Myanmar Character Extraction from Vehicle Images Using Aspect Ratio and Bounding Box

Ohnmar Khin*, Montri Phothisonothai*, and Somsak Choomchuay**
* International College, ** Faculty of Engineering,
King Mongkut's Institute of Technology Ladkrabang (KMITL), Bangkok, Thailand.
Corresponding e-mail: ohnmarkhin.cs@gmail.com

***Abstract*— In this paper, we publicize Myanmar character extraction system using the license plate number as experimental sample. In recent times, the identification of car license becomes a popular task because of the increase in the number of vehicles. To settle this problem, there are a number of techniques in which aspect ratio and bounding box are the suitable technology. We also use edge boxes in order to separate the background from the foreground and accentuate in the foreground. In the investigation, the Myanmar character are segmented firstly and then these are extracted precisely. The experimental results showed that the characters have been correctly extracted by the accuracy rate of 90%.**

Keywords— aspect ratio; bounding box; number plate extraction;

I. INTRODUCTION

A license plate is a metal plate that suppresses numbers, and letters fixed on the front and back view of car and is used to identify the vehicles. This research aims to extract the numbers and characters from Myanmar license plates. The license plate extraction is very important in many applications such as traffic light, car parking and travel and airport parking. The research of vehicle number plate recognition is in advance and the outcome of this research has been practically used by traffic-light development. Many researcher of pattern recognition field continue study of character extraction for a long time. Their research is caused by effort that shifts human visual system to machine. But this performance of the machine is very small in variance to human scholarly.

The character of Myanmar originated in Indian Sanskrit. Myanmar character consists of letters and numbers that are not Arabic numerals. The numbers and characters are make-up of many curves. In other words, most characters and numbers are indistinguishable in shape to one another. Myanmar letters and numbers are shown in Figure 1. In this paper, we would like to extract Myanmar character using aspect ratio and bounding box as learning method. The size of the vehicle image is 2048×1536. In the preprocessing procedure, we extract number plate from the car image and characters from the license plate. The result is shown in Figure 2.

Myanmar letters and numbers that are made of curves and straight lines are shown in Figure 1. In first and second rows are Myanmar numbers, 0-9 and there are 18 letters in the third to sixth rows.

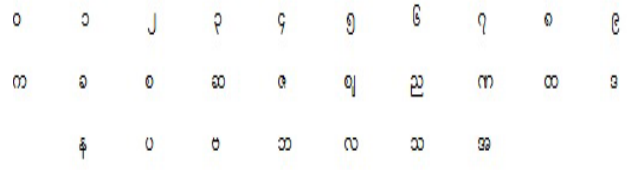


Fig.1. Typical Myanmar Letters and Numbers

I. RELATED WORK

Numerous authors have expressed the many research for the segmentation and extraction of the characters from the image. [1,2] O. Khin, M. Phothisonothai and S. Choomchuay researched the non-fixed LPR (License Plate Recognition) for Myanmar license plate detection system. This research does detect the different angles of dissimilar vehicles with a top correctness. [3] The paper submitted Myanmar printed character using OCR. The characters can be detected by Hopfield neural network algorithm about 98%. [4] The Myanmar keyboard layouts difference are described. Myanmar language use the many keyboards. Zawgyi Myanmar Unicode is popular for keyboard and typing. The current system is not familiar with the users. This work focusses on the keyboard inspection mapping. The result displays that present Myanmar PC keyboards are constant. [5] The authors suggested the comparison method for printed characters and Myanmar handwriting recognition. The discrete Hidden Markov model is researched by the paper. The words, printed characters and handwriting are selected in this experiment. The Hidden Marko Models classifier is proposed for the recognition process. This method accepts the handwriting in print style. The results convey 97% the printed characters and 91.1% of the handwriting. [6] They expressed character recognition of handwriting. The method identifies the handwriting documents writer. Fast Fourier Transform method extracts one character in this method. The median filter is experimented on 40 writers. It provides the identification correct rate. This filter also removes noises in this individual character. [7] In the existing system, Myanmar words sorting is compared with the complication of expanded words sorting. Myanmar ordinary scripts and Myanmar Word

978-1-5386-2615-3/18/$31.00 ©2018 IEEE

Tokenizer are changed into analyzable enlarged scripts based on LIPIDIPIKAR treatise. [8] Myanmar Printed Text are identified and converted in this paper. This paper intends to develop an OCR for the printed text. The printed character identification is the old research. In Myanmar, the printed characters are used over 85%. Myanmar character recognition algorithm is extended from the Hopfield Neural Network method. The result displays Myanmar printed characters about 97%. [9] The paper proposed a method for searching out of vocabulary words from Myanmar text. This method includes two processes. In the first step, the maximal substrings are extracted and post-processing of having three parts are done in the second step. Lastly, OOV words are searched. The research showed that the accuracy does seem favorable. [10] The Positional Gesture Text Input for syllabic scripts are demonstrated. For PC keyboards, text input is not easy. This paper describes text input identification. Many mobile computing devices and writing natures are applicable. [11] The Printed Document Image converts into Machine Acceptable Text Format that are offered. There are multiple document images needed to change into the machine acceptable text format. The input image is expanded and then a novel segmentation method is used for the characters segmentation in the first step. To end, the hierarchical mechanism is utilized for the character recognition image. Many documents are handled for experiment and the results exhibit the proposed method situation.

II. System Architecture

The overall system of character extraction is illustrated in Figure 2. The proposed approach is utilized for characters segmentation and extraction from the license plates that are varying from the size, style, and format. The vehicle license plate which consists of characters and numbers. Myanmar license plates are different in background that are constructed on the type of possession. The present method is trumped up of two major processes: preprocessing and extraction. The preprocessing process begins with binarization and noise removal of the plate. Later these process, the algorithm segments and extracts the characters. Finally, the extracted characters and numbers have been got. The proposed method of overall architecture is shown in Figure 2. The Myanmar character license pate are displayed in Figure 3. Outcomes of the processes are described in the Figure 4.

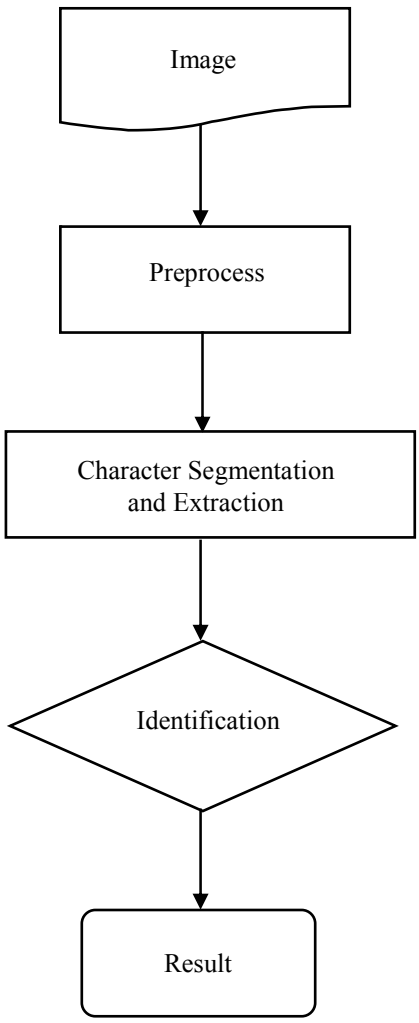


Fig. 2. The overall structure of proposed system

Fig. 3. Myanmar Character License Plates

III. PREPROCESSING

The Myanmar character license plate images are inputted. These input images contain the disparate sizes of characters. Most of the Myanmar characters are horizontal straight upper line. However, numbers are upper and under horizontal straight line. These are described in Figure 3. The input color images are converted into gray scale images and they are binarized using a thresholding technique. The images are also needed to be free noise. Accordingly, the noise is removed from the images. Distinct methods are better for distinct kinds of noise for the removal of the noise. All object containing fewer than 30 pixels are removed and finally, we get the input images without noise. Then further processing will be executed on these images.

IV. CHARACTER EXTRACTION

Myanmar letters and numbers are very similar and complicated. It involves curve, line, circle and dot. The image processing toolbox function is utilized to segment and extract the characters from the license plate image. There are multiple steps in this process. The object extraction is the simple extraction for foreground objects from the image.

Practically, the object extraction is performed the following steps:

Construct `bwlabel and regionprops` of illustrative color for sure forefront and certain background

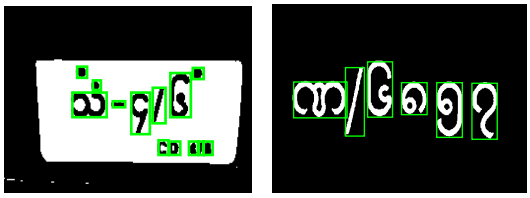
Give all of the image points to forefront or background by a weight nearest neighbor

Implement some standard image processing operations like `bwlabel` and `regionprops` to remove artifacts.

Search the connect foreground components that are either large enough.

It is working on a few other images. It applies multiple enhancement techniques to keep removing it step by step. After finishing, we are successful extracting our regions of interest.

We are doing one more step which are extracting its features with bounding box in the area. Connected components in an image is that all the pixels inside head are connected. If the area of this connected component is greater than threshold area then store its bounding box and display the box around the text region. The text has been detected pretty accurately but have some erroneous extraction. But it has quite good with the complexities. From the character segmented images of Figure 4, the characters and numbers are segmented and extracted with one by one.



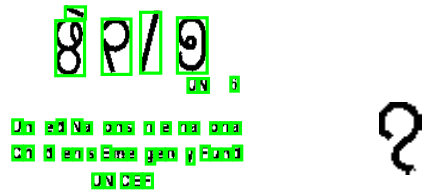


Fig. 4. Letters and Numbers Extraction

V. EXPERIMENTAL RESULTS

An experiment has been carried out on the Myanmar vehicle images. Different types of vehicle images are investigated and the dissimilar character images are experimented. The segmented letters and numbers are depicted in the Figure 4. The extraction results are as follows:

TABLE I. ACCURACY RATE RESULTS

Total 48 letters and 52 numbers	Correct Extraction Rate (%)
	(90/100) ×100 = 90%

VI. DISCUSSIONS

From this study, we can say that almost all character extractions are based on the character sizes and image sizes. In our experiment, the dissimilar license plates are tested. We searched that the characters and numbers sizes affect the proposed approach. The program is experimented with the different character sizes. For the accuracy, medium size characters were used in the system experiment. The larger size characters showed the wrong result. It was also searched that the successful extraction rate was done by the clearance of license plate. The obtained results of the character extraction are displayed in TABLE I. This paper is in progression and we target to make differentiation of other languages or text input method such as Thai languages or Cambodia languages in the near future.

VII. CONCLUSION

This paper has publicized the investigation of Myanmar characters. The main purpose of this research is to examine and extract the numbers and letters from the different Myanmar vehicle license plates. The presented experiments and results are very strong extraction for the various license plates. To verify the present method, the examinations are

worked with the vehicle images. There is the extraction failure. However, we can establish preprocessing that points out an answer to this difficulty. The result using the proposed approach reveals the strong production since 90% of the characters are absolutely extracted.

ACKNOWLEDGMENT

I would like to thank my advisers for their great support for this research work as well as scholarship provided by AUN/SEED-Net (ASEAN University Network/Southeast Asia Engineering Education Development Network).

REFERENCES

[1] O. Khin, M. Phothisonothai and S. Choomchuay, "License Plate Detection of Myanmar Vehicle Images Captured from the Dissimilar Environmental Conditions", *RCCIE 2017*, November 30-December 01, 2017.

[2] O. Khin, M. Phothisonothai and S. Choomchuay, "License Plate Detection of Myanmar Vehicle Images from Dissimilar Angle Conditions", *ICSEC 2017*, *November 15-18, 2017*.

[3] T. Swe and P. Tin, "Recognition and Translation of the Myanmar Printed Text Based on Hopfield Neural Network" *6th Asia-Pacific Symposium on Information and Telecommunication Technologies*, pp. 99-104, 2005.

[4] Y.K. Thu and Y. Urano , "A Comparison of Myanmar PC Keyboard Layouts" *2009 Eighth International Symposium on Natural Language Processing*, pp. 15 - 20, 2009.

[5] K. Sandar, "A Comparison of Recognition for Off-line Myanmar Handwriting and Printed Characters", pp. 2954-2959, 2014.

[6] S.H. Mar and N.L. Thein, "Myanmar Character Identification of Handwriting Between Exhibit and Specimen" pp. 1 – 4.

[7] T.T. Thwin, A.T. Win, P.P. Wai and M.M.S. Thwin, "Proposed Myanmar Word Tokenizer Based on LIPIDIPIKAR Treatise" *2010 2nd International Conference on Computer Engineering and Technology*, pp. 136 - 140, 2010.

[8] H.P.P. Win, P.T.T. Khine and K.N.N. Tun, "OCRMPD: OCR System for Myanmar Printed Document Image with a Novel Segmentation Method and Hierarchical Classification Scheme" pp.285 – 291, 2011.

[9] Y. Win and T. Masada, "Exploring OOV Words from Myanmar Text Using Maximal Substrings" *2016 5th IIAI International Congress on Advanced Applied Informatics*, pp. 657 - 663, 2016.

[10] Y.K. Thu, O. Phavy and Y. URANO, "Positional Gesture for Advanced Smart Terminals: Simple Gesture Text Input for Syllabic Scripts Like Myanmar, Khmer and Bangla" *First ITU-T Kaleidoscope Academic Conference*, 2008.

[11] H.P. P.Win, P.T.T. Khine, "Converting Myanmar Printed Document Image into Machine Understandable Text Format" pp. 96-101, 2011.