Detection of a Key String from Scene Images Using Saliency

Kota Oodaira, Tomo Miyazaki, Yoshihiro Sugaya, and Shinichiro Omachi
Department of Communications Engineering,
Graduate School of Engineering, Tohoku University
Sendai, Japan
hira@iic.ecei.tohoku.ac.jp

Abstract— **Recently the scene text detection and recognition are used in various places. For example, there are an automatic translation system to convert the text written in a foreign language into a native language, and a system to make books available in computer as data and so on. However, it is inefficient to detect and recognize all the texts in a scene image because there is useless information for users in a scene image. In this paper, we present a method to detect a key character string in a scene image. A key character string means an important letter. This method enables users to only detect the text which users need. In this method, we detect character candidates by using edge detection at first. Secondly, we determine the candidates of which a string is composed and detect character string regions. Finally, we determine a key character string region by using saliency map. We tested the effectiveness of this method and succeeded in detection of a key string in 37 out of 56 scene images.**

Keywords-text detection; scene images; saliency map;

I. Introduction

We get most of necessary information from vision, and there is text information everywhere around us. Therefore, we can get a variety of benefits if we detect and recognize those texts. OCR is a technique to read handwritten or printed characters by an image scanner or a digital camera and convert them into digital character codes.

However, since there are many characters around us, we may get useless information and that may cause overlooking of necessary information. In this paper, a method which detects key character strings on scene images is proposed. The aim is to make character detection and recognition beneficial by detecting and recognizing an only character string which is important for a user.

II. PROPOSED METHOD

A. *Features of key strings*

We performed a subjective evaluation experiment to investigate what kind of features key character strings have. In this experiment, subjects answer a character string which they think is important for each of 45 scene images.

We give consideration on features of key character strings. From the results, many subjects answered “WARNING” in Figure 1 and “CHINA AMERICAN INN” in Figure 2 are important. This means the more important a character string is, the more human beings tend to pay attention to it because a character is designed so that they can easily pay attention. So it is considered that “WARNING” and “CHINA AMERICAN INN” are selected because their sizes are large and their colors are vivid. Therefore, it is an important factor whether human beings can easily pay attention to the character string to calculate importance of the string. Then we propose a method using saliency maps that express which part of a scene image human beings pay attention to.

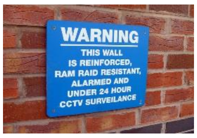


Figure 1. Headline

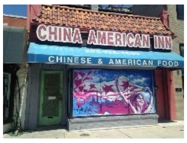


Figure 2. Shop name

Also, “WARNING” in Figure 1 is a headline and “CHINA AMERICAN INN” in Figure 2 is a shop name. Therefore, it is considered that character strings belonging to those categories are important.

It is important to judge whether character strings are important in consideration of their meanings. However, this time we judge whether a character string is important or not by using only visual characteristics.

In addition, we made correct answer data of experiment III from this experiment.

B. *Detection of a key string*

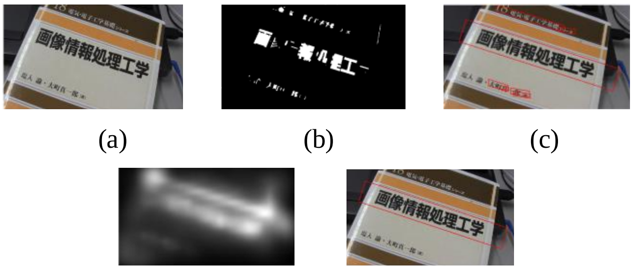
In this section, we propose a method which detect a key character string in a scene image. This method judges whether a character string is important or not after detecting character string regions.

Characters in a scene image have high edge strength. Therefore, we detect character strings by using the Sobel filter. Also, we assume that the edge detected from a character string region becomes the closed contour, so we consider a region whose edge becomes the closed contour to be a character candidate. Next, we create a group of the character candidates belonging to the same character string to be able to detect strings from character candidates. We distinguish whether character candidates belong to a same character string using certain conditions. Those conditions are whether character candidates have similar height value, whether the luminance patterns of character candidates are similar on grayscale image and whether distance between character candidates is small. Finally, the groups of character candidates which belong to strings are created.

After groups of character candidates are created, we estimate the angle of the character string by using the center of gravity coordinates of their character candidates. We consider a median value of angle calculated from coordinates of character candidates belonging to a same string to be the angle of the character string.

However, if the edge detected from a character string region does not become the closed contour, we cannot detect character strings correctly. Therefore, we detect the final character string regions whose character strings have all the characters belonging to them. We calculate differences between maximum and minimum of luminance in a character string region and confirm whether peripheral regions have similar value. We enlarge a character string region in lateral direction of it as far as a similar value is detected. After this process, we can detect a whole character string region.

We detect the character string regions by the processing mentioned above. Finally, we calculate the total value of the saliency map on character string regions and finally define the region having the highest value as a key character string region. These processes are shown in Figure 3.



(a) (b) (c)

(d) (e)

Figure 3. Detection process. (a) Input image. (b) Character candidates. (c) Result of text detection. (d) Saliency map. (e) Detection result of a key string.

III. EXPERIMENT AND RESULTS

We performed an experiment to verify if this method detects the key character string region. In the experiment, we used the images which we collected from the website “flickr.” We mainly assumed shop names and headlines key character strings this time. The reason for which we chose the images is that there are plural strings in each image and we can clearly judge key character strings that are shop names or headlines.

A. *Comparison of existing saliency maps*

First, we examined which saliency map is the most effective for our method. We used 50 images in this experiment.

In this experiment, we decide character string regions in advance and define a character string that has the highest value which is a total of the saliency map in the character string region as a key string.

We used Harel’s method, Hou’s method and both to detect a key character string. Experimental results are presented in Table I. As shown in Table I, using both methods is the most accurate.

Table I. Experimental results

Method	Harel	Hou	Harel + Hou
Accuracy [%]	88	92	94

B. *Detection of a key string*

Secondly, we detected key character strings using the proposed method. We performed the experiment on 56 images using both Harel’s method and Hou’s method.

As a result, we succeeded in the detection of a key character string in 37 out of 56 images. Figure 4 is a successful example. However, in some images, detection was failed. In Figure 5, the region of the object was detected. It is considered that only edge strength is used in finding character candidates. Therefore, it is necessary to add the other elements to our method in order to detect a key character string.

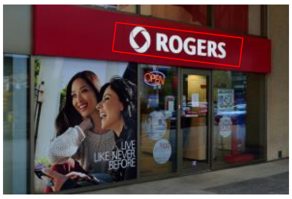


Figure 4. Successful example

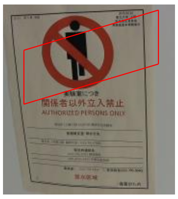


Figure 5. Failure case

IV. CONCLUSION

In this paper, we proposed a method which detects a key string in a scene image. Improvement of the accuracy of detection is a future work.

ACKNOWLEDGMENT

Part of this work was supported by JST, CREST and JSPS KAKENHI Grand Number 16H02841.

REFERENCES

[1] L.itti, C. Koch, and E. Niebur. “A model of saliency-based visual attention for repid scene analysis,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 11, pp. 1254-1259,1998

[2] J. Harel, C. Koch, and P. Perona: “Graph-Based Visual Saliency,” Proceedings of Neural Information Processing Systems (NIPS), 2006

[3] X. Hou, J. Harel, and C. Koch: “Image Signature: Highlighting Sparse Salient Regions,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 1, pp. 194-201, 2012