International Journal of Advanced Engineering and Management Research
Vol. 2 Issue 6, 2017

http://ijaemr.com/ ISSN: 2456-3676

CONVERSION OF JAVA CHARACTER TO ALPHABETS ON DIGITAL IMAGE USING DIAGONAL BASED FEATURE EXTRACTION

Cahya Rahmad
Information Technology Department
State Polytechnics of Malang
East Java, Indonesia

Ersa Rahmawati Aurollah
Information Technology Department
State Polytechnic of Malang
East Java, Indonesia

Putra Prima Arhandi
Information Technology Department
State Polytechnics of Malang
East Java, Indonesia

Rudy Ariyanto
Information Technology Department
State Polytechnics of Malang
East Java, Indonesia

ABSTRACT

The Java character is one form of writing of Javanese culture in addition to using the alphabet that we often encounter in some relics that are in the special region of Yogyakarta (DIY). Characteristics of Java's character is very complex, so it is a bit difficult to understand due to it's not familiar.

Technological developments are demanding to be able to build an application that can help the domestic or international travelers to identify a Java character in which there is historical information from several existing relics at DIY. The main objective of this research is to handwriting recognition in which the application can convert Java's character to alphabets. Identification methods in this research use of digital image processing, namely the extraction of characteristics by method diagonal based feature extraction. The method can recognize the characteristics of each Java character, Especially the basic character in the Java character which consists of 20 types of letters.

Based on the results of the study known that by using the method of extracting the characteristics of image taking positions and the form of writing is to determine the process of calculating the extraction of these characteristics. Each character in the main, character uses 3 different sets of character, so the total is 60 data sets used. The application's accuracy rate reaches 75% in order to properly convert Java character to the alphabet. From 20 data testing, get 15 correct data and 5 wrong data.

Key Words: hand writing recognition; conversion of Java character to alphabets; diagonal based feature extraction.

Introduction

Indonesia is a country rich in culture and language. The wealth that we must preserve and keep when many imported cultures enter Indonesia. Daerah Istimewa Yogyakarta (DIY) is one of the provinces in Indonesia, which became the center of tourism for domestic and foreign. This is bec ause DIY has many tourist sites that are easy to reach. In addition, Javanese culture is still thick, creativity, art and friendly attitude DIY community.

The Java character is one of the forms of writing Javanese culture besides using alphabets which we often encounter in some relics in DIY. The forms of Java character are very complex, it is a little difficult to can because of their unfamiliar shape, so this makes the domestic or foreign tourists need a guide to interpret Java character in which there is historical information from some of the existing relics in DIY. The tool to translate Java character into alphabets that already exist only software that input in manual type. In fact, Javanese character is often found in places such as monuments, etc. Therefore, it takes a system that can recognize of a handwriting found in some tourist attractions in Yogyakarta.

Handwriting Recognition is an emerging as well as a challenging area in the fields of pattern recognition and computer vision. The aim of a handwriting recognition system is to convert human readable characters, which are present in a photographed or digitized sheet of paper and convert it into a machine editable form[1]. Diagonal Based Feature Extraction is one of the feature extraction that is often used to recognize a handwriting. Diagonal Based Feature Extraction method can extract and store special characteristics of an object in the image. These features will then be used as a comparison to recognize objects in digital imagery[1].

From the description of the above problems, then in this study developed a system to convert Java characters to alphabets on digital images. This system can detect handwriting in the form of writing Java characters on a digital image and able to convert to Latin letters. It is expected that this system can be a medium who can facilitate DIY tourists to read and know the historical information in some tourist attractions.

LITERATURE REVIEW

The Proposed Recognition System

In this section, the proposed recognition system is described. A typical handwriting recognition system consists of pre-processing, feature extraction, and distance measurement. The schematic diagram of the proposed recognition system is shown in Fig.1

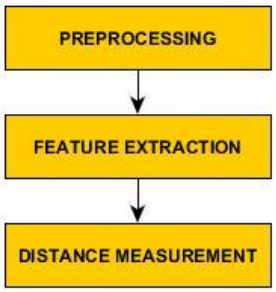


Figure 1. Schematic diagram of the proposed recognition system.

The preprocessing is a series of operations performed on the scanned input image. It essentially enhances the image rendering it suitable for feature extraction. The various tasks performed in the image in preprocessing stage are shown in Fig.2 [2].

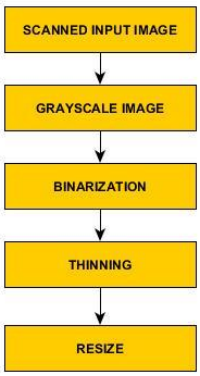


Figure 2. Pre-processing of Handwriting Recognition.

Scanned Input Image

In scanned input image, the recognition system acquires a scanned image as an input image. The image should have a specific format such as JPEG, BMT etc. This image is acquired through a scanner, digital camera or any other suitable digital input device.

Grayscale Image

The handwriting images are being normalized for handwriting recognition. These are transformed into gray scale images. The conversion from color image to grayscale image is given by the following equation:

(1)

Where Rn,Gn, Bn denote the red, green, and blue color constituents of the nthpixel of the color image and Gn is the gray level value of nthpixel of the gray scale image, and the resolution of the image is M □□N [3].

Binarization

Binarization process converts a gray scaleimage into a binary image using global threshold technique.

Thinning

Zhang and Suen's thinning algorithm [4]: It performs thinning of binary images by repeating two sub iterations: one deletes the southeast boundary points and the northwest corner points while the other one deletes the northwest boundary points and the southeast corner points. The two sub-iterations are repeated until no more points validate the deleting rules. Point deleting is done according to a specific set of rules.

1) $2 \leq N_{(x}$ ***(2)***

(3)

(4)

(5)

Resize

Change the size of the digital image in pixels. In this research must be resized 30 x 30 pixels.

Java Characters

The Javanese language is one of the regional languages that are part of the national culture of Indonesia, which lives and remains used in the community concerned. Carakan (Javanese alphabet) used in Javanese spelling consists of over 20 basic characters that are syllabic shown in Fig.3. Each main character has a pair, a character that serves to connect a consonant closed syllable with the next syllable. The Javanese language is always constantly experiencing progress, so the spelling also needs to be adjusted with these developments, especially in writing the character of Java is increasingly unknown to the public. Therefore, the Office of Education and Culture of the Special Province of Yogyakarta (DIY) in the budget year 1992/1993 decided to set the guidelines for writing Javanese script. It has also been agreed by the Provincial Government of Central Java and the Provincial Government of East Java [5].

Handwriting Recognition

Handwriting Recognition is an emerging as well as a challenging area in the fields of pattern recognition and computer vision. The aim of a handwriting recognition system is to convert

human readable characters, which are present in a photographed or digitized sheet of paper and convert it into a machine editable form [1].

The handwriting recognition process faces certain challenges [6]-[9]. The biggest one is that the handwritten images‟ various dimensions have to be normalized and/or processed in order for them to fall within the system's boundary specific requirements. Thus, the challenges of handwritten digit recognition do not just result from the many different ways in which a single digit can be written, but also arise from the various requirements that are imposed by the applications that are specifically used [10], [11]. In addition, there are the varying degrees of thickness of people's handwriting and the writing's different positions when it comes to the sample's margins [12]-[14]. Moreover, Akhtar et al. [10] indicated that as people's various writing styles will depend on their age, qualifications, mode, background, etc. handwritten digit recognition is a relatively complex research task, and the format of handwritten digits will have a big impact on classification and/or identification, as the different subjects will all use different styles of writing. In fact, even digits that have been written by the same person at different times were also found to vary [15], [16]. Therefore, it is almost impossible to develop a generic recognizer that can recognize an infinite number of writers‟ handwritten digits.

PURPOSED METHODOLOGY

Diagonal Based Feature Extraction

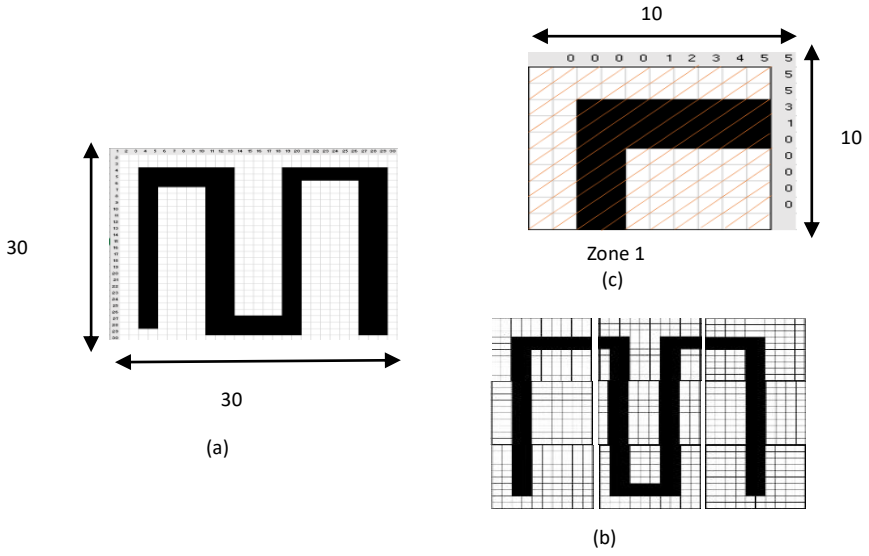


Figure 4. Procedure for extracting feature from the characters.

In this stage, the features of the characters that are crucial for

classifying them at recognition stage are extracted. This is an important stage as its effective functioning improves the recognition rate and reduces the miss distance measurement [17]. Diagonal feature extraction scheme for recognizing off-line handwritten characters is proposed in this work. Every character image of size 30x30 pixels is divided into 9 equal zones, each of size 10x10 pixels (Fig.4(c)). The features are extracted from each zone pixel by moving along the diagonals of its respective 10X10 pixels. Each zone has19 diagonal lines and the foreground pixels present long each diagonal line is summed to get a single sub-feature, thus 19 sub-features are obtained from each zone. These 19 sub-features values are averaged to form a single feature value and placed in the corresponding zone (Fig.4 (b)). This procedure is sequentially repeated for the all the zones. There could be some zones whose diagonals are empty of foreground pixels. The feature values corresponding to these zones are zero. Finally, 9 features are extracted for each character[18].

Data Set

In this research, the dataset contains 60 handwriting of Java characters use 3 different characters. The database is created to different handwriting. In this database, all images captured with a camera phone in July 2017. Figure 5 shows a sample image of the dataset.

Distance Measurement

Once the features are extracted, the next step is to measure the distance between images. The distance between two images is a major concern in image recognition and computer vision. The final step of handwriting recognition is measuring the distance between two images. Image similarity is the distance between the vectors of two images. The distance between feature space representations is used as the basis for recognition decisions. One way or another, distance measurement has a big impact in handwriting recognition area. Distance measurement methods are used in many areas like finance, data mining, voice recognition, handwriting recognition and signal decoding.

Euclidean distance is used for distance measurement between images. Euclidean Distance is defined as the straight-line distance between two points, which examines the root of square differences between the coordinates of a pair of objects [19],[20]. Euclidean Distance can be calculated using the equation below:

$$Euclidean\ Distance = \sqrt{\sum_{n=1}^{i}(Xn - \square\square\square}$$

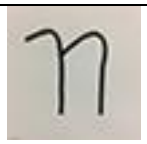
IMPLEMENTATION

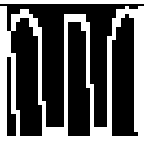
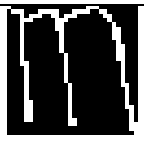
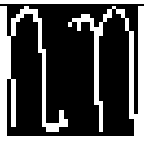
Testing Accuracy

Testing accuracy is used to test the purpose of the conversion of Java characters to alphabet are able to convert Java characters to the alphabet by applying the method diagonal based feature extraction.

Tabel 1.Testing Accuracy

No.	Input Image	Output Image	Ma-nual	Recognition	Result
1.			HA	YA	INCORERCT
2.			NA	NA	CORRECT
3.			CA	CA	CORRECT

No.	Input Image	Output Image	Ma-nual	Recognition	Result
4.			RA	RA	CORRECT
5.			KA	YA	INCORERCT
6.			DA	DA	CORRECT
7.			TA	TA	CORRECT
8.			SA	SA	CORRECT
9.			WA	CA	INCORERCT
10.			LA	LA	CORRECT
11.			PA	PA	INCORERCT
12.			DHA	DHA	CORRECT

No.	Input Image	Output Image	Ma-nual	Recognition	Result
13.			JA	JA	CORRECT
14.			YA	YA	CORRECT
15.			NYA	NYA	CORRECT
16.			MA	MA	CORRECT
17.			GA	GA	CORRECT
18.			BA	BA	CORRECT
19.			THA	THA	CORRECT
20.			NGA	NGA	INCORERCT

Precision can be seen as a measure of exactness or fidelity. High precision means a relevant result. Precision is defined

As [21]:

$$Precision = \frac{True\ Po}{False\ Positive+} \qquad (7)$$

$$Precision = \frac{15}{20} = 0{,}75\ x\ 100\ \% = 75\%$$

CONCLUSION and Future work

Based on the research that has been done, there are several conclusions:

1. Development of Java characters conversion application to alphabet on digital image using diagonal based feature extraction method successfully done.
2. Similarity test using Euclidean distance calculation can give similarity rate between image testing and training image.
3. In the process of testing the application can determine the conversion of Java characters to alphabet with an accuracy of 75%.

In future works we recommended:

1. It is expected that further research will be added with a variant of Java character letter using paired characters, etc.
2. On the application development is expected to be able to apply in the android version. So it can have more value for the benefit of society in general.

REFERENCES

[1] Meenu Alex and Smija Das, "An aprooach towards Malayalam Handwriting Recogniting Using Dissimilar Classifiers," Global Colloquium in Recent Advancement and Effectual Researches in Engineering, Science and Technology (RAREST), 2016.

[2] Rafael C. Gonzalez, Richard E. woods and Steven L.Eddins, *Digital Image Processingusing MATLAB*, Pearson Education, Dorling Kindersley, South Asia, 2004.

[3] Md. Al-Amin Bhuiyan. "Towards Face Recognition Using Eigenface". International Journal of Advanced Computer Science and Applications (IJACSA). Vol. 7, No. 5, 2016.

[4] T. Zhang, C. Suen, A fast parallel algorithm for thinning digital patterns, Commun. ACM 27 (3) (1984) 236–239.

[5] Darusuprapta, Prof. DR.,et al. 2002. *Pedoman Penulisan Aksara Jawa*. Yogyakarta: Yayasan Pustaka Nusantara.

[6] Jackson, J.E. (1991). A User‟s Guide to Principal Components (Wiley).

[7] Jollife, Ian. "Principal component analysis." New York: SpringerVerlag., 2002.

[8] Mohebi, Ehsan, and Adil Bagirov. "A convolutional recursive modified Self Organizing Map for handwritten digits recognition." Neural Networks 60 (2014): 104-118.

[9] Khatri, Sunil Kumar, Shivali Dutta, and Prashant Johri. "Recognizing images of handwritten digits using learning vector quantization artificial neural network." Reliability, Infocom Technologie

[10] Akhtar, Muhammad Suhail, and Hammad A. Qureshi, "Handwritten digit recognition through wavelet decomposition and wavelet packet decomposition." Digital Information Management (ICDIM), 2013 Eighth International Conference on. IEEE, 2013.

[11] Niu, Xiao-Xiao, and Ching Y. Suen. "A novel hybrid CNN–SVM classifier for recognizing handwritten digits." Pattern Recognition 45.4 (2012): 1318-1325.

[12] Wu, J. (2012). "Advances in K-means clustering: a data mining thinking." Springer Science & Business Media.

[13] Khan, M. E. (2015). "K-means clustering."

[14] Juanying Xie, Shuai Jiang 2010, "A simple and fast algorithm for global K-means clustering", 2010 Second International Workshop on Education Technology and Computer Science,

[15] Decong Yu, Lihong, Ma. "Digit Recognition Based on Multi-features", IEEE Systems and Design Engineering Symposium, 2007.

[16] Ye Xu, Dezhan Liu: Clustering research of handwritten digit recognition by genetic algorithm. BIC-TA 2010: 291-297.

[17] S.V. Rajashekararadhya, and P.VanajaRanjan, "Efficient zone based feature extractionalgorithm for handwritten numeral recognition of four popular south-Indian scripts," *Journal of Theoretical and Applied Information Technology*, JATIT vol.4, no.12, mpp.1171-1181, 2008.

[18] J. Pradeep, E. Srinivasan and S.Himavathi, "Diagonal Based Feature Extraction for Handwritten Alphabets Recognition System Using Neural Network", International Journal of Computer Science & Information Technology (IJCSIT). Vol. 3, no. 1, Feb. 2011.

[19] Gawande, M.P. and Agrawal, D.G., Face recognition using PCA and different distance classifiers. **IOSR Journal of Electronics and Communication Engineering**, 9(1), pp.1-5.

[20] Cha SH, Comprehensive Survey on Distance/Similarity Measures Between Probability Density Functions, City, 2007, 1:2, 1.

[21] Dr.P.Subashini, S.Jansi . " Optimal Thinning Algorithm for detection of FCD in MRI Images", International Journal of Scientific and Engineering Research. Vol. 2, Issue 9, September 2011.

AUTHOR PROFILE

Cahya Rahmad: He received BS degrees from Brawijaya University Indonesia in 1998 and MS degrees from Informatics engineering at Tenth of November Institute of Technology Surabaya Indonesia in 2005.He received Doctoral degrees at Saga University Japan in 2013, He is a lecturer in State Polytechnic of Malang Since 2005. His interest researches are image processing, data mining and patterns recognition.

Ersa Rahmawati Aurollah: She received BS degrees from State Polytechnics of Malang 2017, Her interest researches are image processing and handwriting recognition.

Putra Prima Arhandi: He received BS degrees from Brawijaya Universityin 2010 and MS degrees from Informatics engineering at Tenth of November Institute of Technology Surabaya in 2013. He is a lecturer in State Polytechnic of Malang since 2014. His interest researches are image processing, mobile and web application.

Rudy Ariyanto, He received BS degrees from Brawijaya University Indonesia in 1997 and MS degrees from Gadjah Mada University, Yogyakarta Indonesia in 2014. He is a lecturer in State Polytechnic of Malang Since 1999. His interest researches are Machine learning, Decision Support System and pattern recognition.