2020 59th Annual Conference of the
Society of Instrument and Control Engineers of Japan (SICE)
September 23-26, 2020, Chiang Mai, Thailand

Thai Text Detection and Classification Using Convolutional Neural Network

Susanta Malakar† and Werapon Chiracharit

Department of Electronic and Telecommunication Engineering
Faculty of Engineering
King Mongkut's University of Technology Thonburi, Bangkok, Thailand
(Tel: +66-64-063-3001; E-mail: susanta.malakar@mail.kmutt.ac.th)

Abstract: Many foreign people don't know Thai language and most of the time Thai sign images, posters or text images do not have subtitles in English so, it is very necessary to have a system that can translate Thai text to English. In this paper, MSER and convolutional neural network (CNN) have used to understand Thai text in English. Firstly, region of interest has localized from natural image which is some particular Thai text. Then text has extracted and fed to CNN. We used a 7-layer self-designed CNN model that provides the output with an accuracy of 98%. The proposed system takes natural scene image as input and uses MSER, geometrical properties as well as bounding box algorithm to localize the text area then selected localized areas have fed to CNN and provide an output that has the English meaning for the Thai text image. This paper introduces a new approach of text translation by using image classification method. The proposed system can work on particular inputs which are indoor sign Thai text images.

Keywords: MSER, CNN, Transfer learning, Text Detection, Image Classification.

1. INTRODUCTION

Every year many people include students, working professionals and tourists visit Thailand, most of them who don't know Thai language face problems in their daily life to read sign images or signboards. Many places can be found where important sign images are only written in Thai as shown in Fig. 1. This is a big issue for foreigners especially when they are alone. In order to solve this problem, a system that can provide sufficient information about the Thai text in English is required. This is a challenging task because in a natural scene image there could be many Thai text images near the targeted Thai text image which we want to read in English, as shown in Fig. 2 (a). Furthermore, the image could be in any orientation or the brightness and contrast value can vary for different images as shown in Fig. 2 (b).

(a) (b)

Fig. 1 Sign images written only in Thai, (a) Fire Exit, (b)Toilet is this way.

The proposed research problem mainly belongs to detection and classification problems. There is no existing research paper where this problem has been solved but some authors solved text detection problems and some authors solved image classification problems similar to proposed research problem. In this paper, we will use some existing detection and classification methods that can solve this proposed problem. Therefore, creation of the dataset and introducing this new application can be considered as our contribution.

(a) (b)

Fig. 2 (a) Written Exit in Thai on the top, (b) Written Fire Exit

2. RELETED WORK

In this section, we will discuss about the previous related works.

2.1 Classification

Many papers have used variety of classification methods in order to classify images but not all of them will be suitable for our research problem. In [1], the author classified traffic sign images by considering shape and size properties i.e. circular shape, rectangular shape, etc. We cannot use this classification process as the shape of a text image can vary. In [2], the author mainly used SVM classifier to classify the superimposed text region, they also used AdaBoost and Ostu's classifier but the classification accuracy of these classifiers is less than 80%. In [3], the author used neural network, SVM and HOG for text recognition of

† Susanta Malakar is the presenter of this paper.

signboard images and got 70.2% accuracy. All of these classification processes have not shown good results but the next papers have got high accuracy by using CNN. In [4], author used CLSN Classification Network to classify Korean characters and this classifier provides good accuracy. The author in [5] used CNN classifier to distinguished graphical text images from natural scene images and they get high accuracy for different datasets. The author in [6] used deep-CNN for a large scale of images. The author in [7] used transfer learning to classify flower images and got high accuracy results.

2.2 Text Localization

In order to understand a particular text from natural scene image, text localization could be the first important task. There are many existing text localization methods but maximally stable extremal regions (MSER) [8, 9] has shown the best results for text localization.

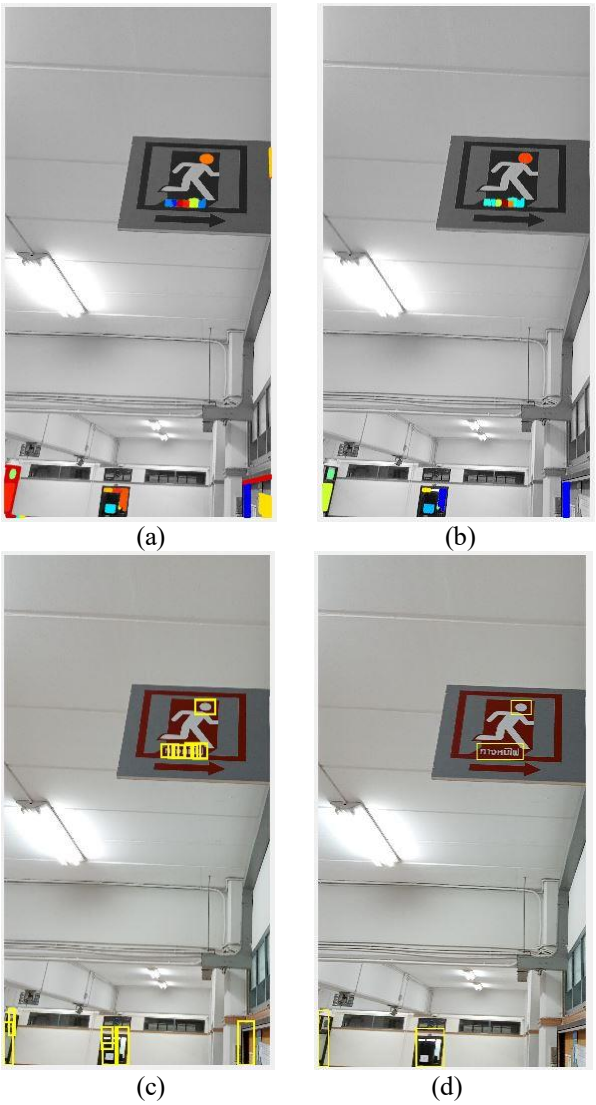
3. PROPOSED METHOD

From the literature review, it can be concluded that, CNN provides better results than other classifiers but in proposed classification problem, a classifier is needed which can give high accuracy because before classification when the target text image will be extracted from natural scene image then, there could be many similar-looking other text images in the natural scene image, therefore, those similar-looking text images will also get high accuracy, that is why the classifier needs to be very robust. In the proposed research problem, text detection from natural scene image is also a major part because the detected image should only contain text area, if the detected image contains pixels of text areas plus more pixels which do not belongs to text area then during classification process, the classifier will get extra features for the extra pixels and the classification accuracy will decrease.

3.1 Text Localization

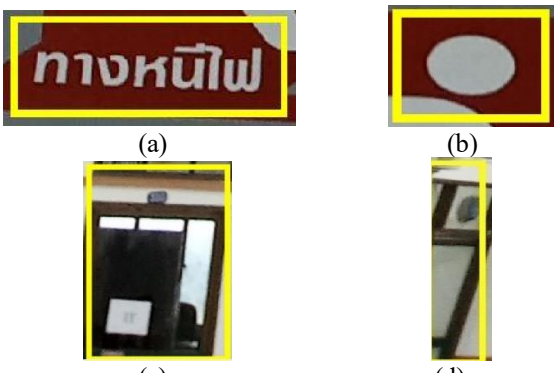
Very simple combinations of techniques have used to localize the text. First, the input RGB image has converted into grayscale image then MSER algorithm has used to get MSER region for a particular threshold value, at this point many unnecessary MSER regions can found which are not text, so in order to remove these non-text areas, some basic geometrical properties of text have used. Then bounding box algorithm has been used. Next, the bounding boxes which are overlapping to each other are merged to get final region of interest images, this whole process has shown in Fig. 3 step by step.

In Fig. 3 (d) the final region of interest has localized but for most of the cases, there will be few unnecessary localized areas. Now each of these localized areas has extracted and fed to CNN. The extracted localized areas have been shown in Fig. 4.



(a) (b) (c) (d)

Fig. 3 Steps of image localization. (a) detected MSER regions, (b) MSER regions after using geometric properties, here some regions have been removed, (c) image after using bounding box algorithm, (d) image with final detected regions.



(a) (b) (c) (d)

Fig. 4 Extracted images from localized area, (a) Text containing image. (b) non-text containing image, (c) non-text image, (d) non-text image.

3.2 Text Classification

After text localization step all the localized areas have fed to the CNN and images which have accuracy higher than 70% is considered as successfully classified because at the end of text localization step many unnecessary localized images can be found with the text images as shown in Fig. 4. Only text images which looks similar to our CNN training data will get accuracy higher than 70% as shown in Fig. 5. The overall view of proposed method has shown in Fig. 6.

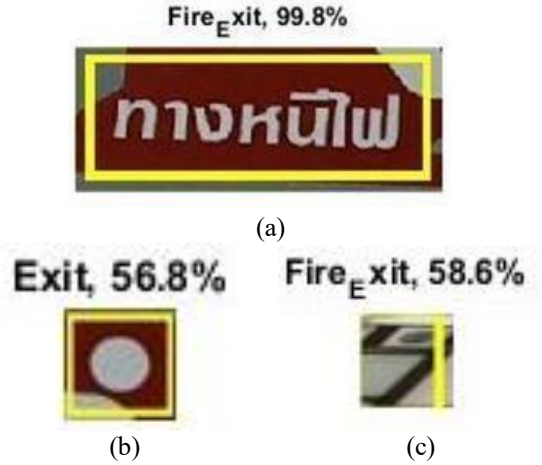


Fig. 5 Classification results of high accuracy and low accuracy. (a) This image shows Fire exit in Thai and system can classify this image in Fire Exit category with accuracy 99.8% whereas (c) and (d) has very low accuracy because these are not real exit and fire exit text image.

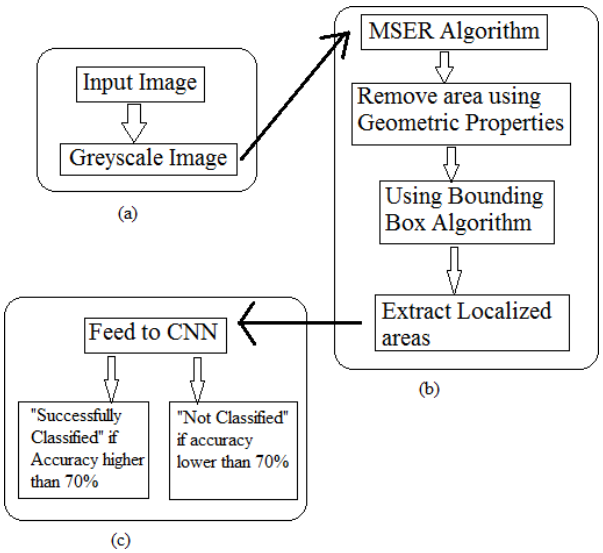


Fig. 6 Overall view of the proposed system. (a) Shows the prepossessing step, (b) shows Text Localization part, (c) shows classification part

For classification we used self-designed convolutional neural network which has only seven layers such as 'image input layer (input size)', 'convolution 2d layer', 'batch normalization layer', 'relu layer', 'fully connected layer', 'softmax layer' and 'classification layer'. In order to evaluate the seven-layer CNN's performance, we use another CNN model which is ResNet-50. We did not implement Resnet-50, in exchange we used transfer learning method because in transfer learning method a pre-trained CNN can be used without implementing the whole network.

4. EXPERIMENTAL RESULT AND DISCUSSION

For experiment we created two different datasets by our-self.

1. Dataset for text detection from natural image and testing with CNN.
2. Dataset for CNN training.

No.1 dataset has 60 natural images and No.2 dataset has 2400 images in six different classes. Most of the images in dataset No.1 has been captured by camera and some has been downloaded from internet. Most of the images in dataset No.2 has been created by downloading images from internet and then edited to get only text area for training. Some images from No.1 dataset and No.2 dataset has been shown in Fig.7 and in Fig.8 respectively.

Fig. 7 some images from dataset No.1

Fig. 8 Some images from dataset No.2.

Some results of the proposed system have been shown in Fig. 9. In Table 1(a) and Table 1(b), we have shown the image classification results for two different CNN. The overall accuracy of image classification can also be considered as the overall system accuracy because the CNN classify those images, which has been detected by MSER so the image classification accuracy depends on the MSER performance. From Table 1(a) and Table 1(b) we can understand that the proposed combination of detection and classification methods can solve our proposed problem. We used two tables

because in order to fit the results of six classes in one table, the table need much more space in horizontally so, in each table we have shown results for three classes.

Fig. 9 some of the final test results.

Table 1(a) Comparison of image classification result

Methods	Accuracy Level				
	Number of Training Images(6 classes)	Over all (%)	Exit (%)	Fire exit (%)	Washroom (%)
MSER + 7 Layer CNN	2400	99.31	99	92	96
MSER + Transfer Learning (ResNet-50)	2400	92.85	90	95	98

Table 1(b) Comparison of image classification result

Methods	Accuracy Level				
	Number of Training Images (6 classes)	Over all (%)	Men Toilet (%)	Women Toilet (%)	Toile (%)
MSER + 7 Layer CNN	2400	99.31	98	97	90
MSER + Transfer Learning (ResNet-50)	2400	92.85	90	93	91

4.1 Limitation

The overall system has some limitations which are,

1. For night images of dark images, the performance will be low.
2. The CNN in the proposed system can only classify six types of images such as, Exit, Fire Exit, Male Washroom, Female Washroom, Washroom sign text images.

5. CONCLUTION

In this paper, a combination of text detection method and image classification method has proposed in order to understand the Thai sign text images in English. The proposed system has been constructed by simple existing methods such as MSER for text detection from natural scene images and CNN for image classification. The novelty of this paper is Thai text can be translated by image classification furthermore, the proposed system can work for variety of fonts. In future we intend to design a deep learning-based system which can provide more information in English for all indoor and outdoor-based Thai sign text images.

ACKNOWLEDGMENT

I would like to thank my friend and lab mates for their help. I also want to thank my brother for supporting me and keep motivating me.

REFERENCES

[1] F. Paulo and P. L. Correia, "Automatic Detection and Classification of Traffic Signs," Eighth International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '07), Santorini, pp. 11-11, 2007.

[2] S. Satwashil1 and V. R. Pawar, “Integrated Natural Scene Text Localization and Recognition”, ICECA, Coimbatore, pp. 371-374, 2017.

[3] M. A. Panhwar, K. A. Memon, A. Abro, D. Zhongliang, S. A. Khuhro and S. Memon, "Signboard Detection and Text Recognition Using Artificial Neural Networks", ICEIEC, Beijing, China, pp. 16-19, 2019.

[4] H. Eun, J. Kim, J. Kim and C. Kim, "Fast Korean Text Detection and Recognition in Traffic Guide Signs", VCIP, Taichung, Taiwan, pp. 1-4, 2018.

[5] M. Ghosh, H. Mukherjee, S. M. Obaidullah, K. C. Santosh, N. Das and K. Roy, "Identifying the Presence of Graphical Texts in Scene Images using CNN", ICDARW, Sydney, Australia, pp. 86-91, 2019.

[6] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-scale Image Recognition”, Proceedings of ICLR, 2015.

[7] E. Cengıl and A. Çinar, "Multiple Classification of Flower Images Using Transfer Learning", IDAP, Malatya, Turkey, pp. 1-6, 2019.

[8] A. Tabassum and S. A. Dhondse, “Text Detection Using MSER and Stroke Width Transform,” 2015 International Conference on Communication Systems and Network Technologies, Gwalior, pp. 568-571, 2015.

[9] S. Choudhary, N. K. Singh and S. Chichadwani, “Text Detection and Recognition from Scene Images using MSER and CNN,” 2018 International Conference on Advances in Electronics, Computers and Communications (ICAECC), Bangalore, pp. 1-4, 2018.