Pattern Recognition Of Javanese Letter Using Template Matching Correlation Method

Irham Ferdiansyah Katili[1], Fairuz Dyah Esabella[2]
Dian Nuswantoro University, Jl. Imam Bonjol No.207 Semarang, Indonesia
E-mail : irhamferdi759@gmail.com[1], fairuzdyahesabella@gmail.com[2]

Ardytha Luthfiarta*[3]
Dian Nuswantoro University, Jl. Imam Bonjol No.207 Semarang, Indonesia
E-mail : ardytha.luthfiartha@dsn.dinus.ac.id[3]

Abstract - *In this modern age, the impact of globalization is increasingly entering and expanding into most societies. One impact of globalization makes people prefer to learn the language and use a foreign language rather than the local language, especially the Java language. It is very influential on the knowledge of the community about the existence or the existence of Javanese Letter, especially in the field of education. In this study, In this research will be made an application to recognize the writing of Javanese Letter based on Optical Character Recognition (OCR). Matching templates correlation can be used as pattern recognition methods. How the Template Matching Algorithm works by matching the template image with the test image after going through the Pre-processing and segmentation process. From the research that has been done by using 10 character template and 20 data testing get accuracy equal to 93.44% and error rate 6.56%. So the Matching Template Algorithm can well recognize the Javanese Letter pattern.*

Keywords - *Javanese Letter, OCR, Segmentation, Template Matching Correlation*

1. INTRODUCTION

Javanese Letter is one of the cultural heritage that must be preserved because it is almost extinct. Maintaining and preserving the culture of the nation is the responsibility of every citizen, as stated in article 32 of UUD 1945 on the state guarantees the freedom of the community in maintaining and developing its cultural values [1]. One impact of globalization and the introduction of foreign culture makes people prefer to learn and use English rather than Javanese. Western culture into Indonesia is caused by several things, one of which is the crisis of globalization. The western culture that goes into Indonesia is growing very fast. It has a huge impact on the social system and the culture of the community. With the rapid influence of these cultures lead to social shocks where people are not able to filter the various influences caused by incoming culture. Without realizing it can destroy the original culture of Indonesian society. But globalization not only has a negative impact on this culture. Globalization also brings the impact of rapid technological development. Technology is one of the weapons that can be used to introduce our culture and can also be used as a medium of learning [2].

The introduction of Javanese Letter to the community can be done using Digital Image Processing discipline. A field of science that studies about how images are processed, shaped and analyzed to produce understandable information. By applying the field of science that is processing the pattern, can be found information that can be used to perform pattern

recognition. Thus the Javanese Letter is easier to recognize. Some research on pattern recognition of Javanese Letter has been done among others done by Madha Christian Wibowo, I Dewa Gede Rai Mardiana, Sandy Wirakusuma in 2015, about the recognition of Javanese handwriting pattern using Multi-Layer Perceptron. The study obtained the final error value of 0.0095844230767128 [3]. Further research by Zaky Mukhoyyar about the introduction of Javanese Letter using K-Nearest Neighbor algorithm. By using the algorithm get 100% accuracy result [4]. In addition to these algorithms, there are several algorithms that are used to perform Java pattern recognition such as Self Organizing Maps, Skew Detection, Backpropagation and Template Matching. Algorithm K-Nearest Neighbor is an algorithm that is often used and quite easy to implement. The drawback of the algorithm is that it is necessary to determine the value of the parameter (k) and distance learning is not clear which one is used to getting the best result. Template Matching is a simple algorithm, efficient, and has a high level of accuracy. In addition to the advantages, there are also shortcomings of template matching algorithm that requires data train or reference data to obtain optimal results. From the comparison of algorithms, the Template Matching algorithm can cover the shortcomings of the K-Nearest Neighbor algorithm by producing a high degree of accuracy[5][6]. So that algorithm is suitable for pattern recognition in this research. From several related research results about pattern recognition, this study aims to implement a template matching algorithm to perform pattern recognition. The goal is to get the best accuracy result and to get the right conclusion.

The author will create an Optical Character Recognition (OCR) system. OCR is a computer system that can read characters including (letters), which come from printing and handwriting. The OCR system can help improve flexibility and intelligence in computer systems. In addition, OCR can help people or big efforts in recognizing ancient digital literature and ancient in Letters. Other benefits can be used in education as a media of learning, especially on the local content of Java language [7]. The method used in this OCR system is Template Matching. Matching Templates is one method of Optical Character Recognition that serves to match each part of an image with a template image. In this algorithm, the input image is matched to the existing template image in the database to look for similarities using a particular rule. Matching can be a good result determining that the image is recognized by one of the template images.

2. RESEARCH METHOD

The type of research undertaken in the writing of this final task is a type of experimental research, in order to recognize and translate the Javanese Letter using Template Matching Correlation method based on Optical Character Recognition. And build the application of Optical Character Recognition.

In this research, the Researcher will use the word of Javanese Script of Proklamasi Script, Javanese Script of Pancasila Script, Javanese Script of Tembang Dhandhanggula, Javanese Script of Tembang Kinanthi which does not contain Javanese "sandhangan" , Javanese "pasangan" and Javanese Number.
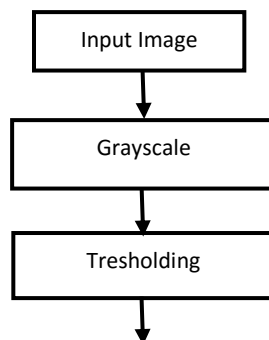
Table 1. Data Testing

No	Data Testing
1.	vtk
2.	dka
3.	al
4.	wsa
5.	grk
6.	kt
7.	ktv
8.	at
9.	sb
10.	an
11.	ma
12.	as
13.	nsy
14.	wrt
15.	ꦧꦠꦏꦩꦏ
16.	ꦏꦩꦛꦩꦏ
17.	ꦩꦸꦩꦩꦩꦩꦥ
18.	ktv
19.	vmlp
20.	lpqwr

This research is implemented by simulation using Matlab application. In this section, will discribe the process of Character Recognition using Template Matching Correlation method.

Character Recognition Process
The process of Recognition in this research is as follow:

Input Image

↓

Grayscale

↓

Tresholding

↓

```
┌─────────────────┐
│   Resize Image  │
└────────┬────────┘
         ▼
┌─────────────────┐
│Feature Extraction│
└────────┬────────┘
         ▼
┌─────────────────┐         ┌──────────────────┐
│  Classification │◄────────│ Template Matching │
└────────┬────────┘         └──────────────────┘
         ▼
┌─────────────────┐
│    Decision     │
└─────────────────┘
```
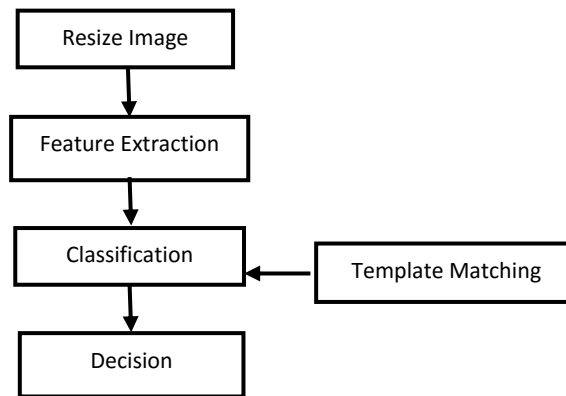
Figure 1. Process of Recognition

Based on figure 3.1, the process of Recognition of Javanese Letter using Template Matching Correlation method will be explained as follows:

1. The first step is to input image testing into the application.
   Sample image:

   Figure 2. Sample image to be tested

2. Then grayscale method, in this process color of the image testing will be changed to grayish.

   Figure 3. Grayscale Image

3. After that, the next process is Thresholding. This process, used to generate binary images. If the pixel below 127 will be initialized to a value of 1 or white if the pixel above 127 will be initialized to 0 or black. After that remove the object below 30 pixels.

   Figure 4. Thresholding Image

4. After Thresholding, the next process is to resize image. This process resizes from the customized input image to the template size. In this research, the size of the template is 52x32 pixels. The original size of every image testing will be resized to 52x32 pixels.

5. The different shapes of the Javanese characters become the characteristic of each letter. These characteristics are determined from the binary form of each character.

0	0	0	0	0	0
0	0	1	1	1	1
0	0	1	1	1	1
0	1	1	1	1	1

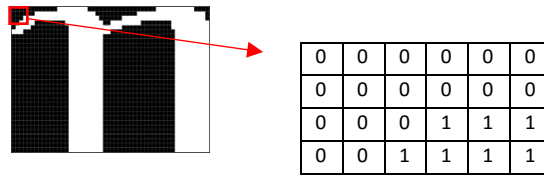| 0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	1	1	1
0	0	1	1	1	1
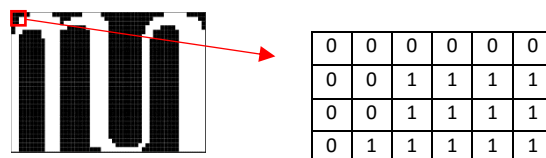
Figure 5. Part of Image Pixel

6. The next process is Classification, this process using Template Matching Correlation method. Template matching method is a statistical technique used to search for two or more matrices that are quantitative. The test image is compared with the training data/template image by finding the highest correlation value. The cross-correlation calculation of Figure 5 as follows:

Pixel 1 :

$$\bar{x}_i = \frac{1}{n} \sum_{k=1}^{n} x_{ik} = \frac{1}{1664} \cdot 0 = 0$$
$$\bar{x}_j = \frac{1}{n} \sum_{k=1}^{n} x_{jk} = \frac{1}{1664} \cdot 0 = 0$$

Pixel 2 :

$$\bar{x}_i = \frac{1}{n} \sum_{k=2}^{n} x_{ik} = \frac{1}{1664} \cdot 0 = 0$$
$$\bar{x}_j = \frac{1}{n} \sum_{k=2}^{n} x_{jk} = \frac{1}{1664} \cdot 0 = 0$$

Pixel 3 :

$$\bar{x}_i = \frac{1}{n} \sum_{k=3}^{n} x_{ik} = \frac{1}{1664} \cdot 0 = 0$$
$$\bar{x}_j = \frac{1}{n} \sum_{k=3}^{n} x_{jk} = \frac{1}{1664} \cdot 0 = 0$$

Until Pixel-n

Correlation :

$$r = \frac{\sum_{k=1}^{n}(x_{ik} - \bar{x}_i).(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{n}(x_{ik} - \bar{x}_i)^2 . \sum_{k=1}^{n}(x_{jk} - \bar{x}_j)^2}} = \frac{(0-0).(0-0)}{\sqrt{(0-0)^2(0-0)^2}} + \frac{(0-0).(0-0)}{\sqrt{(0-0)^2(0-0)^2}} + \frac{(0-0).(0-0)}{\sqrt{(0-0)^2(0-0)^2}} + n$$

The above calculation is a sample of some calculation of matching correlation template, because the image size used 52x32 pixel then summed as much as 1664 times.

Table 2. Testing Result

Testing / Template	l	k	l	r	r
a	-0,4112	-0,4674	-0,4112	-0,0295	-0,0295
n	-0,3379	0,1040	-0,3379	-0,1107	-0,1107
c	-0,3777	-0,3620	-0,3777	-0,2285	-0,2285
r	-0,1973	-0,0912	-0,1973	0,7747	0,7747
k	-0,4108	0,7958	0,4401	-0,0189	-0,0189
f	-0,3016	-0,4396	-0,3096	-0,1943	-0,1943
t	0,3699	-0,6451	0,3699	0,1045	0,1045
s	-0,3021	-0,1694	0-0,3021	-0,4479	-0,4479
w	-0,3890	-0,4039	-0,3890	-0,2718	-0,2718
l	0,7006	-0,4105	0,7006	-0,1916	-0,1916

p	0,3822	-0,3972	0,3822	-0,2437	-0,2437
d	0,4357	-0,5499	0,4357	-0,0924	-0,0924
j	-0,0583	-0,1574	-0,0583	0,0526	0,0526
y	-0,1333	-0,3325	-0,1333	0,0071	0,0071
v	-0,0367	0,1405	-0,0367	-0,0895	-0,0895
m	-0,3336	-0,2684	-0,3336	-0,3014	-0,3014
g	-0,1973	-0,3795	-0,1973	-0,0166	-0,0166
b	-0,5729	-0,0882	-0,5729	0,2711	0,2711
q	-0,3393	0,3399	-0,3393	0,2131	0,2131
z	-0,4712	0,4472	-0,4712	0,1512	0,1512

3. RESULT AND DISCUSSION

Table 3. Table of Data Testing

No	Data testing	Result	Character	Char Detected	True Char	False Char
1		HaTa	2	2	2	0
2		SaBa	2	2	2	0
3		HaNa	2	2	2	0
4		MaHa	2	2	2	0
5		HaSa	2	2	2	0
6		NaSaYa	3	3	3	0
7		WaRaTa	3	3	3	0
8		PaMaCaRa	3	3	3	0
9		NaGaSaRa	4	4	4	0
10		GaHaGaKa	4	4	4	0
11		HaKa	2	2	2	0
12		NyaTaKa	3	3	3	0
13		DhaKaHa	3	3	3	0
14		ThaPaTaHaSaWaya	6	7	5	1
15		WaSaHa	3	3	3	0
16		GaRaKa	3	3	3	0
17		KaTa	2	2	2	0
18		KaTaBya	3	3	2	1
19		ByaMaLaPa	4	4	3	1

20	ꦩꦩꦩꦴꦤꦩ	LaPaTgaWaRa	5	5	4	1
	Total		61	62	57	4

The table above shows the results of the character recognition of data testing with the template. The second column shows the data testing. The third column shows the recognition result using OCR Application. The Fourth column shows the number of character. The fifth column shows the character detected in the OCR Application. The sixth column shows the number of true character. The seventh column shows the number of false character.

From testing the data above, obtained the value of accuracy and error rate in character recognition Java Letter using template matching correlation with the calculation:

Accuracy $= \frac{57}{61} \; x \; 100 \; \% = 93.44 \; \%$

Error Rate $= \frac{4}{61} \; x \; 100 \; \% = \; 6.56 \; \%$

From the above calculation, got the accuracy value of 93.44% and error rate of 6.56% in pattern recognition Javanese Letter. The analysis, with an accuracy of 93.44%, Template Matching Correlation method good for pattern recognition.

4. CONCLUSION

1. In this research Optical Character Recognition can recognize the Javanese Letter pattern using Template Matching method with the accuracy 93.44 % and the error rate of 6.56 %. The result of this research about Optical Character Recognition using Template Matching method is successful for recognize.
2. Application Optical Character Recognition has been done and can translate the Javanese Letter.

REFERENCES

[1] Admin, "LIMC4U," Penjelasan Pasal 32 UUD 1945, 20 July 2015. [Online]. Available: http://limc4u.com/uud-1945/penjelasan-pasal/penjelasan-pasal-32-uud-1945/. [Diakses 10 November 2017].

[2] M. D. Elita, "Mutiara Dana Elita," Pengaruh Globalisasi Terhadap Kebdayaan Indonesia, 14 June 2015. [Online]. Available: http://danamutiaradana.blogspot.co.id/2015/06/pengaruh-globalisasi-terhadap.html. [Diakses 15 November 2017].

[3] M. C. Wibowo, I. D. G. R. Mardiana dan S. Wirakusuma, "PENGENALAN POLA TULISAN TANGAN AKSARA JAWA MENGGUNAKAN MULTI LAYER PERCEPTRON," dalam *Informatic Engineering conference*, Yogyakarta, 2015.

[4] Z. Mukhoyyar, "PENGENALAN KATA AKSARA JAWA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR," dalam *Informatic Engineering Conference*, Semarang, 2017.

Journal of Applied Intelligent System (e-ISSN : 2502-9401 | p-ISSN : 2503-0493)
Vol. 3 No. 2, November 2018, pp. 49 – 56

[5] A. Hidayat dan R. N. Shofa, "SELF ORGANIZING MAPS (SOM) SUATU METODE UNTUK PENGENALAN AKSARA JAW," dalam *Science and Technology Conference*, Bandung, 2016.

[6] A. Anam, S. T. Rasmana dan M. C. Wibowo, "PENGAMBILAN FITUR ANGKA JAWA MENGGUNAKAN SHADOW FEATURE EXTRACTION," dalam *Journal of Electrical and Electronics Engineering* , Surabaya, 2015.

[7] G. H. Wibowo, R. Sigit dan A. Barakbah, "Javanese Character Feature Extraction Based on Shape Energy," *International Journal of Engineering Technology,* vol. V, no. 1, pp. 154-170, 2017.