Comparison Online to Offline Handwritten Jawi Character Recognition Application

Nazaruddin¹ and Sayed Muchallil²*

¹Department of Informatics, Faculty of Mathematic and Natural Science, Syiah Kuala University, Kota Banda Aceh, Aceh – 23111, Indonesia; anzaro@unsyiah.ac.id
²Department of Electrical and Computer Engineering, Faculty of Engineering, Syiah Kuala University, Kota Banda Aceh, Aceh – 23111, Indonesia; sayed.muchallil@unsyiah.ac.id

Abstract

Objectives: This paper investigates processing time comparison of handwritten Jawi characters recognition application between offline and online version. Methods/Analysis: The use of web server for processing offline tasks is considerable. This process is covered by translating offline application code into web-based programming language such as PHP. Total of 100 types of handwritten characters were used for this experiment. These characters were transformed using scaling transformation into four categories. Findings: The result shown online application has better performance than offline. The online application succeeds to extract moment feature of handwritten characters. Novelty/Improvement: The comparison was performed for choosing better Optical Character recognition (OCR) application system: online or offline method.

Keywords: Handwritten Jawi Characters, OCR Performance Comparison, Online OCR, Moment Feature

1. Introduction

Historical documents can be preserved much easier in digital format compared to their original form, paper-based. However, this digitalization needs some image processing phases, such as denoising, binarization, features extraction and pattern classifier. Character recognition can be categorized in features extraction stage.

Character recognition aims to determine digital text or alphabet that can be read by computer from a digital image. Beside character recognition, ancient document preservation also needs document analysis to keep their information. Character recognition plays an important role in document analysis process. In the last decade, character recognition has become an interesting research area. At first, English and Roman characters were the only objects to be investigated in this research field. Recently, some research has been done to recognize Arabic, Korean and Chinese characters. Character recognition specifically including in pattern recognition area and is one part of Optical Character Recognition (OCR).

Converted a digital image into a readable digital text to identify a character is known as OCR. A complete OCR system has some stages. First is pre-processing step. This phase will remove some noise from an image and is also known as denoising. Next, the output of this stage will use for binarization process. This step converts an image matrix into a matrix that only has 0 and 1 value. Then, the output of the latter method can be used for character segmentation, if needed, and followed by features extraction from the separated characters. The last step is pattern classifier that will recognize each character. OCR also consists of artificial intelligent and computer vision.

In addition to document analysis necessity, historical documents content need to be distributed. Internet is the best dissemination process. For this reason, more offline application has been ported to online version. Web server can handle computation process. Thus, this con-
version can be made possible. Developing online system, the system prototype has to be constructed. Next, the online prototype of the system is compared to the existing system or application. This comparison aims to find out whether the online system is working or not. If the results show that the online system does not work correctly, the development process should be started again.

This paper compares online and offline Jawi handwritten character recognition performance based on the processing time for extracting feature from a character image. This comparison was carried on to choose which type of the applications that is more reliable to be applied in an OCR system for Jawi handwritten image.

Jawi characters, Arabic characters with some additional features, are the characters that is used to write Malay language. Malay language has been used in South East Asia since 600 years ago. There are many countries that communicate using Malay language such as Indonesia, Malaysia, Singapore and Brunei. There are many heritages in South East Asia that were written by using Jawi characters. For examples: historical manuscripts, ancient tombstone, and building epigraph. The ancient documents contain full of useful information such as history, religious matter and natural medication.

There are six additional Jawi characters that do not exist in Arabic characters. They are Nya, Ca, Nga, Pa, Ga, and Va. In addition, Jawi characters are not familiarly written with harakah or reading mark.

Arabic character has widely adopted in many countries such as Pakistan with Urdu or Iran with Farsi. Jawi character is also the result of Arabic characters implementation in most South East Asian countries to write Malay Language. Jawi Character research was published in 20005 and followed by 6 in 2002. After this year, more research has been carried on in this area.

Binarization is a process to convert a grayscale image into a binary image. Binary image is a type of image that consists of two level representations. The two level representation are 0 and 1. The 0 value represents black colors or known as background and the 1 value represents white color that known as foreground.

Binarization of an image should use a threshold to determine which part of an image become background and which part is a foreground. There are numerous binarization algorithm such as Otsu thresholding, Sauvola algorithm, Su and Lu algorithm and modification of the existing algorithm such as Nina and NICK. In this paper, the thresholding was chosen manually to get the best result.

Feature extraction is a process to get features as attributes from an image. To identify an object in an image, the features must be extracted from the image. Feature extraction aims to get significant features on an image depending on their intrinsic characteristics.

Hu moment invariant is one of extraction feature methods that is commonly used to get image features. Hu extracted the feature from an image using moment method. Hu moment invariant also known as geometrical moment invariant Hu moment was defined as:

\[
\Phi_1 = \eta_{20} + \eta_{02}
\]

\[
\Phi_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2
\]

\[
\Phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2
\]

\[
\Phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} - \eta_{03})^2
\]

\[
\Phi_5 = (\eta_{30} - \eta_{12})(\eta_{30} + \eta_{12})[3(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2]
\]

\[
\Phi_6 = (\eta_{30} - \eta_{12})(\eta_{30} + \eta_{12})[3(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2]
\]

The rest of this paper including literature review in next section, the experiment method describes in Experimental Detail section, result and discussion is available in fourth section and finally the conclusion presents in fifth section.

2. Materials and Methods

In this experiment, private handwritten database is used. The database contains of Jawi characters that wrote by four writers. The writers had different educational background. For the experiment, we chose ten characters of every writer data which different character randomly. The samples which are taken was considered to represent every shape that are different from each character. This is to ensure that every different shape of each character drawn as samples in this experiment. Figure 1 is examples of the experiment’s character. Figure 2 is example of handwritten Jawi character after binarization process.

The offline application for extracted Hu moments was developed using MATLAB. The MATLAB script which is developed as the applications was taken from RC.
Gonzales's book[15]. The offline application was containing two functions while the online application only consists one functions. In online application, binarization and feature extraction process combined in a function, while in offline application binarization and feature extraction process was pack in different package.

Figure 1. Example of Jawi characters.

Gonzales's book[15]. The offline application was containing two functions while the online application only consists one functions. In online application, binarization and feature extraction process combined in a function, while in offline application binarization and feature extraction process was pack in different package.

Figure 2. Example of Jawi characters after binarization performance.

Comparison of time processing in the application is performed on the same notebook with specification: Intel i5 with speed is ~ 2.4 GHz. Time calculation process start by providing the timer function at the beginning and at the end of the moment invariant calculation script. The similar was applying to binarization process of the gray-scale images.

Finally, the data's time is comparing between online and offline to figure out which application is fastest in binarization process and extracted Hu moment feature from an image. The comparing data would be references of applying OCR systems.

3. Result and Discussion

In this section, we present the experimental result of time comparison between offline and online feature extraction in handwritten Jawi character recognition. The result of the comparison was provided in average value of the result.

3.1 Offline Application

The result of the first dataset of handwritten Jawi character was shown in Figure 3. The offline application was proceed ten characters. The biggest time execution process is at the first process, it need 57 miliseconds. Next character process, it just need 4 miliseconds in average, except the second process that used 7 miliseconds. Overall, the average time of the application process to proceed the first dataset is 4 miliseconds.

Figure 3. The execution time of offline system on dataset 1.

The result of the second dataset of handwritten Jawi character was shown in Figure 4. The biggest time execution process is at the first process, it need 60 miliseconds more 3 miliseconds than the first dataset. Next character process, it just need 7 miliseconds in average, except the second process that used 11 miliseconds. Overall, the average time of the application process to proceed the first dataset is 7 miliseconds. The second dataset had longer execution process than the first.

Figure 4. The execution time of offline system on dataset 2.

The result of the third dataset of handwritten Jawi character was shown in Figure 5. As same as previous process, the offline application was proceed ten characters. The first character process required 63 miliseconds. The second character process had less expected time, it was 15 miliseconds. Furthermore, the next process just expected 12 miliseconds in average. Overall, the average time of the application process to proceed the first dataset is 12 miliseconds.

The fourth dataset result was shown in Figure 6. In this process, the execution times was various than other three...
datasets. The first character had 70 milliseconds, while the second characters proceed for 29 milliseconds. Next characters had different processing time, it vary between 24 and 26 milliseconds. The fourth dataset is the longest required time execution than others dataset.

Figure 5. The execution time of offline system on dataset 3.

Figure 6. The execution time of offline system on dataset 4.

3.2 Online Application

The online application was developed using PHP programming and proceed by using Apache server. This online application was placed on localhost server. Figure 7 showed the result of handwritten Jawi feature extraction using the online application. The online application was tested on ten characters. The first character need 0.08 microsecond for execution process while the second characters required 0.06 microseconds. The longest time for execution process is at the fifth character with 0.9 microsecond execution process. The average period of time that required for execution time is 0.6 microsecond exclude the fifth character.

Figure 7. The execution time of online system on dataset 1.

Figure 8. The execution time of online system on dataset 2.

The result of the third dataset of handwritten Jawi character was shown in Figure 9. As same as previous process, the offline application was proceed ten characters. The first character process required 0.31 microseconds. The second character process had less expected time, it was 0.24 microseconds. There are two characters that expected longer process the the average, the fourth character with execution time 0.72 microseconds and the eight characters with execution time 0.63 microseconds. Overall, the average time of the application process to proceed the first dataset is 0.27 microseconds.

The fourth dataset result was shown in Figure 10. In this process, the execution times had the same pattern as the third dataset. The fourth dataset had two highest execution time: the fourth and the eigth character. Overall, the average time that required for the fourth dataset feature extraction is 0.3 microseconds.

The result indicate that the online application had faster processing time than the offline application. The first dataset extracted the feature by using offline application and online application by ratio 1:10000. The average result of the first dataset.
4. Conclusion

Performance comparison of offline and online handwritten Jawi OCR system was presented. The OCR system consists of two stages: binarization process and feature extraction process. To extract the feature from the character, Hu moment invariant was used. Binarization and feature extraction process were combined in an application in online handwritten OCR system, but both of them were separated in offline OCR system.

The Online handwritten Jawi OCR system has faster execution time than the offline OCR system. This result could be consideration to develop OCR system for online application processing. Furthermore, the online OCR system can use for sharing and did not need another installation process.

5. References