method with *k* values of 1, 3, 5, and 7. The test is applied by using 10 cross-validation sampling. The result of this stage is shown in Table 3. In this discussion, we use abbreviation such as Shape Energy method (SE), Improvement Method (IM) with 10-degree angle (IM10), multiples of 20 degrees (IM20), multiples of 30 degrees (IM30), and multiple of 40 degrees (IM40).

In this test results, the highest performance of the SE is 54.58% and has a fairly stable on *k* with 1, 3, and 5, but decreases when *k* is 7. This indicates that its feature extraction result is not sufficient to represent the uniqueness of each Javanese character. The performance of IM shows better results than previous methods.

Table 3. K-Nearest Neighbor Performance Experiments

Feature Extraction Method		Accuracy			
		k=1	k=3	k=5	k=7
Shape Energy		**54.58%**	51.28%	51.20%	46.58%
Improvement Method	Multiple of 10 degrees	76.97%	**79.55%**	76.42%	73.58%
	Multiple of 20 degrees	**81.90%**	76.98%	77.22%	73.62%
	Multiple of 30 degrees	73.32%	**74.88%**	73.10%	71.28%
	Multiple of 40 degrees	69.20%	69.96%	69.45%	**70.75%**

According to the test results, presented in Table 1, it can be observed more deeply about the results of IM performance against the number of k which is used. Based on the test results, IM40 has a stable performance value about 70% with the highest performance by using k of 7. It is also shown on IM30 with highest performance value is 74.88% by using k of 3. IM10 and IM20 show the same pattern that the performance is almost decreasing by using k with value ranging from 1 to 7. Despite the decrease, the value of the performance shown by the IM10 and IM20 better than the average performance of the IM30 and IM40. This shows that the value of the features, generated by IM10 and IM20, is good enough, in the sense of being able to distinguish the uniqueness of each Javanese character. The best performance result is shown on IM20 that is 81.90% with k value of 1, known as Single

Nearest Neighbor. So IM20 manages to improve the performance of the SE about 27.32%.

5.2 Performance of Rotated Image

In this research, we also do performance test on rotated images in order to prove the improvement method (IM) can be resistant to a rotation, in accordance with the theory in the previous chapter. The test is conducted by taking samples from the random image dataset in which there are 48 images in total. The samples data consist of 11 classes/character which are *cecek*, *da*, *ga*, *ka*, *la*, *layar*, *na*, *pangkon*, *sa*, *ta*, and *taling*.

Figure 10. Sample Image (a) Image of *'ka'*, and others is rotated image of (a) by (b) 90, (c) 180, and (d) 270.

Figure 8 shows the rotation of 90 degrees, 180 degrees, and 270 degrees. The test is conducted by using IM20 as its feature extraction, k-NN with k value of 1, and class precision calculation as its performances. The test results are presented in Table 4.

Table 4. Performance of Rotated Images

Character / Class	Instance	Class Precission of Rotated Image		
		90 degrees	180 degrees	270 degrees
Cecek	3	100.00%	75.00%	66.67%
Da	6	100.00%	50.00%	50.00%
Ga	4	75.00%	80.00%	66.67%
Ka	5	100.00%	100.00%	100.00%
La	5	100.00%	100.00%	100.00%
Layar	4	100.00%	100.00%	100.00%
Na	6	100.00%	80.00%	83.33%
Pangkon	5	100.00%	100.00%	100.00%
Sa	3	100.00%	80.00%	100.00%
Ta	4	83.33%	100.00%	100.00%
Taling	3	100.00%	100.00%	100.00%
Average		96.21%	87.73%	87.88%

Based on the above experimental results, it can be concluded that IM, especially IM20, is a feature extraction that has a resistance to rotated image. This is based on the average value of performance obtained from the test

data that is worth more than 87.73%. The average performance value shown in the image is rotated 180 degrees with the lowest value with 50%, i.e. at *da*, followed by the character *cecek* with 75%. However, it can be noted that in other character data, the majority is well recognized, ranging from 80% to 100%. The images neither rotates in 90 and 270 degrees. In other hand, overall of the result presents a good performance value with an average value of 90.61%.

5. Conclusion

Based on the above experiments, we may conclude that the feature extraction as alternatives in the form of an improvement method based on Shape Energy are successfully performed and are able to distinguish between Javanese characters quite well. The method is based on a combination of curvature energy and shape approach by the angle which 20 degrees multiplier. The performance improvement was successful with 81.90% achievement, 27.32% better than Shape Energy. In addition, the method possesses resistance to rotated images. This is shown in the test results of rotated images successfully identified with a performance value of more than 87.73%.

REFERENCES

[1] Vinita Dutt, Vikas Chaudhry, and Imran Khan, **Pattern recognition: an overview**, American Journal of Intelligent Systems, Vol 2.1, pp. 23-27, 2012.

[2] V.K Govindan, A.P Shivaprasad, **Character Recognition - A Review,** *Pattern Recognition*, Vol. 23, pp. 671–683, 1990.

[3] A. M. Sulaiman, **HANACARAKA : Javanese Script that Began to be Abandoned**, Master, ISI(Surakarta), 2011.

[4] M. Hossain, M. Amin, and H. Yan, **Rapid Feature Extraction for Optical Character Recognition,** *arXiv preprint arXiv:1206.0238*, 2012.

[5] O. D. Trier, A. K. Jain, and T. Taxt, **Feature Extraction Methods for Character Recognition - A Survey**, *Pattern Recognition*, Vol. 29, no. 4, pp. 641–662, 1996.

[6] Mozaffari, Saeed, Karim Faez, and Majid Ziaratban, **A Hybrid Structural/Statistical Classifier for Handwritten Farsi/Arabic Numeral Recognition,** *MVA,* pp. 218-221, 2005.

[7] B. Ariani, M. Pertiwi, A. R. Barakbah, and Y. Setiowati, Pengenalan Pola Penulisan Kata Aksara Jawa pada Piranti Bergerak dengan Shape Independent Clustering**,** D4 Final Project, EEPIS(Surabaya), 2013.

[8] Rudi Adipranata, **Java Characters Recognition using Evolutionary Neural Network and Combination of Chi2 and Backpropagation Neural Network**, Disertation, Petra Christian University, 2014.

[9] Gregorius Satia Budhi, and Rudy Adipranata, **Handwritten Javanese Character Recognition using Several Artificial Neural Network Methods**, *Journal of ICT Research and Applications*, Vol. 8, No. 3, pp.

195–212, 2015.
[10] Bow, Sing T, **Pattern recognition and image preprocessing**, *CRC press,* 2002.
[11] Sonka, Milan, Vaclav Hlavac, and Roger Boyle, **Image Processing, Analysis and Machine Vision**, 2014.
[12] Otsu, Nobuyuki, **A Threshold Selection Method from Gray-Level Histograms,** IEEE Transactions on Systems, Man, and Cybernetics, Vol 9, No. 1, pp. 62-66, 1979.
[13] Chen, Yung-Sheng, and Wen-Hsing Hsu, **A Modified Fast Parallel Algorithm for Thinning Digital Patterns**, *Pattern Recognition Letters*, Vol. 7, No. 2, pp. 99–106, 1988.
[14] Wibowo, Galih Hendra, Riyanto Sigit, and Aliridho Barakbah, **Feature Extraction of Character Image using Shape Energy**, *International Electronics Symposium,* pp. 471–475, 2016.
[15] Hand, David J., Heikki Mannila, and Padhraic Smyth, **Principles of Data Mining**, *MIT press*. 2001.
[16] Desi Amirullah, Ali Ridho Barakbah, and Achmad Basuki. **Semantic Songket Image Search with Cultural Computing of Symbolic Meaning Extraction and Analytical Aggregation of Color and Shape Features**, *EMITTER International Journal of Engineering Technology*, Vol 3, No. 1, 2015.