A Method of Character Detection and Segmentation for Highway Guide Signs

Jiawei Xu1,* and Chongyang Zhang1

1School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China

*2433041042@qq.com

Abstract. In this paper, a method of character detection and segmentation for highway signs in China is proposed. It consists of four steps. Firstly, the highway sign area is detected by colour and geometric features, and the possible character region is obtained by multi-level projection strategy. Secondly, pseudo target character region is removed by local binary patterns (LBP) feature. Thirdly, convolutional neural network (CNN) is used to classify target regions. Finally, adaptive projection strategies are used to segment characters strings. Experimental results indicate that the proposed method achieves new state-of-the-art results.

1. Introduction

The acquisition of visual information of the road is a part of the study of unmanned vehicles. The method proposed in this paper is to detect and segment the characters of Chinese highway signs. Domestic and foreign researchers have proposed some methods for the acquisition of visual information of the road. The extraction of the traffic sign area of interest can be achieved by a combination of colour features, shape features and texture features [1,2,3,4]. For the acquisition of character information in natural scene, there are methods based on connected component [5], texture feature [6] and both of them [7]. Some articles have proposed methods for locating characters on traffic signs in Chinese road scenes. Gu et al. [4] propose a method to obtain rectangular traffic sign regions by colour and shape features, and locate Chinese characters by one vertical projection and one horizontal projection on the binary image. Liu et al. [8] propose a method to get the binary image of the road scene by improving the SCW algorithm. By analyzing the connected components, the rules are defined, and the signs and characters are located. Ni et al. [9] consider that the traffic sign area is obtained by combining shape feature and classifier. Then the image is binarized by K-means colour clustering, and the connected components are analyzed to segment characters. Chen et al. [10] use the colour feature and the HOG feature combined with the support vector machine to obtain the traffic area of interest, and obtains the characters and symbols by analyzing the geometric and location information of the connected components. For Chinese traffic signs with characters, a Chinese character may consist of multiple connected components, and the information of the layout is complicated. For these characteristics, the above mentioned methods have some limitations. In this paper, our method aims to overcome the difficulties above and obtain the character information of the highway signs in China. The character areas we want to detect include highway number, distance character, direction symbol, and Chinese place name. Firstly, the areas of highway sign are obtained by colour segmentation and filtering based on geometric features, and the tilted signs need to be...
corrected. Then a multi-level projection strategy is used to obtain character regions on the binary image. Next, the LBP [11,12] features of character regions are extracted, and are compressed by the model of piecewise function. Target character regions and pseudo target character regions are classified by the AdaBoost classifier [13,14]. Finally, the target character regions are classified by CNN [15,16,17,18,19], and four different character regions such as highway number, distance character, direction symbol and Chinese place name are obtained. Different types of characters are segmented by adaptive projection strategies. The framework of our method is shown in Figure 1. Section 2, 3, 4 introduce the details of the method. Section 5 introduces the experiment. Conclusions are made in section 6.

![Algorithm flowchart](image)

**Figure 1.** Algorithm flowchart

2. Location of Highway Guide Sign
A highway guide sign is a rectangular panel of dark green, which can be roughly detected by colour segmentation. In order to avoid the influence of light and other factors, a robust colour segmentation method is proposed. Pixel values of the image in the RGB colour space and the HSV colour space are obtained. Then the brightness of the image is divided into three intervals: \( v_1 \), \( v_2 \), and \( v_3 \). In the three different intervals, the separation thresholds of the colour space are obtained statistically, and the dark green regions are obtained by the thresholds of the two colour spaces. After colour segmentation, part noises are removed by area feature. Then the operations of morphological dilation, cavity filling, and morphological corrosion are performed. According to the aspect ratio of the region and the area ratio of the connected component to the external rectangle, noises are filtered out. Finally, the regions of highway guide signs can be obtained. Figure 2 shows some test results.

![The result of highway guide signs detection in some scene of highway](image)

**Figure 2.** The result of highway guide signs detection in some scene of highway

3. Detection of target character regions

3.1. Multi-level projection strategy
A multi-level projection strategy is proposed to obtain the possible character regions from the highway guide signs. First of all, a binary image of the guide sign is obtained by binarization, and corner points of the sign are found by Hough transform on the binary image. According to the corner points, affine transformation can be used to correct the tilt sign. By analysis of the layout, we can see that the layout is very complex, and a Chinese character may consist of multiple connected components. In order not to destroy the integrity of characters, a plurality of combined projection operations can be carried out,
and the complicated layout can be completely cutoff. Each combined projection operation consists of a vertical projection operation and a horizontal projection operation. Projection is done on a binary image. Vertical projection operation means the accumulation of pixels which consists of connected components, and is applied to each row of image. A row of the image, of which projection value is 0, can be cut off. Similarly, Horizontal projection operation is for image columns. After a combined projection operation, it is necessary to judge whether the obtained regions satisfy the condition for continued projection. If the condition is satisfied, a combined projection operation will continue until complete segmentation is performed. By analysis of a large number of layouts, regions that meet the following three situations need a combined projection operation.

- A region accounts for a larger proportion of the layout.
- A region is vertical strip.
- The upper part of the region consists of Chinese characters, and the lower part of the region consists of English alphabets. The Chinese characters are usually a name of a place, and the English alphabets are Pinyin or English name of the place. This is a fixed pattern.

After cutting, individual components are merged into character regions by character lines and layout rules. As shown in Figure 3, the target regions include highway numbers, distance characters, direction symbols, and Chinese place names in the white font. Others are pseudo target regions.

![Figure 3](image)

**Figure 3.** The character regions are obtained by a multi-level projection strategy

### 3.2. LBP feature extraction and feature dimension compression

LBP feature [12] is an operator used to describe the local texture feature of an image, which has the advantages of rotational invariance and grayscale invariance. The original LBP operator defines a 3*3 square window and takes the value of the center pixel of the window as the threshold. Comparing the adjacent pixel values to the threshold, if the pixel value is greater than the threshold, then the pixel position is set to 1, otherwise 0. Thus, you can get an 8 bit binary number and convert it to decimal number, which is the LBP value of the central pixel. As shown in formula (1).

$$LBP(x, y) = \sum_{n=0}^{7} 2^n S(v_n - v_c), \quad S(x) = \begin{cases} 1, & \text{if } (x \geq 0) \\ 0, & \text{if } (x < 0) \end{cases} \tag{1}$$

$S(x, y)$ is a symbolic function, and $(x, y)$ is the central pixel of the window. $v_n$ is the pixel value of neighborhood pixels, and $v_c$ is the center pixel value. A circular neighborhood is used instead of a square neighborhood, and a circular field with a radius of $R$ is allowed to have any number of pixels, which can be adapted to different sizes of texture features. For the central pixel, the points of its neighborhood can be computed by the following formula (2). $R$ is the radius of the circle region, and $N$ is the number of pixels in the neighborhood.

$$\begin{align*}
x_n &= x_c + R \cos \left(\frac{2\pi n}{N}\right) \\
y_n &= y_c + R \sin \left(\frac{2\pi n}{N}\right)
\end{align*} \tag{2}$$

$$f(x, y) = \left[ \begin{array}{c} f(0,0) \\ f(0,1) \\ f(1,0) \\ f(1,1) \end{array} \right] \approx \left[ \begin{array}{c} 1-y \\ -x \end{array} \right] \begin{bmatrix} 1 & \quad 0 \\ 0 & \quad 1 \\ -1 & \quad 0 \\ 0 & \quad -1 \end{bmatrix} \left[ \begin{array}{c} 1 \quad y \end{array} \right] \tag{3}$$
If the calculated coordinates and pixel coordinates are biased, the bilinear interpolation is used for approximate processing, such as formula (3). Figure 4 is the image of the character regions and their LBP.

Figure 4. Images of some character regions and their LBP images: G42, Zhongshan East Road, 500m, arrow

The LBP feature of an image can be obtained as follows. The image is divided into some blocks which size is $n \times n$, and the LBP of each block is calculated. Then we calculate the frequency of each value of the LBP, build the histogram of LBP of the block, and normalize the histogram. Finally, the histograms of each block are concatenated into a feature vector, which is the LBP texture feature vector of the image. In order to reduce the dimension of the feature, we propose a LBP feature based on the reduced model of the piecewise function (RdLBP), as shown in formula (4).

$$f(LBP(x, y)) = \begin{cases} 
0, & LBP(x, y) \in [0, n), \\
1, & LBP(x, y) \in [n, 2n), \\
2, & LBP(x, y) \in [2n, 3n), \\
\text{and so on.}
\end{cases}$$

(4)

3.3. AdaBoost classifier

The AdaBoost algorithm [13] is a way to generate a strong classifier from a series of weak classifiers. Suppose there are a feature vector set $X$ and a label set $Y$, then each feature vector corresponds to a label. The AdaBoost algorithm generates a weak classifier set $K$ after N iterations. Each weak classifier can decide which tag the feature vector belongs to. After the training of $n-1$ iterations, $n-1$ weak classifiers are combined to produce a strong classifier. Then a new strong classifier can be generated by combining the $n$-th week classifier with the previous strong classifier, such as formula (5). The weight $\alpha_n$ and the weak classifier $k_n$ need to make the strong classifier $F_n$ better than the $F_{n-1}$ classifier.

$$F_n = F_{n-1} + \alpha_n k_n$$

(5)

$$F_n = \text{sign}(\sum_{n=1}^{N} k_n(x))$$

(6)

The AdaBoost algorithms include Discrete AdaBoost, Real AdaBoost, LogitBoost and Gentle AdaBoost. And the Gentle AdaBoost is used here. The weak classifier $k_n$ is obtained by regression fitting of weight-based weighted least square method from label to feature vector. After normalizing the weights, the final strong classifier is shown in formula (6). Figure 5 is a graph showing the result of removing the pseudo targets in Figure 3.

Figure 5. Highway guide signs after removing pseudo targets

4. Character segmentation
There are four types of target character regions: Highway number, distance character, direction symbol and Chinese place name. In order to achieve better cutting effect of characters, the target character regions are classified by CNN, and then adaptive projection strategies are adopted.

### Table 1. The network hierarchy of Caffenet

| Conv1 | Conv2 | Conv3 | Conv4 | Conv5 | Fc6 | Fc7 | Fc8 |
|-------|-------|-------|-------|-------|-----|-----|-----|
| data  | norm1 | norm2 | conv3 | conv4 | pool5 | InnerProduct | InnerProduct | InnerProduct |
| conv1 | conv2 | conv3 | conv4 | conv5 | fc6  | fc7  | fc8  |
| ReLU  | ReLU  | ReLU  | ReLU  | ReLU  | ReLU | ReLU | Softmax |
| pool1 | pool2 | pool3 | pool4 | pool5 | fc6  | fc7  | fc7  |
| norm1 | norm2 |

### 4.1. Classification based on CNN

CNN is a multi-layer artificial neural network [18], which usually consists of feature extraction layer and feature mapping layer. The input of each neuron is connected to the local accepted domain of the previous layer and extracts the feature of the region. Once the local feature is extracted, the positional relationship between it and other features is determined. Each computational layer of the network consists of multiple feature mappings, each of which is a plane, and the weights of all the neurons on the plane are equal. Each convolutional layer follows a computational layer for local averaging and secondary extraction, and this unique feature extraction structure reduces the feature resolution. Caffenet network structure [15] is adopted here. It consists of eight weighted layers. The first five layers are convolutional layer, and the latter three are inner product layer. The structure of the network is shown in Table 1.

Convolution is an operation of analytic mathematics. Assuming that \( f(x) \) and \( g(x) \) are integrable functions [17] on the set \( R \), the formula (7) defines the convolution of function \( f(x) \) and \( g(x) \).

\[
F = \int_{-\infty}^{\infty} f(a) \cdot g(x - a)da
\]

(7)

Convolution of different positions of image can obtain different eigenvalues. The convolutional layer is the core layer of convolutional neural network, and convolution operation is performed by parameter setting. The pooling layer is basically the same as the convolutional layer, with the aim of reducing the data dimension, similar to downsampling. The LRN layer normalizes the local region of an input and achieves the effect of lateral inhibition. The activation layer activates the input elements, using the ReLU activation function [19].

\[
F(x_i) = \begin{bmatrix}
    p(y_1 | x_i) \\
    p(y_2 | x_i) \\
    ... \\
    p(y_k | x_i)
\end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{l_{ij}}} \begin{bmatrix}
    e^{l_{i1}} \\
    e^{l_{i2}} \\
    ... \\
    e^{l_{ik}}
\end{bmatrix}
\]

(8)

The innerproduct layer is actually also a convolutional layer, but its convolutional kernel is the
same size as the original data. Dropout is a strategy of preventing over-fitting, which allows the network to hide the weight of the nodes which belong to default hidden layer randomly. The ultimate goal of the Softmax layer is to obtain the probability likelihood values for each class, and the output function [17] follows the formula (8).

\[ P(y_i | x_i) \text{ indicates the probability that } x_i \text{ belongs to } y_i, \text{ and } w \text{ is parameter of the model.} \]

\[
\sum_{j=1}^{d} e^{w_{ij}x_i}
\]

normalizes the probability distribution.

The model is obtained by training of the network. Firstly, the corresponding values of output and loss function are calculated by forward propagation, and then the gradient of loss is calculated by back propagation, and the new weights and parameters are obtained. Finally, the appropriate model is obtained. In the prediction, the original image is fed into a convolutional neural network, and the category and probability value are obtained.

4.2. Different types of character regions segmentation

After getting the category of character area, adaptive projection strategies are adopted. A Chinese character may contain more than one connected component. When cutting Chinese place names, it is necessary to consider the aspect ratio of Chinese characters, and cutting is based on the horizontal projection strategy. Generally speaking, the width to height ratio of Chinese characters is about equal to 1. Distance characters and highway numbers contain both English letters and numbers, in which a single character contains only one connected component, and the aspect ratio of different characters fluctuates considerably. We can directly cut each character by horizontal projection. A direction symbol contains only one symbol, and does not need to be cut again. After cutting all the characters, the non-character parts of the region above and below are removed by vertical projection. The horizontal projection and vertical projection here are the same as those in front. Figure 6 shows an instance.

**Figure 6.** An instance of cutting different character regions: Beijing, S48, 3km, arrow

5. Experiment

5.1. Experimental condition

The data used in the experiment is photos of actual highway scene collected by unmanned vehicles, and the resolutions of these photos are 2448*2050. Pictures are collected in the morning and afternoon. In the morning, from Nanjing to Wuxi, the weather is fine. In the afternoon, from Wuxi to Nanjing, the weather is cloudy. Throughout the process, the speed is about 90km/h, and the direction and intensity of illumination vary. The experimental software condition is Windows 7+vs2013+opencv 2.4.10, and the hardware condition is Intel Core i5-2450M CPU and 4GB of RAM.

5.2. Experimental results

In this experiment, Precision and Recall are used to describe the experimental results, as shown in the formula (9).

\[
\begin{align*}
\text{Precision} &= \frac{TP}{TP + FP} \\
\text{Recall} &= \frac{TP}{TP + FN}
\end{align*}
\]

\( TP \) is the correct detected region, \( FP \) is the wrong detected region, and \( FN \) is the correct region.
for missed detection. As to the detection of character regions, this method combined with machine learning can detect more character region information than [4,8,9,10]. The correct regions include the highway number regions, the distance character regions, the direction symbol regions and the Chinese place name regions. The wrong regions are some other regions. The experimental results are shown in Table 2. Firstly, the possible character regions are obtained by multi-level projection. Then each region is normalized to 60*60, and each region is divided into some blocks of which size is 10*10, and the LBP feature of each region is obtained statistically. Next the dimension of the LBP feature is reduced, and the size of the dimensionality reduction is 8. The dimension of the original LBP feature is 9216, and the dimension of the LBP feature is 1152 after reduction. Finally, the AdaBoost classifier is used to remove the pseudo target character regions.

| Method                     | Number of character regions | Accuracy (%) | Recall (%) | Time (ms) |
|----------------------------|-----------------------------|--------------|------------|-----------|
| multi-level projection     | 1498                        | 69.69        | 97.46      | 44        |
| multi-level projection +LBP+AdaBoost | 1498                        | 96.30        | 97.19      | 45        |
| multi-level projection +RdLBP+AdaBoost | 1498                        | 96.48        | 97.19      | 35        |

Experimental results show the detection effect is not very good when the pseudo targets are removed only by the geometric feature and not by the machine learning method. For example, some other character regions such as low discrimination regions of Pinyin will affect the detection. It can be seen that the use of LBP combined with AdaBoost is very effective. At the same time, the experimental results before and after dimension reduction are similar, but the time of execution is obviously shortened after dimension reduction.

For the experiment of character segmentation, the research[20] proposed a method based on connected component and a method based on projection. Compared with the segmentation method in this paper, the experimental results are shown in Table 3.

In the experiment, if each character in character region is correctly segmented, the region is considered to be correctly segmented. The experimental results show that the segmentation method proposed in this paper is more accurate than the other two methods.

| Method                                          | Accuracy (%) |
|-------------------------------------------------|--------------|
| A segmentation method based on connected component | 81.02        |
| A segmentation method based on projection        | 94.69        |
| The segmentation method of this paper            | 97.76        |

6. Conclusion
In this paper, the method of detecting and segmenting the character for highway guide sign is effective. We can detect the useful characters on the sign, including the place name, distance, direction and so on. This provides visual information for unmanned vehicles and supports double validation of visual information and satellite navigation information on expressways. However, there are many factors affecting the quality of images, and the environment of the road changes drastically. Further research is needed to do to improve the adaptability and robustness of the algorithm. And after getting the characters, we need to recognize the characters and understand the semantics, which needs further research.

Acknowledgements
We acknowledge the support of the National Science and Technology Major Project of China, under
References

[1] S. Salti, A. Petrelli, F. Tombari, N. Fioraio, and L. D. Stefano. Traffic sign detection via interest region extraction. *Pattern Recognition*, 48(4):1039-1049, 2015.

[2] W. Chen, C. Li, and Z. Wang. Road Traffic Sign Detection Using Color and shape. *Journal of Xiamen University(Natural Science)*, 46(5):635-640, 2007.

[3] Y. Chen, F. Ye, F. Xiao, and Q. Li. Detection and Recognition of Traffic Signs Based on HSV Vision Model and Shape Features. *J. Jianghan Univ.(Nat. Sci. Ed.)*, 44(2):119-125, 2016.

[4] M. Gu, Z. Cai, and X. Ren. Detection algorithm of traffic sign with character in nature scenes. *Journal of Central South University(Science and Technology)*, 44(5):120-127, 2013.

[5] S. Tian, S. Lu, B. Su, and C. Tan. Scene Text Segmentation with Multi-level Maximally Stable Extremal Regions. In *Proceedings of the 22nd International Conference on Pattern Recognition*, pages 2703-2708, 2014.

[6] J. Fabrizio, B. Marcotegui, and M. Cord. Text detection in street level images. *Pattern Analysis and Applications*,16(4):519-533, 2014.

[7] G. Zhou, Y. Liu, Z. Tian, and Y. Su. A New Hybrid Method to Detect Text in Natural Scene. In *Proceedings of the 18th IEEE International Conference on Image Processing*, pages 2605-2608, 2011.

[8] F. Liu, and H. Sha. An Algorithm for the Detection of Text on Road Signs Based on Improved SCW. *Journal of Image and Graphics*, 12(10):1869-1872, 2007.

[9] C. Ni. Study on Traffic Sign Text Detection and Recognition Based on Video. *Beijing Jiaotong University*, 2015.

[10] Y. Chen. Study on Traffic Panel Layout Understanding from Natural Scenes. *Beijing Jiaotong University*, 2016.

[11] J. Ren, X. Jiang, and J. Yuan. Learning LBP structure by maximizing the conditional mutual information. *Pattern Recognition*, 48(10):3180-3190, 2015.

[12] M. Ghorbani, A. T. Targhi, and M. M. Dehshibi. HOG and LBP: Towards a robust face recognition system. In *International Conference on Digital Information Management*, pages 138-141, 2016.

[13] C. Yan, Y. Wang, J. Li, and Z. Zhang. Theory deduction of AdaBoost classification. *Journal of Southeast University(Natural Science Edition)*, 41(4):701-705, 2011.

[14] D. Yang, Z. Zhao, and Q. Bai. Face detection based on AdaBoost and Space Support Vector Domain classifier. *Journal of Circuits and Systems*, 18(2):321-325, 2013.

[15] Y. Jia, E. Shelhamber, J. Donahue, S. Karayev, J. Long, R. Girshick, et al. Caffe: Convolutional architecture for fast feature embedding. In *EprintArxiv*, pages 675-678, 2014.

[16] R. Girshick. Fast R-CNN. In *IEEE International Conference on Computer Vision*, pages 1440-1448, 2015.

[17] F. Li. Convolutional Neural Network and Its Applications. *Dalian University of Technology*, 2014.

[18] H. Lu, and Q. Zhang. Applications of Deep Convolutional Neural Network in Computer Vision. *Journal of Data Acquisition and Processing*, 31(1):1-17, 2016.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. *International Conference on Neural Information Processing Systems*,25(2):1097-1105, 2012.

[20] Q. Weng. Research on Optimized Preprocessing and Character Segmentation of Irregular Vehicle Plate. *Zhejiang University*, 2014.