

FEATURES EXTRACTION ANALYSIS USING FREEMAN CHAIN CODE IN TO INTRODUCTION OF RUSSIA LETTER

Yosafat Dhima Harfana Putra, Guruh Fajar Shidik

Handwriting recognition is a handwriting recognition with the help of a software that has artificial intelligence to recognize the handwriting. Freeman Chain Code is a method of recognizing the outline based on the direction of the wind. This final project discusses the influence of the type font letters and also the rotation of letters when identified by the method of Freeman Chain Code. Examples of letters used in this study using cyrillic russia where each letter is tested with different font types and font rotation is also different to know if the effect is currently recognized. From the research, the rotation of letters and fonts also affect the level of success while recognizing the Russia letter which has a success rate of 70,88%.

Index Terms— Freeman Chain Code, Cyrillic, font

I. INTRODUCTION

The alphabet cyrillic first consisting of 43 letters (according to k .Mirchev) [1]. Based on the research j.e.tiernery and n.Revell in 1994 about *printed cyrillic character recognition system* if character of the alphabet Russia with 33 character (excluding digits and punctuation similar to English) [2]. Russia not popular in Indonesian compared with English, Chinese, Japan. Hence Russia still have not perceptible by Indonesians.

With the desire to know the accuracy of a method of to recognize character letters Russia so can also be a number of problems, then required instrument to recognize character letters Russia and one method that can be used to recognize character letters Russia is the method Freeman Chain Code.

Freeman chain code often used in processing image to represent a line, a curve or extent edge of an area. Freeman chain code often used as requiring a few memory. Research on application freeman chain code have are mostly done, including zingaretti (1998), who proposed the algorithms that can express a value bitmap who multi value used a code chain [3]. Koplowitz (1995) introduce a method of based on code chain to detect angular points of a curve [4]. A chain code is a lossless compression algorithm for monochrome images. The basic principle of chain codes codes is to separately encode each connected component, or “blot”, coordinates are transmitted. The encoder returns to the starting position, at which point the blot image [5].

Euclidian distance is used to find minimum distances. K-nn

(K-Nearest Neighbor) is used to classify the numerals [6].

Euclidean distance is distance the most commonly used in all application. Euclidean distance to count the difference the root of square between coordinates of a pair of objects and also used to for calculating image pixel of intensity on average [7].

II. METHODOLOGY

The one conducted to research is pre-processing, segmentation , features extraction , classifications, result.

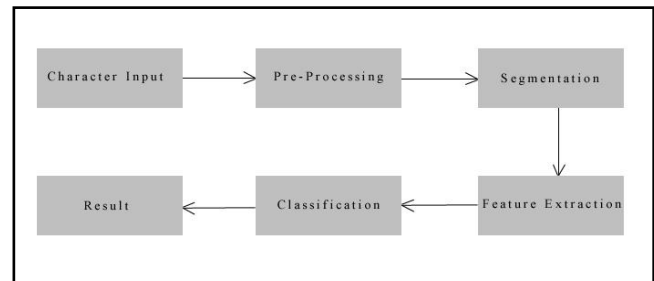


Fig. 1. The process experiment

A. Pre-Processing

This step is first step for begin research. Process on preprocessing is filtering, grayscaling, dan binary. To the process conducts emulation size pixels at all the data that measured as all. After size pixels same all then followed by filtering that serves to sharpen edge on image letter F. To find filtering can use the formula.

$$\frac{1}{1+a} = \begin{bmatrix} -a & a-1 & -a \\ a-1 & a+5 & a-1 \\ -a & a-1 & -a \end{bmatrix} \quad (1)$$

The next process is grayscaling. Grayscaling serves to change image rgb be the image gray. Grayscaling can formulated as below.

$$fo(x,y) = \frac{fi^R + fi^G + fi^B}{3} \quad (2)$$

The next step the binary process. Binary serves to change

the image gray be image binary having matrik 0 and 1 and had a black and white color. Binary can formulated as below.

$$g(x, y) = \begin{cases} 1 & \text{if } f(x, y) \geq T \\ 0 & \text{if } f(x, y) < T \end{cases} \quad (3)$$

After the pre-processing followed by the process segmentation.

B. Segmentation

The process segmentation is the process of separation setting of letters for obtaining edge of letters the detection the edge of using a technique. Process segmentation is edge detection, morphology, filling, bwperim, closing and followed by boundaries. The process first segmentation is edge detection. Edge detection is the search is a line the edge of an image. To find edge detection is change all image white to a black color while white near 1 pixels with black unchanged because will be a line its banks. Edge detection can formulated as follows.

$$H(x, y) = [-1 \ 1] \quad (4)$$

The result edge detection can be seen below.

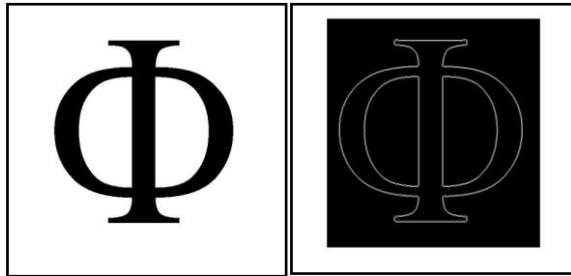


Fig. 2. Before and after edge detection

The next step is process morphology in order to clarify again line the edge that more clearly by using dilations, these were then followed by filling, bwperim, and closing. First done to the process morphology by doing dilations. Dilations is thickening on image binary. To research it does twice dilations, the first using strel periodicline. Morphology dilations by using strel periodicline that is added point pixels acts as an image a line that binary are much more thick and thickening his depends value that is inserted. The result morphology dilations by using strel periodicline can be seen below.

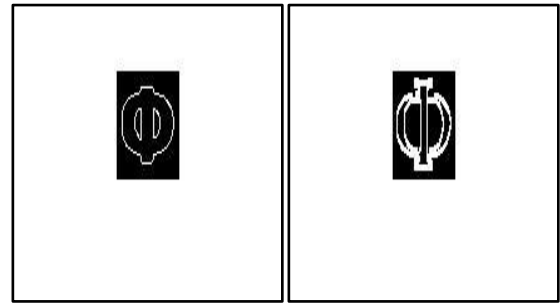


Fig. 3. Before and after strel periodicline

After the morphology dilations by using strel periodicline next perform the process dilations by using strel square. Strel square is to do thickening with binary around one of his image binary. The result morphology dilations by using strel square can be seen below.

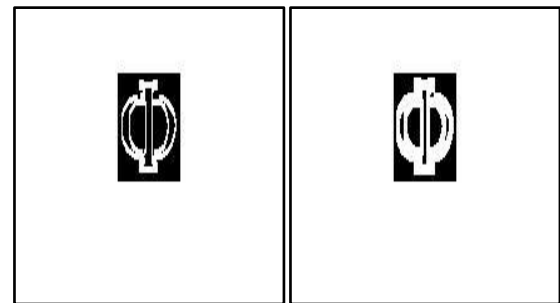


Fig. 4. Before and after strel square

The next step is filling. Filling it serves that hole is in image covered that when detect much better. The result filling can be seen below.

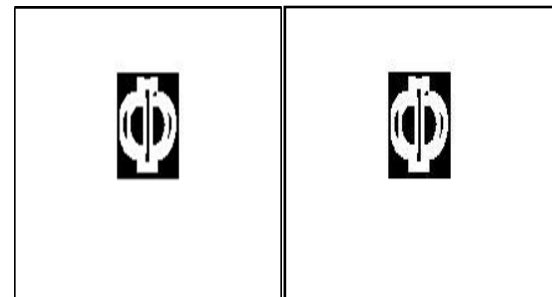


Fig. 5. Before and after filling

The next step is bwperim. Bwperim is working to obtain the perimeter from the process of filling. The result bwperim can be seen below.

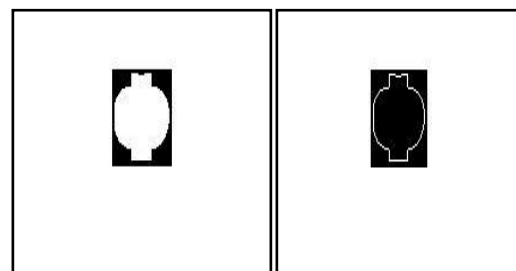


Fig. 6. Before and after bwperim

The next step is closing. Closing is used to connect a nearby object and removing the at an object. The result closing can be seen below.

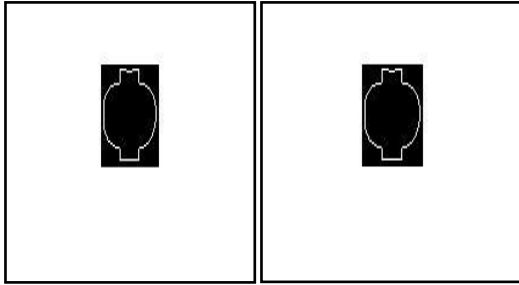


Fig. 7. Before and after closing

C. Feature Extraction

At this step will conduct extraction features to the letters Russia by using the method Freeman Chain Code to get a pattern on of all the letters to comparison to recognize patterns letters. Freeman chain code having 2 directions of the 8 direction and the 4 direction, in this research wearing 8 direction can be seen in the picture [8].

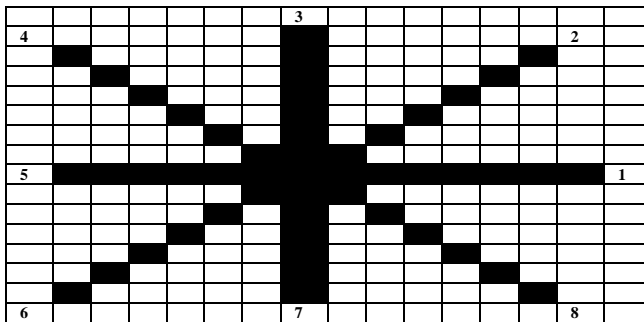


Fig. 8. 8 Direction Freeman Chain Code

D. Classification

At this step to compare data training with testing to know the level of accuracy. In the k – nearest neighbor rule, a test sample is assigned the class most frequently represented among the k nearest training samples [9].

Step to find k-nearest neighbour (k-nn) first determine the value of k in k-nearest neighbour (k-nn) , these were then followed by seeking d or distance. Distance used on a calculation k-nearest neighbor (k-nn) using euclidean distance. To the formula can be seen below [10].

$$d_{(x,y)} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (5)$$

After finding out of his euclidean distance then do ranked in the distance of euclidean k that had been already been determined to match the distance proximity euclidean with k already decided in the first place.

III. EXPERIMENT

A. Data Set

This paper used 1148 images from ms.word and internet, all image to resize 50x50 pixels, have 4 type font which consists of 3 data training, 1 data testing and every type font have 41 image letters then every image to rotate. The sample can be seen below.

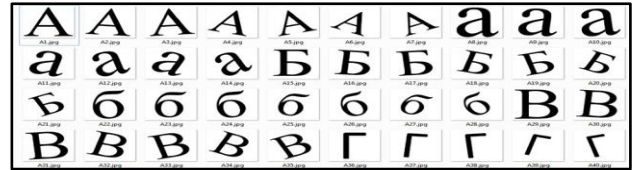


Fig. 9. Sample data Russia letters

B. Experiment Result

The results obtained research on letters russia with the methods freeman chain code by using classifications k-nearest neighbor (k-nn) can be seen in table below.

TABLE I
EXPERIMENT RESULT

Klasifikasi	Accuracy
<i>k-Nearest Neighbor (k-NN)</i>	70.88%
<i>Decision Tree</i>	2.44%
<i>Naïve Bayes</i>	4.04%

IV. CONCLUSION

The introduction of freeman chain code influenced by size font letters, rotation letters, the form of font letters. Normalization do not deprive of features freman chain code in letter because will affect the percentage accuracy when the introduction of letters. To increase the accuracy when the introduction of letters more again the amount of data training. The result of classification k-nearest neighbor (k-nn) better in compare with the classification decision tree and naïve bayes for the introduction of letters Russia is 70,88%. This research is only used in letter russia (cyrillic). Research it aims to get beneficial for research a typeface other or the same letters the method used to better.

REFERENCES

- [1] I. J. of R. Studies, "Short History of The Cyrillic Alphabet," vol. 2, no. 2, 2013.
- [2] N. R. J.E. Tierney, "Printed Cyrillic Character Recognition System," *Print. Cyrillic Character Recognit. Syst.*, pp. 3856–3861, 1994.
- [3] P.Zingaretti,M.Gasparroni,L.Vecchi, 1998, *Fast Chain Coding Of Region Boundaries, IEEE Trans. Pattern Anal. Mach. Intell.* 20 (4) page 407-414.
- [4] J.Koplowitz, S.Plante, 1995, *Corner Detection Of Chain code Curves, Pattern Recognition* 28 (6) (1995) 843-852.

- [5] S. M. M. G. G. Rajput, "Isolated Handwritten Marathi Numerals Recognition Based upon Fourier Descriptors and Freeman Chain Code," vol. 6, pp. 289–298, 2010.
- [6] U. Ravi Babu, A. Kumar Chintha, and Y. Venkateswarlu, "Handwritten Digit Recognition Using Structural, Statistical Features and K-nearest Neighbor Classifier," *Int. J. Inf. Eng. Electron. Bus.*, vol. 6, no. 1, pp. 62–68, Feb. 2014.
- [7] D. S. S. K.M.Ponnmoli, "Analysis of Face Recognition using Manhattan Distance Algorithm with Image Segmentation," vol. 3, no. 7, pp. 18–27, 2014.
- [8] P. Annapurna, S. Kothuri, and S. Lukka, "Digit Recognition Using Freeman Chain Code," vol. 2, no. 8, pp. 362–365, 2013.
- [9] Aman Kataria1 , M. D. Singh2, "A Review of Data Classification Using K-Nearest Neighbour Algorithm," *IJETAE*. 2013.
- [10] D. S. S. K.M.Ponnmoli, "Analysis of Face Recognition using Manhattan Distance Algorithm with Image Segmentation," vol. 3, no. 7, pp. 18–27, 2014.