IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 15, 2020, accepted March 31, 2020, date of publication April 8, 2020, date of current version April 23, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.2986500*

# Mixed Vertical-and-Horizontal-Text Traffic Sign Detection and Recognition for Street-Level Scene

**JIEFENG GUO[1], (Member, IEEE), RONGXUAN YOU[1], AND LIANFEN HUANG[2]**
[1]School of Electronic Science and Engineering, Xiamen University, Xiamen 361005, China
[2]School of Informatics, Xiamen University, Xiamen 361005, China

Corresponding author: Lianfen Huang (lfhuang@xmu.edu.cn)

This work was supported in part by the Natural Science Foundation of Fujian Province of China under Grant 2019J01046, and in part by the Science and Technology Plan Project of Xiamen City of China under Grant 3502Z20183008.

**ABSTRACT** Much effort has been dedicated to text-based traffic sign detection and recognition. However, there are still two problems. First, unlike English traffic signs with only horizontal text, Chinese traffic signs have both horizontal and vertical text. To the best of our knowledge, there is nothing in the literature about simultaneous recognition of both horizontal and vertical text in Chinese text-based traffic signs. Second, most existing methods focus on wild and expressway scenes; few focus on street scenes. To solve these problems, we propose a mixed vertical-and-horizontal-text traffic sign detection and recognition algorithm for street-level scene. First, an effective combination of different red, green and blue components is used to distinguish the traffic signs from many objects of similar color in the very complex street scenes. Second, unlike English letters, the strokes of many Chinese characters are unconnected, which may result in that a character will be detected as two or more characters. Unlike the English text lines, which are only horizontal, the Chinese text lines on text-based traffic signs are usually both in horizontal and vertical directions. Our proposed method uses the position and structural information of the characters to form the text lines. A dataset of Chinese text-based traffic signs is collected. Experimental results indicate the effectiveness of the proposed method.

**INDEX TERMS** Text recognition, text-based traffic sign recognition, traffic sign detection, text boxes.

## I. INTRODUCTION

When a map company wants to build a map for a city or an area of a city, the text-based traffic sign detection and recognition system can be used to collect the information of road automatically and quickly. This system is deployed on the vehicle camera. To build a map, the vehicle equipped with the text-based traffic sign detection and recognition system just needs to take a turn on the road, the information of road can be collected automatically and quickly. Then, those collected text-based information will be used to correlate with the road location provided by the global positioning system (GPS). Using the text-based traffic sign detection and recognition system, it will be faster and cheaper to build a map for a city.

The design of detection and recognition algorithms must consider the scenes. The scenes of text-based traffic sign detection and recognition can be classified into two categories: 1) wild and expressway scenes; 2) street scenes. From Figure 1, it can be seen that the wild and expressway scenes [1] have simple backgrounds with large areas of the sky, the boundless mountains, etc. Conversely, the street-level scenes have complex backgrounds with the high buildings, the crowds of pedestrians and vehicles, etc. However, the previous works [2], [3] focused on text-based traffic signs detection in the wild and expressway scenes. On the crowded street-level road, there are many objects that have similar appearances to text-based traffic signs (e.g. the abundant street billboards and the rectangular vehicles), which are often detected falsely as text-based traffic signs. Therefore, we propose a text-based traffic sign detection and recognition algorithm, which can be applied to the street scenes.

The majority of the existing works [4]–[10] focus mainly on the scene text detection. Some works [11], [12] focus on text-based traffic signs in English. Only a few works [13], [14] focus on text-based traffic signs in Chinese. Extracting Chinese text from text-based traffic signs is more difficult than extracting English text. Actually, as shown in the Figure 2(a), most of these works are dedicated to the horizontal text detection of text-based traffic signs. However, as shown in Figure 2(b), both vertical and horizontal texts are mixed on Chinese text-based traffic signs.

To the best of our knowledge, there is nothing in the literature about simultaneous recognition of both horizontal and vertical texts in Chinese text-based traffic signs. Therefore,

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Awais Javed.


(a) (b) (c)

**FIGURE 1.** **Three kinds of scenes including text-based traffic signs. (a) Street scene. (b) Wild scene. (c) Expressway scene. (b) and (c) are from the Traffic Guide Panel Dataset [3].**


(a) (b)

**FIGURE 2.** **Text-based traffic signs with vertical or/and horizontal text. (a) includes only horizontal text. (b) includes both vertical and horizontal text. The red and green bounding boxes denote the horizontal and vertical text, respectively.**

the component-based text line forming method is proposed to solve the problem. Compared with the recognition of English letters, the recognition of Chinese characters is more difficult for the following three reasons: 1) The number of Chinese characters is far greater than the number of English letters. 2) Many Chinese characters are very similar to each other. 3) The strokes of many Chinese characters are unconnected, which causes a character to be detected as two or more characters. Our aim is to detect the text-based traffic signs in street scenes and recognize the Chinese texts on these traffic signs, as shown in Figure 3.

The flowchart of the proposed text-based traffic sign detection and recognition algorithm is illustrated in Figure 4. Many computer vision methods are used to obtain the traffic information in intelligent transportation systems [15], [16]. Simply using structural information, such as edges, lines, corners, etc., to detect text-based traffic signs is not suitable for street scenes. Therefore, an efficient color segmentation method [17] was proposed to detect the blue text-based traffic signs. However, two problems are inherent in this method:

Problem 1 - dark colored objects: In HSV color space, the hue component is suitable for detecting the color of blue. However, some dark colored objects that have relatively high value of blue components, such as the blue-green vehicles and the glass windows of buildings, are easy to be mistakenly detected as candidate regions. To solve this problem, we propose a new color segmentation method to distinguish the dark blue objects from the blue text-based traffic signs.

Problem 2 - object size: This method detects all blue objects as text-based traffic signs, no matter whether the blue objects are large or small. However, these small objects are almost impossible to obtain the useful information because of the small scale. Therefore, these small regions should be removed.

Actually, the Chinese characters on text-based traffic signs are always in the form of a text line. Therefore, text lines detection is a suitable way to solve this problem. As shown in Table 1, the majority of the existing works [3], [12], [17] only consider the detection of horizontal English text lines on traffic signs. Although Wang's method [5] can detect Chinese text lines on traffic signs, only horizontal text lines can be detected. However, there are both horizontal and vertical text lines in the Chinese traffic signs. To solve this problem, a new component-based text lines forming method is proposed. In addition, the strokes of English letters are connected, but the strokes of many Chinese characters are not connected. Wang's method only uses the spatial location information of text lines to detect text lines. Instead, the proposed method makes full use of the position and structural information of the Chinese characters. Therefore, the unconnected parts belonging to a Chinese character are merged into a character. Then, these characters are merged into a text line. The relationship between the horizontal and vertical text lines is analyzed to determine whether the overlap is horizontal or vertical.

The main contributions of this paper are summarized as follows: 1) Unlike English traffic signs with only horizontal text, Chinese traffic signs have both horizontal and vertical text. To the best of our knowledge, there is nothing in the literature about simultaneous recognition of both horizontal and vertical text in Chinese text-based traffic signs. We propose a mixed vertical-and-horizontal-text traffic sign detection and recognition for street-level images. 2) Most existing methods focus on the wild and expressway scenes; few focus on street scenes. An effective model of different red, green and blue components is used to distinguish the traffic signs from many objects of similar color in the street scenes, which are much more complex than the wild and expressway scenes. 3) Unlike English letters, the strokes of many Chinese characters are unconnected, which may result in that a character will be detected as two or more characters. The proposed method uses the position and structural information of the characters to form the text lines to avoid this situation.

The rest of the paper is organized as follows: Section II introduces the method of text-based traffic signs detection. Section III illustrates the method of characters detection and

(a) (b) (c)

**FIGURE 3. Illustration of Chinese text-based traffic signs detection and recognition. (a) Input image. (b) Expected detection result. (c) Expected recognition result.**

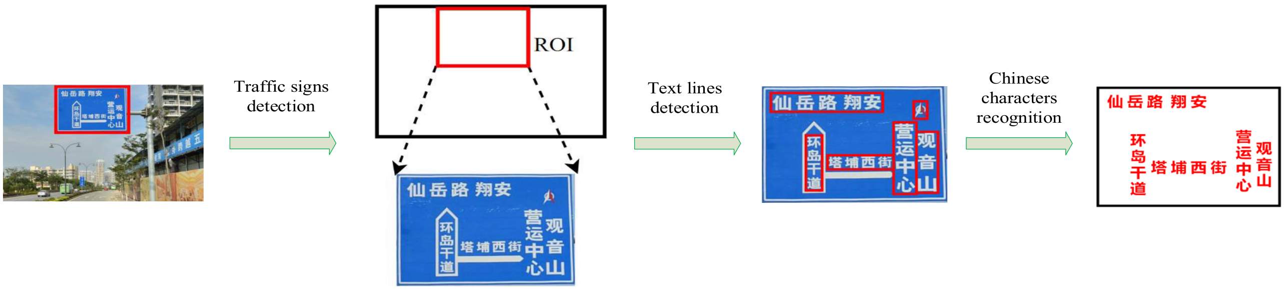



**FIGURE 4. Flowchart of the proposed method for text-based traffic signs detection and recognition. A street scene image is used as an input. Text-based traffic signs are obtained by a cascaded color segmentation method. After applying a new component-based text line forming method, Chinese characters are recognized by an efficient Chinese characters recognition algorithm.**

**TABLE 1. Comparison of the proposed method with other methods.**

| Method | Language | Strokes of characters | Merging into a character | Direction | Text Recognition |
|---|---|---|---|---|---|
| Wang's [5] | Chinese | unconnected | No | horizontal | No |
| Zhu's [12] | English | connected | No | horizontal | No |
| Rong's [3] | English | connected | No | horizontal | Yes |
| González's [17] | English | connected | No | horizontal | Yes |
| Proposed | Chinese | unconnected | Yes | both horizontal and vertical | Yes |

recognition on traffic signs. Section IV presents and analyzes the experimental results. Section V gives the summary and an outlook for future work.

## II. CASCADED COLOR SEGMENTATION METHOD FOR STREET-LEVEL TRAFFIC SIGNS DETECTION

We propose a cascaded color segmentation method to detect text-based traffic signs in the street scene images [18]. In González's method [17], some dark colored objects, which are similar to the blue objects, are retained. To distinguish the dark blue objects from the blue text-based traffic signs, a new cascaded color segmentation algorithm is proposed. The flowchart of the proposed method for text-based traffic signs detection is illustrated in Figure 5. First, the candidate regions are obtained from street scene images by the cascaded color segmentation. Then, the candidate regions of small size are removed. Second, the Bag of Visual Words (BoVW) histogram [19] and Histogram of Oriented Gradients (HOG) features [20] have been extracted from the saved candidate regions, respectively. Finally, the images of the candidate regions are classified by using Support Vector Machine (SVM) [21].

### A. CASCADED COLOR SEGMENTATION

In street scenes, many text-based traffic signs are usually designed with blue backgrounds and white text and/or signs. Therefore, an cascaded color segmentation method is proposed to detect the traffic signs with blue backgrounds in the street scenes.

Kulkarni [22] proposed the color segmentation $M_1$, which is computed as follows:

$$M_1(x, y) = \begin{cases} 255 & if\ R(x, y) \leq T_0 \\ 0 & else \end{cases} \tag{1}$$

where $R(x, y)$, which is in the range 0 to 255, is the pixel of $x^{th}$ row and $y^{th}$ column in the red component, R, of the input image. In [22], $T_0 = 90$ is the suitable threshold. This method is simply denoted as $M_1$. This method not only preserves the majority of the blue regions in the image, but also removes light blue regions, such as the sky and light blue glass windows.

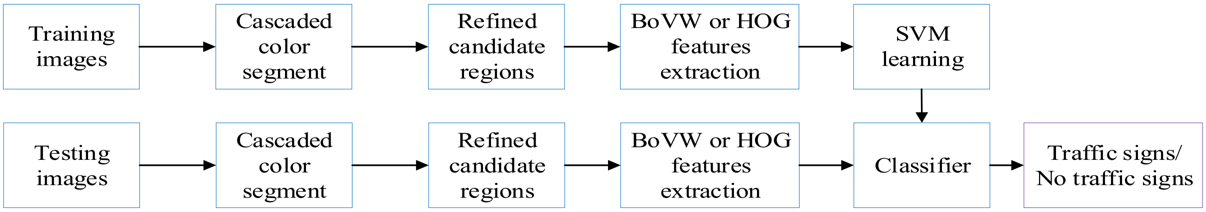



**FIGURE 5.** Flowchart of text-based traffic sign detection.

Gómez-Moreno *et al.* [23] proposed the color segmentation $M_2$, which is computed as follows:

$$M_2(x, y) = \begin{cases} 255 & if\ H(x, y) \geq T_1\ and\ H(x, y) \leq T_2 \\ 0 & else \end{cases} \quad (2)$$

where $H(x, y)$, which is in the range of 0 to 180, is the pixel of $x^{th}$ row and $y^{th}$ column in the hue component of the input image. In [23], $T_1 = 100$ and $T_2 = 140$ are the suitable thresholds. This method is simply denoted as $M_2$. The blue regions are detected by this method, which effectively separates blue from other colors.

As shown in Figure 6(a) and Figure 6(b), in the street scene, some dark regions and white regions are not removed by $M_1$ and $M_2$. Therefore, we propose $M_3$, which is computed as follows:

$$M_3(x, y) = Otsu(|R(x, y) - B(x, y)|)\ AND\ Otsu(|G(x, y) - B(x, y)|) \quad (3)$$

where $B(x, y)$ is the pixel of $x^{th}$ row and $y^{th}$ column in the blue component, B, of the input image. $G(x, y)$ is the pixel of $x^{th}$ row and $y^{th}$ column in the green component, G, of the input image. $Otsu(*)$ is the result using Otsu's segmentation method [24] on the image. This method is simply denoted as $M_3$. In an image, the values of the red, green and blue color components are denoted as $R$, $G$, and $B$, respectively. In the dark green regions, the $R$, $G$, and $B$ are all quite small. The $G$ is slightly larger than that of both the $R$ and the $B$, which are almost the same. Therefore, for the dark green objects, the result of subtracting the $B$ from the $R$ is small; whereas, for the blue objects, the result is relatively large. $Blue_{pre}(x, y)$, which is a combination of the $M_1$, $M_2$ and $M_3$ using a logical "AND" operation, is computed as follows:

$$Blue_{pre}(x, y) = M_1(x, y)\ AND\ M_2(x, y)\ AND\ M_3(x, y). \quad (4)$$

As shown in Figure 6(c), after applying the previous method, the approximate locations of the candidate regions are obtained. However, some small regions having colors similar to the text-based traffic signs are also retained. In addition, a few dark colored and white colored regions are retained. To solve this above problem, we propose $M_4$, which is computed as follows:

$$M_4(x, y) = Abs(2 * B(x, y) - R(x, y) - G(x, y)) \quad (5)$$

where $Abs(*)$ represents the operation of absolute value. This method is simply denoted as $M_4$. For dark colored objects, the value of the result of the $M_4$ is small. Similarly, for some white colored regions, the value of the result of the $M_4$ is also small. Then, the $Blue_{mask}$ is computed as follows:

$$Blue_{mask} = Otsu(Blue_{pre}(x, y)\ AND\ M_4(x, y)) \quad (6)$$

where $Blue_{mask}$ is the final result of the proposed cascaded color segmentation method. As shown in Figure 6(c) and Figure 6(e), $M_4$, which further improves the performance of rejecting dark colored objects and the very small regions, is an important supplement to the previous method.

In HSV color space, the blue regions can be detected by (2), which effectively separates blue from other colors. However, there are some light and dark blue objects that are mistaken for traffic signs. The saturation refers to the purity of color. Generally, the saturation of blue color in traffic sign is higher than that in sky or shade area. Therefore, the saturation can be used to distinguish traffic signs from those objects with light and dark color. However, under the lighting condition of various traffic scenes, the saturation of blue traffic signs changes dramatically. Therefore, setting an experience threshold cannot segment the blue traffic signs very well. The Otsu's method is used to adaptively obtain the threshold of saturation component. Therefore, in the HSV color space, the $HSV_{mask}$ is computed as follows:

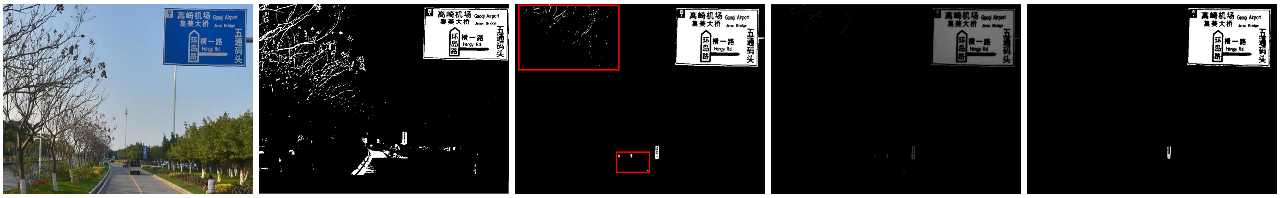
$$HSV_{mask} = Otsu(S(x, y))\ AND\ M_2(x, y) \quad (7)$$

where $HSV_{mask}$ is the result of the proposed color segmentation method. $S(x, y)$ is the pixel of $x^{th}$ row and $y^{th}$ column in the saturation component, S, of the input image. The detail results are shown in Table 2 and Table 3 in the section IV.

In this step, all the contours are extracted in the binary image. The outer contours are obtained to find the minimum bounding boxes. However, as shown in Figure 7(a), these candidate regions of text-based traffic signs are non-connected regions, which means a candidate region is detected as two or more candidate regions. Therefore, as shown in Figure 7(b), these bounding boxes are merged into a new rectangular box as a candidate region. The two connected bounding boxes are merged into a new one, $Rect$, by using:

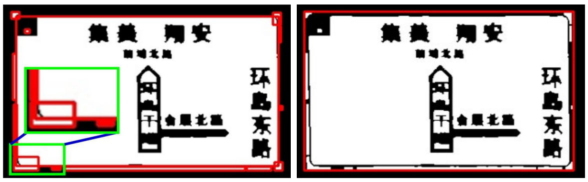
$$Rect = Rt_1 \cup Rt_2\ if\ IoU_r(Rt_1, Rt_2) \geq T_i \quad (8)$$

where $Rt_1$ and $Rt_2$ are the two bounding boxes of candidate regions, respectively. The $IoU_r(*)$ operation stands for the area of intersection over union between two boxes. $T_i$ is set at 0.25.



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

**FIGURE 6.** **Procedures of the proposed cascaded color segmentation method for text-based traffic sign detection. (a) Input image. (b) Result of $M_1$ AND $M_2$. (c) Result of $M_1$ AND $M_2$ AND $M_3$. (d) Result of (c) AND $M_4$ (e) Result of applying Otsu's method on (d).**



(a) (b)

**FIGURE 7.** **Bounding boxes merged algorithm. (a) Many candidate bounding boxes. (b) Merged bounding box.**

### *B. FEATURE EXTRACTION AND CLASSIFICATION*

Actually, after applying the cascaded color segmentation method, there still remain some very small blue regions, which can be classified into the following two categories: 1) the very small text-based traffic signs; 2) the non-traffic sign regions, such as billboards, etc. Although the very small text-based traffic signs contain route information, because of the small scale, it is almost impossible to obtain the useful information. Therefore, a part of the candidate regions is first refined by restricting the size and aspect ratio of the regions as follows:

$$Region_{area} \geq T_{area} \tag{9}$$

$$Region_{aspect} \geq T_{aspect} \tag{10}$$

where $Region_{area}$ and $Region_{aspect}$ represents the size of the bounding box and the aspect ratio of the bounding box, respectively. $T_{area} = 225$ and $T_{aspect} = 7$ are suitable values of the thresholds for filtering the small regions and some non-traffic sign regions.

To improve classification accuracy, good features, which can handle intensity, rotation, scale, and affine transformations, must be introduced into the classification. Therefore, the BoVW histogram and HOG features were chosen to express the candidate regions.

On one hand, the HOG features of the candidate regions are extracted from images. First, the image is segmented into multiple sub-regions. Then, by calculating the gradient direction histogram of these sub-regions, the HOG features are obtained. On the other hand, using the BoVW technique, the features are also extracted. Firstly, Scale-Invariant Feature Transform (SIFT) [25] features are extracted from the images and calculated to generate 128-dimension vector descriptors. Then, using k-means clustering, the sampled features are clustered to quantize the space into a discrete number of visual words. The number of clustering centers is defined as $k = 500$. After performing the clustering analysis based on k-means, each cluster center, equivalent to a visual word, is considered as the representative of several similar local regions. The image is represented by the histogram of the visual words, which counts how often each of the visual words occurs in the image.

The candidate regions are obtained from the dataset which is collected from the street view. These candidate regions can be classified into two classes: the traffic sign and non-traffic sign. Therefore, we label the text-based traffic sign to 0 and the non-traffic sign to 1. The linear SVM classifier is used to classify the two categories of the candidate regions. Two kinds of features are extracted from the training dataset by using both the BoVW technique and the HOG algorithm, respectively. Firstly, the two features, which are preprocessed as input data, are fed into the linear SVM classifiers, respectively. Secondly, a proper SVM classifier is trained. When a candidate region obtained from a testing image is judged to be a non-traffic sign in a linear SVM classifier, the candidate region is eliminated immediately. Finally, the retained candidate regions are judged to be traffic signs regions.

## III. CHINESE CHARACTERS RECOGNITION ON TEXT-BASED TRAFFIC SIGNS

In this section, a component-based text line forming method is proposed to detect the text line in the text-based traffic sign. Then, using an efficient Chinese character recognition method, Chinese characters in text line are recognized.

### *A. COMPONENT-BASED TEXT LINE FORMING METHOD*

Locating each Chinese character directly from text-based traffic sign is a problem because a Chinese character, having unconnected strokes, is detected as two or more characters. Actually, the Chinese characters on text-based traffic signs are always in the form of lines of text. Therefore, to obtain text lines in text-based traffic signs, a component-based text line forming method is proposed. The proposed method combines top-down and bottom-up strategies and uses information about the location, size, and aspect ratio of the candidate regions to form text lines, as shown in Figure 8. Compared with the symbols and words in the natural scene, the symbols and words on such traffic signs have more regularity.

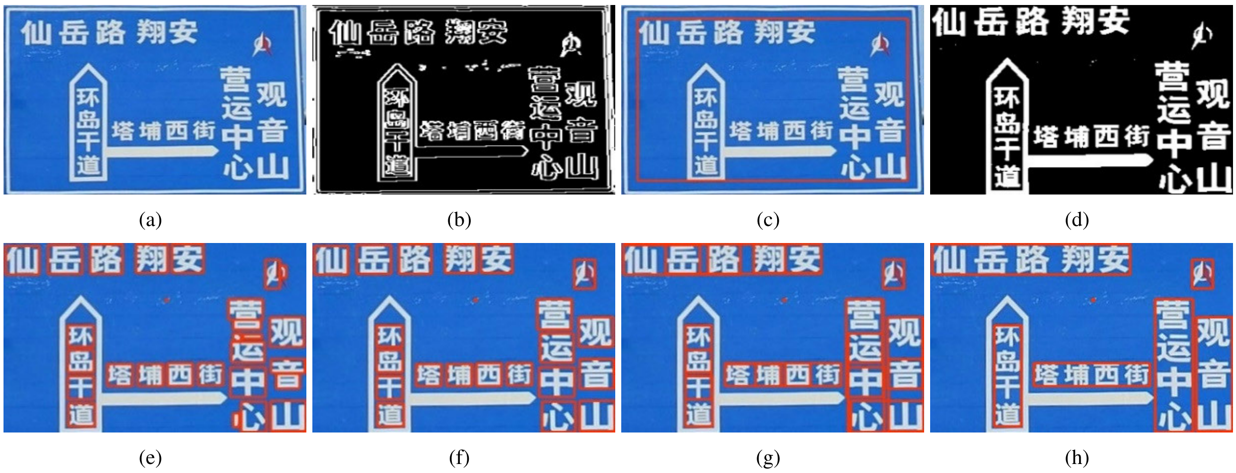



**FIGURE 8. Procedure of component-based text line forming algorithm. (a) Input image. (b) Result of applying enhancement algorithm on (a). (c) Shrunk bounding box. (d) Result of applying Otsu's threshold method on (c). (e) Detected rough characters. (f) Detected characters. (g) Detected text regions. (h) Formed text line.**

Therefore, to make the method simple and effective, we use the scaling method to make the detected traffic signs to an appropriate size.

To improve the performance of detecting text lines in a text-based traffic sign, it is necessary to shrink the border of the whole traffic sign to the smallest size, which exactly includes all the characters. To locate the text block in a text-based traffic sign, the edge enhancement algorithm is used according to the following:

$$O(x, y) = K(Canny(Gray(I(x, y)))) \tag{11}$$

where $I(x, y)$ is the text-based traffic sign. $Gray(*)$ is the operation for converting from an input color image to a gray-scale image. $Canny(*)$ is a method of gray-scale image edge detection, where, if the pixel is regarded as edge, the result is 255; otherwise, the result is zero. $K(*)$ indicates the morphological operation of expansion to enhance the edge of the character candidate regions. Therefore, as shown in Figure 8(b), processed by the edge enhancement algorithm, all the contours are extracted from the binary image. By analyzing the outer contours, the minimum bounding boxes are obtained as the candidate regions. However, the edge of the whole traffic sign often forms a bounding box, but the bounding box is not the real text area. Furthermore, the bounding box also interferes with the character segmentation. Therefore, we want to shrink the border of the whole traffic sign to the smallest size, which exactly includes all the characters. The maximum number of holes that a Chinese character can have is 4, as in the case of "田". The bounding box which contains more than four smaller bounding boxes will be removed. Then, merge all the remaining bounding boxes to a largest one, which is a shrunk bounding box of the whole traffic sign.

As shown in Figure 8(c), the shrunk bounding box of the whole traffic sign is denoted in red color. Actually, the shrunk bounding box has relatively simple backgrounds. Therefore, it is suitable to apply Otsu's segmentation method to obtain the binary image of the character regions, as shown in Figure 8(d). Then, by finding the minimum bounding boxes of each candidate region, the rough characters are obtained. However, as shown in Figure 8(e), we can see that some Chinese characters, such as "音", "营", etc., are easily segmented into several regions, because the strokes of these characters are unconnected. To solve this problem, taking into consideration the appearance of the Chinese characters, the aspect ratio of most Chinese characters is similar to that of a square. Therefore, the aspect ratio is used as the prior features. In this step, the adjacent candidate regions are obtained by searching for the adjacent areas. Then, as shown in Figure 9, two parts of a Chinese character are merged into a new character according to the following:

$$C = A \cup B \tag{12}$$

where $A$ and $B$ are the two bounding boxes of two parts of a Chinese character region, respectively. The "∪" operation stands for the union between two boxes. $C$ is the result of merging $A$ and $B$. In order to better define the relationship between two parts of a Chinese character region, $Ratio_R$ is computed as follows:

$$Ratio_R = max(W/H, H/W) \tag{13}$$

where $W$ and $H$ are the width and height of a candidate region, $R$, respectively. $Ratio_R$ is the maximum between the result of dividing $W$ by $H$ and the result of dividing $H$ by $W$. The merged character regions must satisfy all the following constraints:

$$Ratio_C < Ratio_A \text{ and } Ratio_C < Ratio_B \tag{14}$$

$$Ratio_C - Ratio_A > T_r \text{ and } Ratio_C - Ratio_B > T_r \tag{15}$$

$(W_1, H_1)$ and $(W_2, H_2)$ are the width and height of the two parts of Chinese character regions ($A$ and $B$), respectively. $Ratio_A$, $Ratio_B$, and $Ratio_C$ have the same definitions as in (13). $T_r$ is set at 0.5.

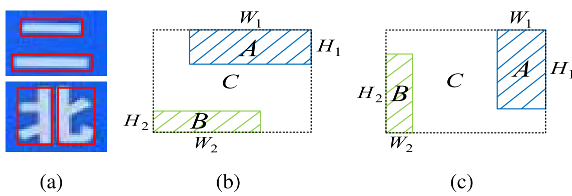



**FIGURE 9.** Illustration of merging two adjacent parts belonging to a Chinese character. (a) Two examples of detected two adjacent parts of a Chinese character in the real scene. (b) Condition to merge two adjacent parts in horizontal direction. (c) Condition to merge two adjacent parts in vertical direction.

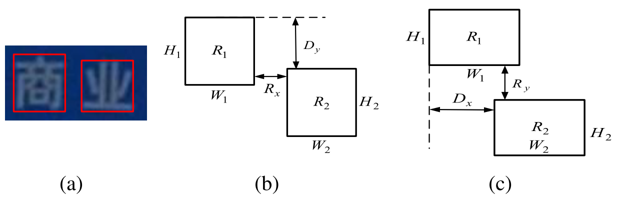



**FIGURE 10.** Two adjacent Chinese characters merged into a text region. (a) An example of detected two adjacent characters in the real scene. (b) Condition to merge two characters into a text region in horizontal direction. (c) Condition to merge two characters into a text region in vertical direction.

As shown in Figure 8(f), all of the Chinese characters with unconnected strokes are detected correctly. On text-based traffic signs in general, rather than appearing isolated, characters are in alignment. Therefore, a better choice is to detect text lines on text-based traffic signs, because complete information can be recognized easily from the text lines. By analyzing the relationships between the adjacent characters, the adjacent characters are grouped into text lines. However, unlike English text lines, which are only horizontal, Chinese text lines on text-based traffic signs are usually both in horizontal and vertical directions. Therefore, it is important to know which Chinese characters on text-based traffic signs are horizontal and which ones are vertical. To solve this problem, as shown in Figure 10, the adjacent characters must be grouped into horizontal or vertical text regions by analyzing the positions of these characters. The vertical and horizontal deviation, $D_y$ and $D_x$, between two adjacent candidate regions, $R_1$ and $R_2$, are computed as follows:

$$D_y = |y_1 - y_2| \tag{16}$$
$$D_x = |x_1 - x_2| \tag{17}$$

where $y_1$ and $y_2$ are the ordinates of the top left corner vertex of $R_1$ and $R_2$, respectively. $x_1$ and $x_2$ are the abscissae of the top left corner vertex of $R_1$ and $R_2$, respectively. To determine whether two characters are horizontally adjacent, the conditions to be met at the same time are expressed as follows:

$$R_x < Th_1 * Min_w \tag{18}$$
$$D_y < Min_h/Th_2 \tag{19}$$
$$D_h < Min_h/Th_3 \tag{20}$$

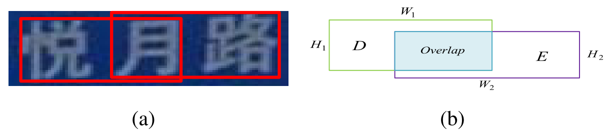



**FIGURE 11.** Two overlapping text regions merged into a text line. (a) An example of detected two text regions in the real scene. (b) Condition to merge two overlapping text regions into a text line.

where $R_x$ refers to the horizontal distance between the two adjacent candidate regions. The $Min_w$ and $Min_h$ are the smaller width and height of the two adjacent regions, respectively. $D_h$ is the height difference between two adjacent regions. $Th_1$, $Th_2$, and $Th_3$ are three thresholds set at 2.0, 7.0, and 6.5, respectively. To determine whether two characters are vertically adjacent, the conditions to be met at the same time are expressed as follows:

$$R_y < Th_1 * Min_h \tag{21}$$
$$D_x < Min_w/Th_2 \tag{22}$$
$$D_w < Min_w/Th_3 \tag{23}$$

where $R_y$ refers to the vertical distance between the two adjacent candidate regions. $D_w$ is the width difference between two adjacent regions. Then, merge the two adjacent character regions to a text region horizontally or vertically. As shown in Figure 8(g), each new text region consists of only two character regions. In addition, these text regions overlap with their adjacent ones. Therefore, a text region merging algorithm is proposed to form a text line. To obtain the appropriate text lines containing complete text information, the two overlapping text regions are merged into a new text line as follows:

$$F = D \cup E \tag{24}$$

where $D$ and $E$ are the bounding boxes of the two text regions, respectively. $F$ is the result of merging $D$ and $E$. The bounding boxes of the two text regions must satisfy the following constraints:

$$Overlap > 0 \tag{25}$$
$$Ratio_F > Ratio_D \tag{26}$$
$$Ratio_F > Ratio_E \tag{27}$$

where $Overlap$ is defined as the area of the overlap between the two overlapping text regions, as shown in Figure 11. $Ratio_D$, $Ratio_E$, and $Ratio_F$ have the same definitions as in (13).

As shown in Figure 8(g), some characters have both horizontal and vertical adjacent characters, which results in the overlapping text lines that must be judged whether they are vertical or horizontal. Actually, on text-based traffic signs, most vertical text lines will contain three or more Chinese characters. Therefore, this feature is used to determine whether or not the overlapping regions are vertical. If the vertical text lines contain three or more Chinese characters,

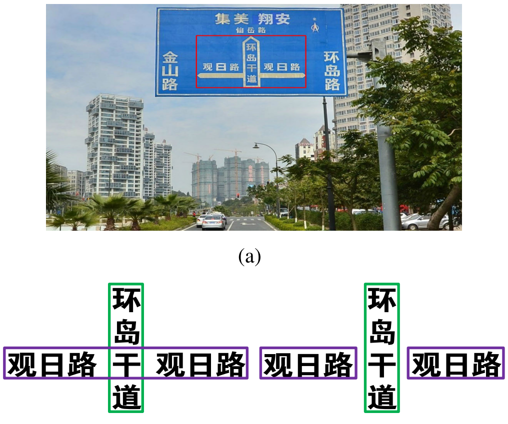



**FIGURE 12.** Either vertical or horizontal text line. (a) Real image including both vertical and horizontal text line in street scene. (b) Detected text line in red box of (a). (c) Correct classification of crossed part into vertical text line in (b).

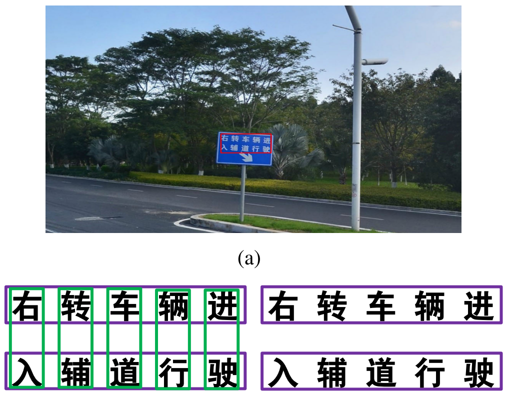



**FIGURE 13.** Either vertical or horizontal text line. (a) Real image including both vertical and horizontal text line in street scene. (b) Detected text line in red box of (a). (c) Correct classification of crossed part into horizontal text line in (b).

as shown in Figure 12, the overlapping regions will be vertical. Conversely, as shown in Figure 13, the overlap will be horizontal if the vertical text lines contain less than three Chinese characters. As shown in Figure 8(h), all the characters are grouped into text lines.

### B. CHINESE CHARACTER RECOGNITION METHOD

The Chinese characters on text-based traffic sign are printed with a fixed font, which slightly reduces the difficulty to recognize the characters. However, the scale of Chinese characters changes with the scale of different traffic signs. Therefore, two different types of features that are insensitive to the scale, the template and HOG features, are used to recognize the Chinese characters.

On one hand, to obtain the template feature, the character region image is re-scaled to be $16 \times 16$. Then, the colored image is converted into a gray image. A binary image, which has been obtained by Otsu's threshold method, is transformed into the feature vector of a histogram. The template feature of a 256-dimension vector, as shown in Figure 14, is obtained according to the following:

$$Hist(x, y) = ToHist(Otsu(Gray(I_c(x, y)))) \tag{28}$$

where $I_c(x, y)$ is a Chinese character. $Gray(*)$ is the operation for converting from an input color image to a gray-scale image. $ToHist(*)$ is the operation for converting the matrix to a histogram. On the other hand, to obtain the HOG feature, the character region image is re-scaled to be $64 \times 64$, and then the $64 \times 64$ block is divided into sixteen $16 \times 16$ sub-blocks, as shown in Figure 14. Then, the descriptor of the HOG feature is computed for each sub-block. Therefore, a HOG feature of a 1,764-dimension vector is obtained.

To recognize Chinese characters on text-based traffic signs, a dataset, consisting of 7,000 different Chinese character samples, was collected. Before recognizing Chinese characters on text-based traffic signs, the template features and the HOG features of the dataset are obtained. The template and the HOG features are cascaded to coarsely and finely recognize Chinese characters on text-based traffic signs. Firstly, the Euclidean distance between the template features of a character on text-based traffic signs and each Chinese character in the dataset is computed. As a result, the first fifteen most likely Chinese characters are obtained. Second, the most appropriate result is obtained by calculating the similarity measurement of the HOG features between the Chinese character to be detected and each Chinese character in the dataset. Actually, some non-Chinese character candidate regions are also retained, resulting in that these candidate regions are incorrectly recognized as Chinese characters. To solve this problem, the similarity measurement of the HOG features

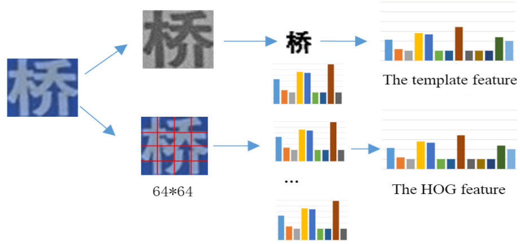



**FIGURE 14.** Feature extraction of Chinese characters.

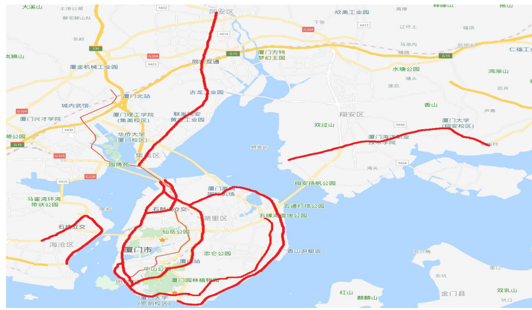
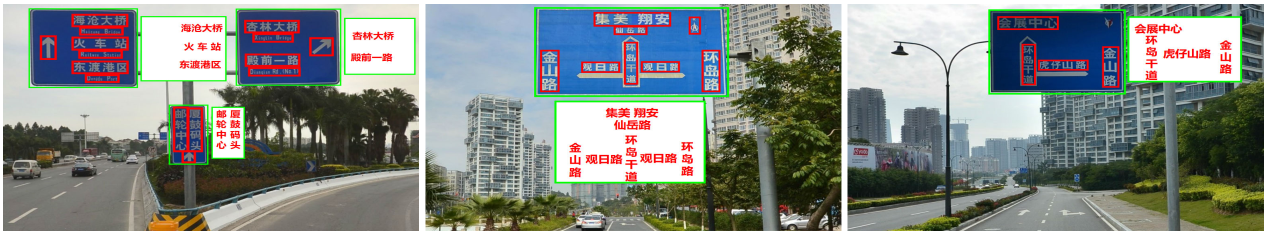
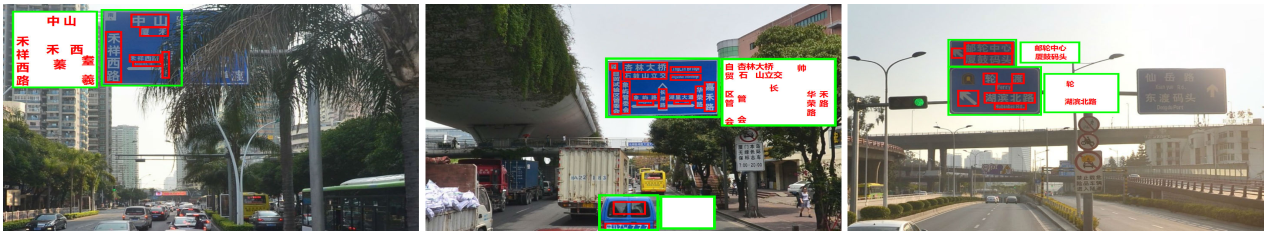


**FIGURE 15.** Roads from which the images have been obtained.




(a) (b) (c)

**FIGURE 16.** Successful examples of text-based traffic sign detection and recognition in street scene.



(a) (b) (c)

**FIGURE 17.** Failed examples of text-based traffic sign detection and recognition in street scene.

(a) (b)

**FIGURE 18.** Text-based traffic signs detection in street scene images. (a) Result of González's method [17]. (b) Result of the proposed method.

between a Chinese character on text-based traffic signs and each Chinese character in the dataset is compared. Experimental results show that $T_c = 0.5$ is the suitable threshold for removing the non-Chinese characters.

## IV. EXPERIMENT

### A. DATASET

To evaluate the performance of the detection and recognition of text-based traffic signs in the street scene, a dataset was collected from the Street View Service developed by Tencent Cooperation. A total of $1,025$ images with size of $1,080 \times 720$ were collected. All training and testing images were collected from the image set from the roads shown in red in Figure 15. All the collected images are annotated with text-based traffic signs. Among the $1,025$ images, 528 images were used for training; the others were used for testing.

These images were labeled with the bounding boxes of the text-based traffic sign region and the character region. The size of the smallest and largest labeled text-based traffic sign is $15 \times 15$ and $540 \times 360$, respectively. The size of the smallest and largest labeled character is $4 \times 4$ and $270 \times 180$, respectively. In order to evaluate the performance of the proposed method at different distances, we also labeled each of text-based traffic signs with different distance. For the middle distance, the size of the smallest and largest labeled text-based traffic sign is $30 \times 30$ and $100 \times 100$, respectively. The text-based traffic sign with the size of smaller than $30 \times 30$ and larger than $100 \times 100$ is regarded as short and long distance, respectively.

To recognize Chinese characters in text-based traffic signs, 2,696 bounding boxes of Chinese character regions have been labeled to evaluate the performance of Chinese characters recognition.

### B. EVALUATION OF TRAFFIC SIGNS DETECTION

In this paper, *Recall*, *Precision*, and $F_1$-*measure* are the important criteria for the evaluation of traffic signs detection. Therefore, *Recall*, *Precision*, and $F_1$-*measure* are calculated, respectively, as follows:

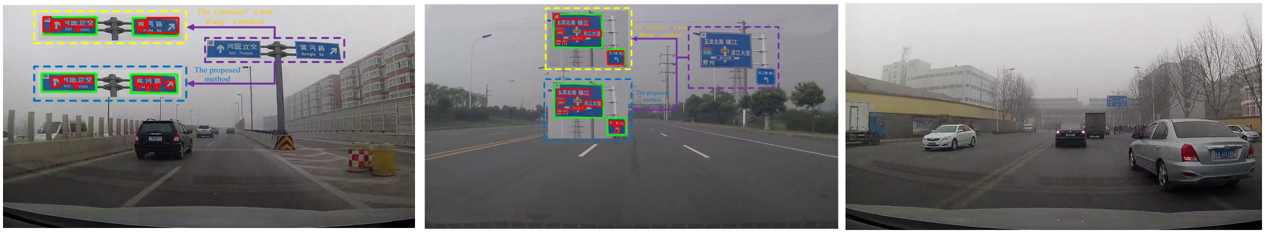
$$Recall = \frac{TP}{TP + FN} \tag{29}$$

$$Precision = \frac{TP}{TP + FP} \tag{30}$$

$$F_1 = 2 * \frac{Recall * Precision}{Recall + Precision} \tag{31}$$

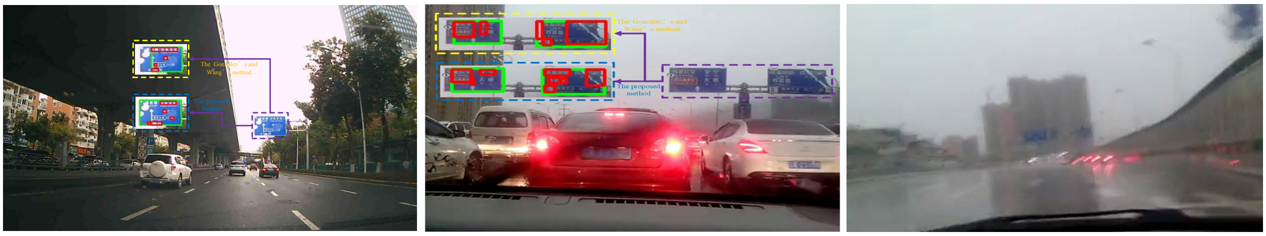
where $TP$, $FN$, $TN$, and $FP$ are the number of true positives, false negatives, true negatives and false positives, respectively. To evaluate whether or not a detected box is a hit, the following expression is adopted:

$$IoA = \frac{Intersection(G, D)}{Area(G)} \tag{32}$$



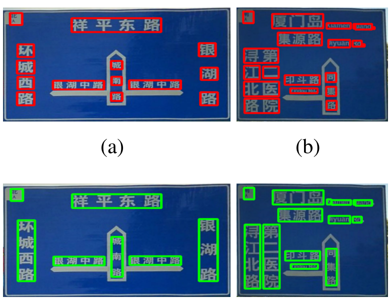
(a) (b) (c)

**FIGURE 19.** Results of traffic signs and text line detection on the foggy day. Yellow dotted box is the results of González's [17] and Wang's [5] method. González's method is used for the traffic sign detection. Wang's method is used for the text line detection on the traffic signs. Blue dotted box is the results of the proposed method.



(a) (b) (c)

**FIGURE 20.** Results of traffic signs and text line detection on the rainy day. Yellow dotted box is the results of González's [17] and Wang's [5] method. González's method is used for the traffic sign detection. Wang's method is used for the text line detection on the traffic signs. Blue dotted box is the results of the proposed method.




(a) (b)

(c) (d)

**FIGURE 21.** Comparison between Wang's method [5] and the proposed method. (a) and (b) are results of Wang's method. (c) and (d) are results of the proposed method.

**TABLE 2.** Performance comparison between the proposed method and González's method with feature extractor of BoVW under different distances.

| Distance | Method | Precision | Recall | $F_1$-measure |
|---|---|---|---|---|
| Short | Proposed | 0.773 | **0.865** | **0.816** |
| | Proposed-hsv | 0.406 | 0.765 | 0.531 |
| | González's [17] | **0.789** | 0.833 | 0.811 |
| Midium | Proposed | **0.769** | **0.759** | **0.764** |
| | Proposed-hsv | 0.295 | 0.616 | 0.399 |
| | González's [17] | 0.672 | 0.739 | 0.704 |
| Long | Proposed | **0.428** | 0.569 | **0.488** |
| | Proposed-hsv | 0.197 | 0.382 | 0.260 |
| | González's [17] | 0.356 | **0.715** | 0.476 |
| All | Proposed | **0.697** | 0.767 | **0.731** |
| | Proposed-hsv | 0.301 | 0.591 | 0.398 |
| | González's [17] | 0.619 | **0.782** | 0.691 |

where $G$ and $D$ are the ground-truth box and detected box, respectively. $Intersection(*)$ stands for the intersection between two boxes. $Area(*)$ is used to calculate the area. If $IoA$ is greater than 0.5, the detected box, $D$, is considered a hit.

Some of correctly detected text-based traffic signs are shown in Figure 16(a), Figure 16(b) and Figure 16(c). The proposed method detects all the text-based traffic signs with blue background in the street scene. The cascaded color segmentation method, which makes full use of color information, improves the performance of detection of blue objects. In addition, small blue objects, which usually do not contain any useful information because of the small scale, are regarded as negative samples. However, as shown in Figure 17(a), the method fails to detect the whole traffic sign because of partial occlusion by leaves. Even though the text-based traffic sign is partially occluded, a part of it can still be detected. Because of complex and messy backgrounds, the blue non-traffic sign candidate regions are mistaken for traffic signs shown in Figure 17(b). Additionally, varied illumination causes drastic changes in the color of text-based traffic signs. Therefore, as seen in Figure 17(c), varied illumination also leads to failed detection of text-based traffic signs.

For the street scene images of the dataset that we collected, the experimental results using the proposed method of text-based traffic sign detection and using González's method [17] are compared in Table 2 and Table 3. From Table 2 and Table 3, it is seen that, compared with González's method,

**TABLE 3. Performance comparison between the proposed method and González's method with feature extractor of HOG under different distances.**

| Distance | Method | Precision | Recall | $F_1$-measure |
|---|---|---|---|---|
| Short | Proposed | 0.813 | **0.844** | 0.828 |
| | Proposed-hsv | 0.631 | 0.706 | 0.666 |
| | González's [17] | **0.818** | 0.843 | **0.830** |
| Midium | Proposed | **0.805** | 0.765 | **0.785** |
| | Proposed-hsv | 0.490 | 0.590 | 0.535 |
| | González's [17] | 0.692 | **0.782** | 0.734 |
| Long | Proposed | **0.523** | 0.777 | **0.625** |
| | Proposed-hsv | 0.482 | 0.505 | 0.493 |
| | González's [17] | 0.417 | **0.792** | 0.546 |
| All | Proposed | **0.739** | 0.799 | **0.767** |
| | Proposed-hsv | 0.514 | 0.601 | 0.554 |
| | González's [17] | 0.658 | **0.808** | 0.725 |

the proposed method shows great improvement in precision and a slight reduction in recall. The proposed-hsv denotes the proposed color segmentation method which only uses the HSV color space. The main reason is that an effective combination of different red, green and blue components is used to distinguish traffic signs from many objects of similar color in a very complex street scene. The small blue objects, which are regarded as negative samples, are removed to improve the performance of text-based traffic sign detection, as shown in Figure 18. In addition, as shown in Table 2 and Table 3, the results of using HOG features to detect text-based traffic signs are better than the results using BoVW features. The experimental results show that the traffic signs with a larger area acquire more feature descriptors to improve the accuracy of the visual words histogram. Accordingly, the SVM classifier has better performance. Conversely, the traffic signs with the smaller and blurring regions cannot obtain adequate descriptors for calculation. However, the number of HOG features is fixed. Thus, using HOG features, better results are obtained in the smaller and blurring regions. As shown in Table 2 and Table 3, the best detection result is achieved when the text-based traffic sign is at the short distance. The reason is that the close-range text-based traffic signs have larger size and more information than those at the long distance. Accordingly, the SVM classifier has the better performance. Conversely, a long-distance text-based traffic sign with smaller and blurring regions is difficult to extract abundant information. Therefore, at the stage of recognition, the text-based traffic signs at the short and middle distances are recognized.

The proposed method is also compared with Rong's method [3], as shown in Table 4. Rong *et al.* propose a localization network with You Only Look Once (YOLO) detector to find all the text-based traffic signs. From the Table 4, the proposed method achieves a better result than Rong's method on the same dataset which we collected. The Rong's method achieves a good result on precision. However, the *Recall* of Rong's method is much lower than the proposed method. The main reason is that the Rong's method with YOLO detector is difficult to detect the small text-based traffic signs at the middle or long distance.

**TABLE 4. Performance comparison of text-based traffic signs detection between the proposed method and Rong's method.**

| Method | Precision | Recall | $F_1$-measure |
|---|---|---|---|
| Proposed | 0.739 | **0.799** | **0.767** |
| Rong's [3] | **0.870** | 0.527 | 0.656 |

**TABLE 5. Performance comparison of Chinese text line detection among three methods.**

| Method | Precision | Recall | $F_1$-measure |
|---|---|---|---|
| Proposed | **0.942** | **0.969** | **0.955** |
| Wang's [5] | 0.898 | 0.932 | 0.915 |
| Rong's [3] | 0.755 | 0.604 | 0.707 |

### C. EVALUATION OF TEXT DETECTION AND RECOGNITION IN TRAFFIC SIGNS

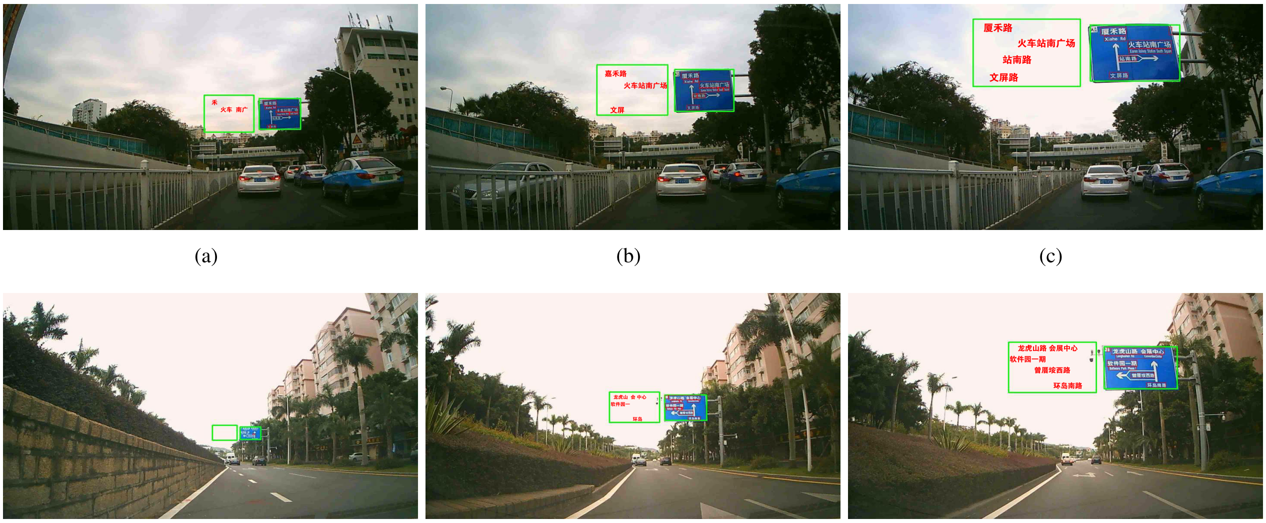
To evaluate the detection of text lines, Intersection over Union ($IoU$) is used to describe whether the detected bounding boxes are true or false. $IoU$ is defined as follows:

$$IoU = \frac{Intersection(G, D)}{Union(G, D)} \tag{33}$$

where $Union(*)$ stands for the union area between two boxes. The threshold of $IoU$ is set at 0.5. From Table 5, it is seen that $Precision$, $Recall$, and $F_1$-$measure$ for text line detection in text-based traffic signs can attain 94.18%, 96.89%, and 95.52%, respectively. From Table 5, the proposed method achieves a better result than Rong's method [3] and Wang's method [5]. As shown in Figure 21, Wang's method detects only the horizontal text lines. Rong *et al.* use the YOLO detector to detect the text lines on text-based traffic signs. However, the orientation of the Chinese text lines can be horizontal or vertical on text-based traffic signs, which makes the YOLO detector difficult to detect text lines exactly. The component-based text line forming method uses the position information to detect both horizontal and vertical text lines. Therefore, the proposed method achieves the better performance of the Chinese text line detection.

Several successful examples of text line detection on text-based traffic signs are showed in Figure 16(a), Figure 16(b), and Figure 16(c). The proposed method successfully detects almost all the text lines on the text-based traffic signs. However, as seen from Figure 17(a), a few text line detections failed due to partial occlusion. As shown in Figure 17(b), because of the blurred text-based traffic signs, some text line detections failed. Besides, as shown in Figure 17(c), varied illumination also causes the failed text line detection. This is because that the contrast ratio between the characters and the background is low, making it difficult to obtain the characters to form the text line. In addition, the various sizes of characters are also the important factor for text lines detection.

As shown in Figure 19(a), Figure 19(b), Figure 20(a), and Figure 20(b), we can see that the proposed method can detect most text-based traffic signs on some challenging weather conditions. The main reason is that the proposed method uses the stable color information to detect text-based traffic signs on some challenging weather conditions. However, it is



(a) (b) (c)

(d) (e) (f)

**FIGURE 22.** Results of the proposed method on the traffic scene video. (a), (b) and (c) are the 1,090$^{th}$, 1,150$^{th}$ and 1,220$^{th}$ frames of one in-vehicle video, respectively. (d), (e) and (f) are the 80$^{th}$, 170$^{th}$ and 210$^{th}$ frames of another in-vehicle video, respectively.

**TABLE 6.** Recognition Performances of Chinese Characters in Traffic Signs under different distances.

| Distance | CTP | TP | Recognition rate |
|---|---|---|---|
| Short | 962 | 1176 | 81.8% |
| Midium | 1071 | 1520 | 70.5% |
| All | 2033 | 2696 | 75.4% |

still difficult to detect traffic signs on the extreme weather, as shown in Figure 19(c) and Figure 20(c). It is difficult for people to see the words on the text-based traffic signs in such extreme weather. In addition, it is also difficult to detect text lines on the text-based traffic signs because of the blurring traffic signs. Compared with González's [17] and Wang's [5] method, we find that our method has slightly improvement on the traffic signs and text line detection in foggy and rainy weather. However, in extremely heavy rain and fog, all methods are difficult to complete the detection task.

A Chinese character recognition method is proposed to recognize the Chinese characters on text-based traffic signs. The rate of Chinese character recognition is defined as follows:

$$Accuracy = \frac{CTP}{TP} \tag{34}$$

where $CTP$ is the number of correctly recognized true positives. The character candidate regions are recognized by using the proposed efficient Chinese character recognition algorithm.

Shown in Figure 16(a), Figure 16(b), and Figure 16(c) are several results of recognizing the Chinese characters on text-based traffic signs. As seen, the proposed method achieves good results in recognizing the Chinese characters on text-based traffic signs. The Chinese characters on text-based traffic signs are printed with a fixed font, which slightly reduces the difficulty in recognizing them. Therefore, the Chinese character recognition algorithm, which uses the stable features to recognize the Chinese characters, can effectively recognize the Chinese characters on text-based traffic signs. However, as seen in Figure 17(a), due to partial occlusion, some Chinese characters are incorrectly recognized. Only a part of the Chinese characters on the detected text-based traffic sign can be recognized. As seen in Figure 17(b), some Chinese characters are incorrectly recognized on the blurred traffic sign. Varied illumination causes the failed text-based traffic signs detection. As a result, as shown in Figure 17(c), the characters on these text-based traffic signs cannot be detected and recognized.

The results of Chinese characters recognition are shown in Table 6. For the short and middle distance, the recognition rates of Chinese characters are 81.8% and 70.5%, respectively. For all the distance, the final recognition rate of Chinese characters is 75.41%. From the results, the proposed Chinese character recognition method can recognize most Chinese characters in the traffic signs.

### D. EVALUATION OF THE PROPOSED METHOD ON THE IN-VEHICLE VIDEO SYSTEM

The goal of the proposed method is to be applied to the in-vehicle video system. To evaluate the performance of the proposed method on the intelligent vehicle video system, a series of natural traffic scene videos are collected. As shown in Figure 22(c) and Figure 22(f), it is seen that the system can detect text-based traffic signs and text lines. Almost all the Chinese text lines at short distance are recognized. As shown in Figure 22(b) and Figure 22(e), the system can also well detect text-based traffic signs and text lines at medium distance. However, some text lines may not be recognized because of the blurred traffic signs. The system can detect most of the traffic signs at long distance. However, as shown in Figure 22(d), the text lines of some traffic signs may not be detected because of the poor quality of the in-vehicle video. As shown in Figure 22(a), only some simple Chinese characters can be recognized correctly.

## V. CONCLUSION AND FUTURE WORK

An effective text-based traffic sign detection algorithm, based on street scene images, was proposed. The experimental results show that the cascaded color segmentation algorithm is effective by fully exploring the color information. A new component-based text line forming method was proposed for recognizing the text on text-based traffic signs. The proposed method also solves the problem of whether a line of text in the text-based traffic sign is horizontal or vertical. An efficient Chinese characters recognition algorithm to recognize the Chinese characters in text-based traffic signs was proposed.

In the future, we will continue to improve the stability of the algorithm and analyze the effects of the occlusion and lighting conditions on the traffic signs detection and recognition. At the same time, the Chinese characters location and recognition method for text-based traffic signs should be further improved.

## REFERENCES

[1] Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li, and S. Hu, "Traffic-sign detection and classification in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 2110–2118.

[2] J. Guo, J. Lu, Y. Qu, and C. Li, "Traffic-sign spotting in the wild via deep features," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Changshu, China, Jun. 2018, pp. 120–125.

[3] X. Rong, C. Yi, and Y. Tian, "Recognizing text-based traffic guide panels with cascaded localization network," in *Proc. Europ. Conf. Comput. Vis. Workshops (ECCV Workshops)*, Amsterdam, The Netherlands, Oct. 2016, pp. 109–121.

[4] X. Ren, Y. Zhou, Z. Huang, J. Sun, X. Yang, and K. Chen, "A novel text structure feature extractor for chinese scene text detection and recognition," *IEEE Access*, vol. 5, pp. 3193–3204, 2017.

[5] S. Wang, L. Huang, and J. Hu, "Text line detection from rectangle traffic panels of natural scene," *J. Phys., Conf. Ser.*, vol. 960, Jan. 2018, Art. no. 012038.

[6] T. He, W. Huang, Y. Qiao, and J. Yao, "Text-attentional convolutional neural network for scene text detection," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2529–2541, Jun. 2016.

[7] H. I. Koo, "Text-line detection in camera-captured document images using the state estimation of connected components," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5358–5368, Nov. 2016.

[8] M. Liao, B. Shi, and X. Bai, "TextBoxes++: A single-shot oriented scene text detector," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 3676–3690, Aug. 2018.

[9] F. Liu, C. Chen, D. Gu, and J. Zheng, "FTPN: Scene text detection with feature pyramid based text proposal network," *IEEE* Access, vol. 7, pp. 44219–44228, Apr. 2019.

[10] Y. Xu, Y. Wang, W. Zhou, Y. Wang, Z. Yang, and X. Bai, "TextField: Learning a deep direction field for irregular scene text detection," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5566–5579, Nov. 2019.

[11] Á. González and L. M. Bergasa, "A text reading algorithm for natural images," *Image Vis. Comput.*, vol. 31, no. 3, pp. 255–274, Mar. 2013.

[12] Y. Zhu, M. Liao, M. Yang, and W. Liu, "Cascaded segmentation-detection networks for text-based traffic sign detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 209–219, Jan. 2018.

[13] S.-J. Hsieh, K.-H. Chen, D.-S. Chen, and Y.-W. Wang, "A detection and recognition system for chinese highway traffic panels," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE-TW)*, Taipei, Taiwan, Jun. 2017, pp. 55–56.

[14] X. He, R. Wang, X. Li, X. Chen, C. Guo, L. Wen, C. Gao, and L. Liu, "HTSTL: Head-and-Tail search network with scale-transfer layer for traffic sign text detection," *IEEE Access*, vol. 7, pp. 118333–118342, Aug. 2019.

[15] X. Chen, L. Qi, Y. Yang, Q. Luo, O. Postolache, J. Tang, and H. Wu, "Video-based detection infrastructure enhancement for automated ship recognition and behavior analysis," *J. Adv. Transp.*, vol. 2020, pp. 1–12, Jan. 2020.

[16] X. Chen, Y. Yang, S. Wang, H. Wu, J. Tang, J. Zhao, and Z. Wang, "Ship type recognition via a coarse-to-fine cascaded convolution neural network," *J. Navigat.*, to be published, doi: 10.1017/S0373463319000900.

[17] Á. González, L. M. Bergasa, and J. J. Yebes, "Text detection and recognition on traffic panels from street-level imagery using visual appearance," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 228–238, Feb. 2014.

[18] J. Guo, R. You, C. Feng, and L. Huang, "Detection of street-level traffic panels based on cascaded color segmentation," in *Proc. 13th Int. Conf. Comput. Sci. Edu. (ICCSE)*, Colombo, Sri Lanka, Aug. 2018, pp. 1–6.

[19] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Minneapolis, MN, USA, Jun. 2007, pp. 1–8.

[20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, San Diego, CA, USA, Jun. 2005, pp. 886–893.

[21] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, Jun. 1998.

[22] N. Kulkarni, "Color thresholding method for image segmentation of natural images," *Int. J. Image, Graph. Signal Process.*, vol. 4, no. 1, pp. 28–34, Feb. 2012.

[23] H. Gómez-Moreno, S. Maldonado-Bascón, P. Gil-Jiménez, and S. Lafuente-Arroyo, "Goal evaluation of segmentation algorithms for traffic sign recognition," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 4, pp. 917–930, Dec. 2010.

[24] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[25] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 2, Kerkyra, Greece, Sep. 1999, pp. 1150–1157.

**JIEFENG GUO** (Member, IEEE) received the B.S. degree from Xiamen University and the Ph.D. degree from the University of Chinese Academy of Sciences. He is currently an Assistant Professor with the School of Electronic Science and Engineering, Xiamen University, Xiamen, China. His current research interest is mainly in the field of image processing and computer vision.

**RONGXUAN YOU** was born in 1994. He received the B.S. degree in electronics and information engineering from Jimei University, Xiamen, China, in 2017. He is currently pursuing the M.S. degree with Xiamen University, Xiamen. His research interest mainly includes text/traffic sign detection and recognition.

**LIANFEN HUANG** received the B.S. degree in radio physics and the Ph.D. degree in communication engineering from Xiamen University, in 1984 and 2008, respectively. She was a Visiting Scholar with Tsinghua University, in 1997, and with The Chinese University of Hong Kong, in 2012. She is currently a Professor with the School of Informatics, Xiamen University, Xiamen, China. Her current research interests include signal processing and image processing.