Feature Point Based Text Detection in Signboard Images

Chien-Cheng Lee and Shang-Fei Shen

Department of Communication Engineering, Yuan Ze University
135 Yuan-Tung Road, Chung-Li,
Taoyuan County 320, Taiwan
cclee@saturn.yzu.edu.tw

Abstract

This paper presents a method of using feature points to locate text area for signboards on street view images. The FAST corner detection was applied for the first step. FAST corner detection is fast and stable enough to retrieve potential text regions on street view images. The characteristics of each feature point color space were used to compute the color histogram and related information. For the second step, we used a gravity clustering method to find clusters of text area on signboard images and got the possible positions of the text area. For the third step, the distribution density was estimated and the average distance of feature points was calculated on the possible text area. The average distance was used to build text pattern regions. These regions were processed by the following steps: morphological closing, image binarization, and minimum bounding box finding to obtain a complete text region. Experimental results have shown the advantages and effectiveness of the proposed method in the text detection in the signboard images.

Key words: feature point detection; text detection; signboard images

Introduction

Text recognition can be used in many applications in smart life. It helps a machine to sense the human society and understand surrounding things. Text detection and localization are important preprocessing steps before text recognition. Text detection roughly classifies text and nontext regions. Text localization determines the accurate boundaries of text strings. After the text areas are localized, the optical character recognition (OCR) software can recognize the text easy.

Although many methods have been proposed for text detection and localization, most of them are texture-based [1, 2] , edge-based [3, 4], or connected component-based [5, 6] methods. However, many difficulties are still needed to be solved, such as complex background, unknown text color, and degraded text quality. These difficulties emerge significantly in natural scene images because there are many uncertainties in the natural scene.

In this study, we focus on natural scene images. We proposed a novel feature point based text detection approach to locate the text area for signboards on street view images. In Section II, we presented our feature point based text detection method of signboard images. Comparison results are presented in Section III. Finally, in Section IV, we draw conclusions.

Methods

Our approach can be subdivided into three major steps: 1) feature point detection, 2) feature point clustering, and 3) minimum bounding box localization. A FAST (Features from Accelerated Segment Test) corner detection used in the first step is a well-known algorithm; therefore, the first step was just briefly summarized here and the majority of this study focuses on the later steps.

First, the image color space was converted to HSV color space; then, the FAST corner detection was applied to extract feature points. Second, these feature points were clustered according to their color space distribution. Finally, a following image processing procedures was performed to obtain the complete minimum bounding box of the text area.

A. Color Space Transform and Feature Point Detection

The idea of our proposed approach is that the text color is different from the background color so the text can be seen easily. To facilitate the differentiation of the text from the background, the color space of an image was converted to HSV color space first. Then, the FAST corner detection was applied in the hue image to extract feature points. An example result of the color space transformation and feature point detection is shown in Fig. 1.

B. Feature Point Clustering

After the feature point extraction, we computed the color histogram for each feature point by selecting an $M \times N$ block centered around the point. Next, kernel density estimation with Gaussian kernel was used to smooth the histogram to find a smoother histogram, as shown in Fig. 2. Then, the maxima of the histogram could be found easily.

If the feature point belongs to text, the smoothed color histogram should have a bi-modal shape, as shown in Fig. 3. Otherwise, the feature points without bi-modal shape should be filtered out. We made $k_1(i)$ and $k_2(i)$ be the first highest and the second highest peaks of the smoothed histogram of feature point $p(i)$. The point set $M(i)$ is defined as:

$$M(i) = \begin{cases} point\big(k_1(i), k_2(i)\big), if\ k_1(i) < k_2(i) \\ point\big(k_2(i), k_1(i)\big), if\ k_1(i) \geq k_2(i) \end{cases} \qquad (1)$$

(a) (b) (c)

Fig. 1 Example result of the color space transformation and feature point detection. (a) original image, (b) hue image of HSV color space, (c) feature points extracted from the hue image.

From the map of the point set $M(i)$, it can be observed that the feature points form a cluster when these points have the similar bi-modal histogram, as shown in Fig. 4(a). That is, these feature points have the similar foreground and background color in the image. In other words, these points belong to the same text block. In this study, gravitational clustering algorithm [7, 8] was used to cluster the point set. The clustering result is shown in Fig. 4(b). The feature points in the most compact cluster (red points in Fig. 4(b)) belong to the same text block. These points were superimposed on the original image and are shown in Fig. 4(c).

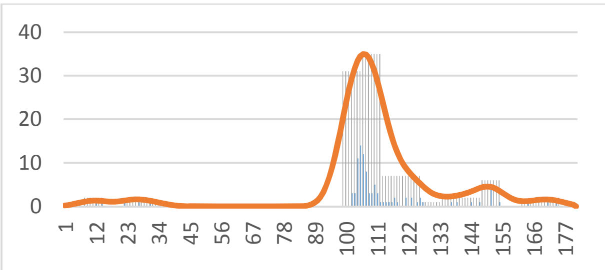


Fig. 2 Example of color histogram. The smooth solid line is the result of kernel density estimation.

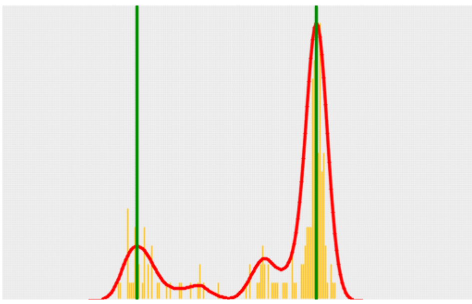


Fig. 3 Bi-modal shape.

C. Minimum Bounding Box Localization

From Fig. 4(c), we can see that most of these points construct the text block. However, there are some unwanted points outside the text block. To remove the points outside the text block and find the minimum bounding box of the text block, the well-known segmentation method of vertical and horizontal projection was used in this study. The vertical and horizontal projection histograms were denoted by $V(x)$ and $H(y)$, and the value indicated the number of feature points in the $Column(x)$ and $Row(y)$, respectively, as shown in Fig. 5(a). Then, a text element T with width T_w and height T_h is defined as:

$$T_w = \frac{\sum_{x=R_{i1}}^{R_{i2}} V(x)}{R_{i2}-R_{i1}} \tag{2}$$

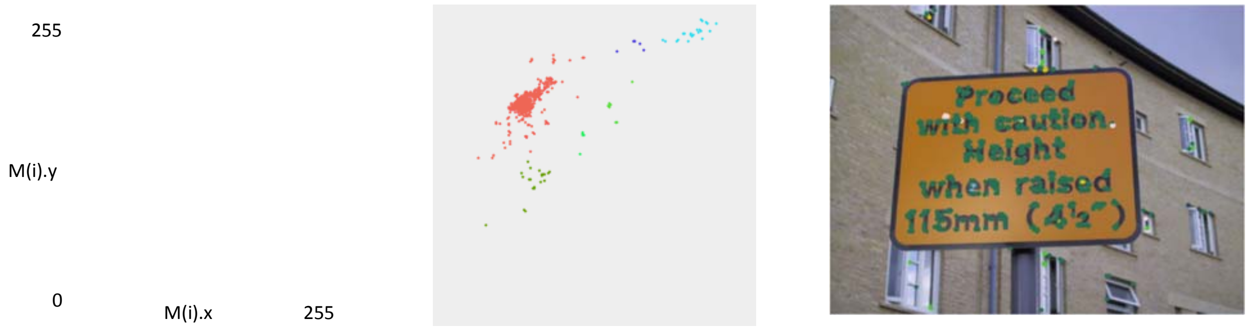
$$T_h = \frac{\sum_{y=C_{j1}}^{C_{j2}} H(y)}{C_{j2}-C_{j1}} \tag{3}$$

where interval $[R_{i1}, R_{i2}]$ represents the continuous row segment i with $V(x) > \delta$, interval $[C_{j1}, C_{j2}]$ represents the continuous column segment j with $H(y) > \delta$, and δ is a pre-defined threshold.

Using the text element, a text map image I is determined as follows:

$$I(x,y) = \begin{cases} 255, & pixel\ (x,y) contained\ in\ T\ centered\ around\ p(i) \\ 0, & otherwise \end{cases} \tag{4}$$

Next, a morphological closing operation with 5x5 structuring element was performed on the text map image five times, as shown in Fig. 5(a). Finally, minimum bounding boxes were localized for the connected components which their sizes were greater than a given threshold. The final result is shown in Fig. 5(b).



(a) (b) (c)

Fig. 4 (a) point set $M(i)$. (b) clustering result. (c) superimposed image of the original image with the most compact cluster points.

Experimental Results

In this experiment, the proposed method was compared with two datasets. One was ICDAR 2013 dataset, the other was our collection from streets in Taiwan. In the ICDAR dataset, our performance was medium. The recall, precision, and Hmean evaluation scores were about 0.7. However, we used the same evaluation to test the Taiwan street scene dataset. The scores were about 0.8. We also compared with the BUCT_YST algorithm. Table I and II show the comparison results of our method and the BUCT_YST method for these two datasets. Some result images are shown in Fig. 6.

The comparison results show that our method demonstrates good performance in Taiwan street scene images, but gets medium performance in ICDAR dataset. It seems reasonable because most texts in ICDAR dataset were western characters. The western characters contain less junction, turning, and end points than Chinese characters. Hence, the number of feature points in western characters is less than that in Chinese characters. These results lead to medium performance. In contrast, BUCT_YST algorithm got low scores in Taiwan street scene images. It seems that the BUCT_YST algorithm was based on neural network training. The performance degradation came from the unseen in the training phase.

TABLE I
PERFORMANCE COMPARISON FOR ICDAR DATASET

Method	Recall	Precision	Hmean
BUCT_YST	0.74	0.85	0.79
Our method	0.74	0.70	0.72

TABLE II
PERFORMANCE COMPARISON FOR TAIWAN STREET SCENE

Method	Recall	Precision	Hmean
Our method	0.79	0.81	0.80
BUCT_YST	0.35	0.34	0.34

Conclusions

This paper proposes a new text detection method. The method was based on the feature point extraction and the color distribution of the feature points. According to our experiments and analysis, the proposed method can be applied on signboards on street view images. The proposed method was also robust to various font sizes, font styles, and contrast levels. A known limitation of the current method is that the performance will be degraded if the text contains less feature points, such as junction, turning, and end points. These issues will be addressed in our future research.

Acknowledgments

Thanks for the Ministry of Science and Technology (Grant number: MOST 104-2218-E-155-002) for funding this work.

References

[1] S. A. Angadi and M. M. Kodabagi, "A texture based methodology for text region extraction from low resolution natural scene images," *International Journal of Image Processing (IJIP),* vol. 3, pp. 121-128, 2010.

[2] K. I. Kim, K. Jung, and J. H. Kim, "Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 25, pp. 1631-1639, Dec 2003.

[3] X. Liu and J. Samarabandu, "Multiscale edge-based text extraction from complex images," *IEEE International Conference on Multimedia and Expo,* pp. 1721-1724, 2006.

[4] X. Liu and J. Samarabandu, "An edge-based text region extraction algorithm for indoor mobile robot navigation," *IEEE International Conference on Mechatronics and Automation,* vol. 2, pp. 701-706, 2005.

[5] H. Wang, "Automatic character location and segmentation in color scene images," *IEEE International Conference on Image Analysis and Processing,* pp. 2-7, 2001.

[6] H. Wang and J. Kangas, "Character-like region verification for extracting text in scene images," *12th International Conference on Document Analysis and Recognition (ICDAR),* pp. 957-962, 2001.

[7] W. E. Wright, "Gravitational clustering," *Pattern Recognition,* vol. 9, pp. 151-166, 1977.

[8] J. Gomez, D. Dasgupta, and O. Nasraoui, "A New Gravitational Clustering Algorithm," in *Proceedings of the 2003 SIAM International Conference on Data Mining*, ed, pp. 83-94.

(a) (b) (c)

Fig. 5 (a) vertical and horizontal projection (b) text map image after morphological closing. (c) minimum bounding boxes for the text area.

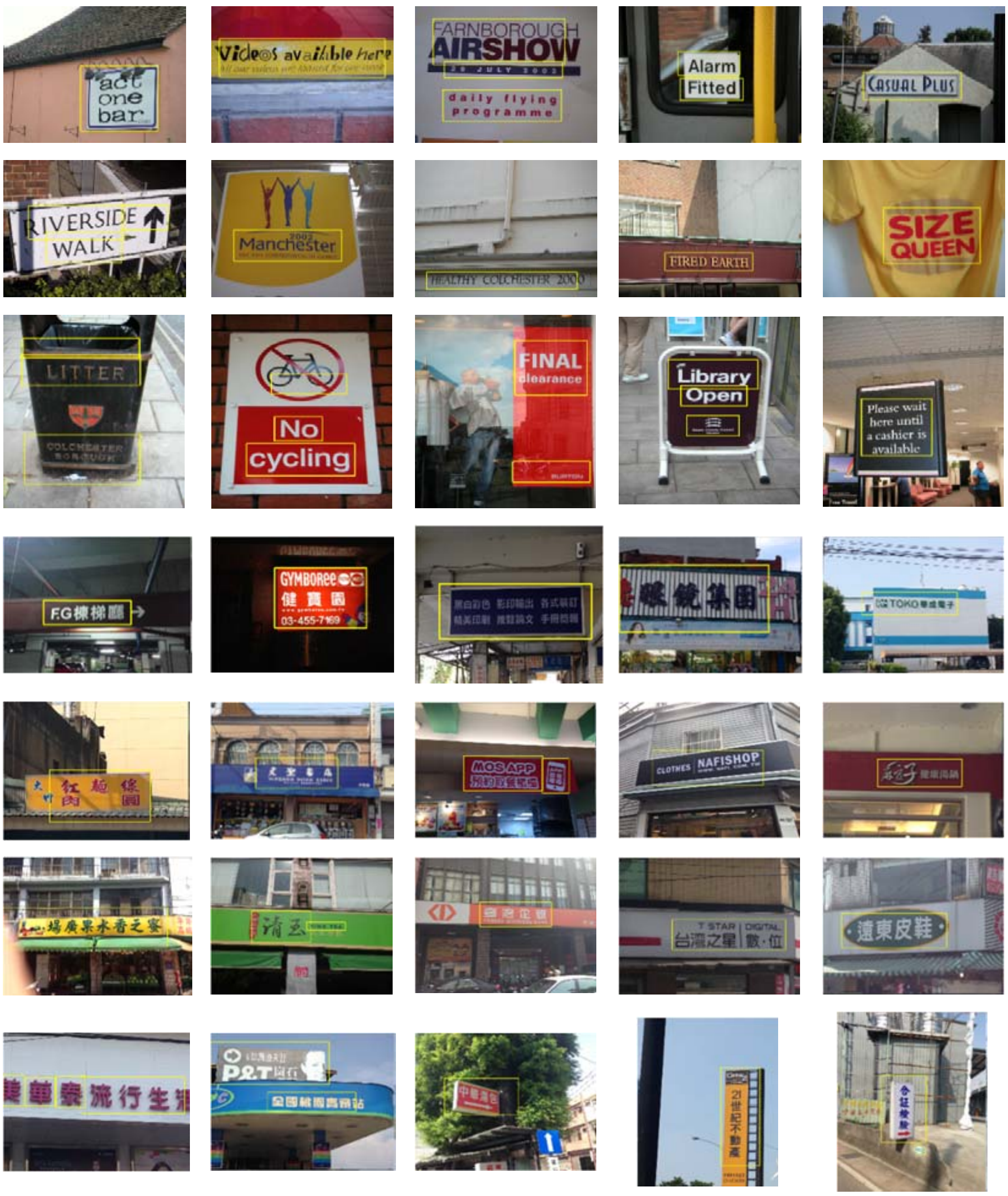


Fig. 6 Examples of experimental results.