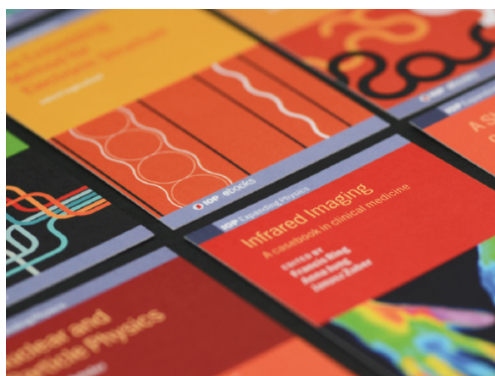


PAPER • OPEN ACCESS

Recognition of Speed Signs in Uncertain and Dynamic Environments

To cite this article: Zhilong Zhu et al 2019 *J. Phys.: Conf. Ser.* **1187** 042066

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

Recognition of Speed Signs in Uncertain and Dynamic Environments

Zhilong Zhu^{1,2,3}, Gang Xu^{1,2*}, Hongmei He³, Juanjuan Jiang^{1,2}, Tao Wang¹

¹ College of Electrical Engineering, Anhui Polytechnic University, Wuhu 241000, China

² Anhui Key Laboratory of Detection Technology and Energy Saving Devices, Wuhu 241000, China

³ Manufacturing Informatics Centre, SATM, Cranfield University, Cranfield, MK43 0AL, UK

Email: zhuzhilong919@ahpu.edu.cn; jiangjuanjuan@ahpu.edu.cn; h.he@cranfield.ac.uk

*Corresponding Author: Gang Xu; email: xugang@ahpu.edu.cn

ABSTRACT: The speed limit signs recognition directly affects the safety of autonomous vehicles. Vehicles are usually running in an uncertain and dynamic environment. The performance of the recognition system is affected by various factors such as the different sizes of pictures, illumination condition and position circumstances, which can lead to misclassification. This makes the speed sign recognition challengeable. To improve the recognition rate of the speed signs in such environments, this work firstly applies the method of the saliency target detection based on the background-absorbing Markov chain, to extract the node in an image, then uses SPP-CNN to classify the extracted nodes with ten-folder validation. The recognition rate is up to 9.32%, higher than that obtained directly by SPP-CNN working on raw dataset.

1. INTRODUCTION

Autonomous vehicles are future trend in automotive sector. Speed sign recognition is important for making good traffic order and ensuring the safety of road users. Traffic signs include text-based and symbolic signs, such as text, numbers, arrows, etc., representing warnings, prohibitions, and other information, using specific colors (e.g. red, yellow, blue, white and black), and specific shapes (e.g. triangles, rectangles, circles and octagonal). Different countries have different traffic signs [1].

The initial road traffic sign recognition (TSR) was based on the color space in the image [2]. Since the first paper on road TSR appeared, various computer vision methods have been developed to detect and identify road traffic signs. The identification of traffic signs is generally divided into three steps: segmentation, detection, recognition. The purpose of segmentation is to obtain the location of traffic signs, and methods include color-based and shape-based. A color-based method is to extract the specific color region; a shape-based method generally relies on edge detection to obtain the shape of traffic signs. Detection determines whether there exists a signpost from the output of segmentation, and the purpose of detection is to reduce the search space to obtain interesting area. Recognition is to



classify it into different groups. The speed sign recognition is a critical challenge in the real-time TSR.

However, due to the complexity of the roads and the surrounding environments, speed signs can appear in a variety of scenes, such as the change of sign color due to long-term outdoor environments, dust pollution in the air, the impact of weather conditions, the impact of light, obstruction, view changes, damage, etc. This brings great challenges to automatic speed sign recognition.

Most of the existing TSR systems use the inner region of the signs or the local features such as histograms of oriented gradients, and scale-invariant feature transform for recognition. Although the current technology has made the great progress in the road sign recognition, there is still space to improve the performance of recognition techniques in both effectiveness and efficiency, as well as robustness in dealing with the rotation, illumination, pollution, damage and scale variations in uncertain and dynamic environments.

Deep learning has achieved the success in several areas, such as image processing, natural language processing, etc. Especially convolution neural networks (CNNs) have been widely applied in image-based pattern recognition. In 2014, Girshick [3] first applied R-CNN to categorize the target objects. It calculates the deep CNN to each candidate scheme, and thus generates a lot of information unrelated to the class, increasing the computing and time complexity. In order to improve the efficiency, He et al. [4] developed the Spatial Pyramid Pool Network, and calculated the convolution feature graph of the whole image to produce the feature vector. This accelerated the R-CNN by nearly 100 times. Lee et al. [5] used CNN to recognize the traffic signs based on the boundary estimation and end-to-end training. Zang et al. [6] proposed a framework for the recognition of traffic signs using CNN-based fusion of space and time characteristics. Zhu et al. [7] created a traffic sign dataset, covering illumination and weather conditions, including 100000 examples of 30,000 traffic signs, and they use CNN to detect and classify traffic signs.

In order to improve the recognition rate and real-time performance of Speed Sign Recognition (SSR), we propose a new method that consists of two stages: *Segmentation*, using the Background-Absorbing Markov Chain (BAMC) model [8] [9] to remove background, and *Recognition*, using Spatial Pyramid Pooling based convolution neural networks (SPP-CNN) on the images without background. The experiments will be done on a real-world database, built by the researchers, including various tiled positioned, scaled, and light blurred images.

2. PROPOSED METHOD

2.1. Detection of saliency targets based on BAMC

SLIC segmentation [8] was used to divide the input image into m pixel blocks (nodes). Generally, the pixel blocks at the 4 boundaries (top, bottom, left, right) of an image do not cover the target. Initially k nodes on three selected boundaries are copied as absorption nodes (virtual nodes) in the absorbing Markov Chain. A graph model $G(V, E)$ is established with n nodes, represented by the set of V , $n = k + m$, E is the set of edges, which represent the relations between nodes as the rules: (1) the nodes are correlated to the adjacent pixel blocks; (2) the nodes of all the pixel blocks at all boundaries are correlated with each other; (3) the copied k virtual nodes are not correlated with each other; (4) if node i is a node correlated to a node on the inner boundary of the image, a new node i' is copied, then node i is correlated with node i' and all nodes, correlated with i , are correlated with i' .

Therefore, the correlation matrix A of the nodes in the graph model $G(V, E)$ can be calculated:

$$\alpha_{ij} = \begin{cases} w_{ij} & j \in N(i), 1 \leq i \leq n \\ 1 & i = j \\ 0 & \text{else} \end{cases}, w_{ij} = e^{-\frac{\|x_i - x_j\|}{\sigma^2}}. \quad (1)$$

where, $N(i)$ is the set of nodes correlated with the node i , the w_{ij} is the weight between correlated nodes i and j , x_i and x_j as a color mean value to the corresponding hyper-pixel block in the color space, σ is a constant. The m pixel blocks in the image are considered as the transfer nodes in the absorption Markov Chain, the k pixel blocks are the absorption nodes. The degree matrix that records the sum of the weights connected to each node is written as $D = \text{diag}(\sum_j a_{ij})$. The transition matrix $P = D^{-1} \times A$. With

the k -absorbing states and the m -transfer states, P can be reconstructed as

$$P = \begin{bmatrix} Q & R \\ 0 & I \end{bmatrix}, \quad (2)$$

where, $Q \in [0,1]_{m \times m}$, transferring matrix, representing the probabilities between the m transferring nodes; $R \in [0,1]_{m \times k}$, representing the probability between the m transferring nodes and the k absorbing nodes; 0 as a $k \times m$ zero matrix, I is the unit matrix, representing the k absorbing nodes.

For an absorption chain, the basic matrix $N = (I - Q)^{-1}$ can be obtained, n_{ij} is the expected absorption time from the transfer node i to the transfer node j . According to n_{ij} , we can calculate the vector $y = N \times J$, which consists of the absorption time by all transfer nodes, where J is a m -dimension column with all elements are 1. According to the Markov Chain, y is smaller, the corresponding pixel block of the node is more likely to be the background; y is larger, the absorption time is longer, the pixel block may be a significant target. We use the average value of y to represent the significant value s , that is $s(i) = \bar{y}(i)$, $i = 1, 2, 3, \dots, m$, the number of transition state nodes in the graph, $\bar{y}$ is the normalized absorption time vector.

Boundary selection [9] [10] is based on the saliency values of the four boundaries. The boundary with largest saliency value could include the target; hence, the pixel blocks on the three boundaries are copied as virtual nodes (i.e. absorption nodes) to the set C .

After getting the absorption nodes from the three boundaries, we need to identify more absorption node in the image. The absorption time of all super pixel blocks in M is calculated, and the initial saliency value $S_i^{initial}$ of node i is obtained. If the initial saliency value is less than the threshold T , the node i is more likely background, then node i , as an absorption node, is added to set C .

$$C = C \cup \{i \mid S_i^{initial} \leq T\} \quad (3)$$

Algorithm 1 presents the pseudocode of the BAMC process, where, the parameters include: Im is an image to be processed; b indicates the size of a $b \times b$ super pixel blocks; M contains all the m super pixel block nodes; C contains all the copied pixel block nodes as background nodes.

The image is first segmented by the SLIC algorithm, and the background pixel block nodes on the three boundaries are identified. For each layer, identify the correlated nodes in the set (M) of super pixel block nodes, and assess if they are belong to background nodes in terms of the defined rules of pixel block relationships and the criteria in Formula (3), move these nodes to the set C , and remove them from M , if they are background nodes. Finally, return the segmented sets of absorption nodes (C) and target nodes (M).

Algorithm 1 BAMC(Im, b, T, L)

```

1:  $[M]_m = \text{SLIC}(Im, b, L)$ ;
2:  $[C] = \text{BoundarySelection}(B)$ ;
3: for ( $i=1 \dots L$ ) do
4:    $G = \text{constructGraph}(M, C)$ ;
5:    $A = \text{calMatrix}(G)$ ;
6:    $D = \text{calDegreeMatrix}(A)$ ;
7:    $P = D^{-1} \times A$ ;
8:    $Q = \text{reconstruct}(P)$ ;
9:    $N = (I - Q)^{-1}$ ,  $y = N \times J$ ;
10:  for ( $j=1 \dots m$ ) do
11:     $S_{ij} = \bar{y}_j$ ;
12:    if ( $S_{ij} < T$ ) then
13:       $C \leftarrow j$ ;
14:       $M \leftarrow M \setminus j$ ;
15:    end if
16:  end for
17: end for
18: return ( $M, C$ );
```

2.2. Classification with SPP-CNN

The last full connected layer of CNN requires fixed-length vectors [11]. Spatial pyramid pooling can maintain spatial information by pooling in local spatial bins. These spatial bins have sizes proportional to the image size, so the number of bins is fixed regardless of the image size [4] [12]. To make the CNN robust for any size of images, we use spatial pyramid pooling in the deep convolution neural network (SPP-CNN) to recognize the speed signs regardless the size of images produced by the BAMC algorithm. A typical five-layer CNN with the SPP layer as the last pool layer is developed. The CNN passes the features from last convolution layer to the SPP layer. The fixed size of outputs from the SPP layer is passed to the final full connect layer.

3. EXPERIMENT RESULTS AND DISCUSSION

3.1. Experiment setup

Both training and testing of the SPP-CNN were done on a PC with AMD thread ripper 1900x CPU, GTX 1070 Ti Hybrid GPU, 32G memory, the SPP-CNN model is trained on GTX1070 GPU.

The Databases of speed limit sign used in our experiments include 587 samples, which are collected by us on a variety of different roads in the United Kingdom. It includes city streets, countryside, high ways, various luminous environments and positions. Most of samples have a complex background with various target sizes. Two data set were produced as follows: a. the first speed limited sign data set is the original images, denoted as FSLSD as shown in Figure. 1; b. the set of black-and-white traffic speed limit sign extracted from original images with the BAMC algorithm, denoted as the second dataset (SSLSD) in the second line of Figure. 2. From 587 samples, we have chosen 59 samples with different sizes, 52 light-blurred samples, and 46 positions tilted samples for experiment assessment.



Figure 1. FSLSD is obtained by annotate in the original image



Figure 2 SSLSD is obtained based on Background-Absorbing Markov chain

In the segment experiment, the initial number of pixel blocks is set $m=250$, the threshold value is set $T=0.015$. As compared effect of the fusion, the number of the layer $L = 3$. The images in Figure 2 are some results produced by Algorithm 1.

In order to verify how the proposed method improve the performance of the speed limit sign recognition, we conducted the experiments as follows: a. For each dataset, we use the five-layer

SPP-CNN, using ten-folders crossing validation. Namely, for each run, 10% samples are randomly picked as test set, and the rest 90% of images as the training set. b. take the 59 samples with different sizes, 52 light-blurred samples and 46 positions tilted samples as the test set respectively, and calculate the recognition rate.

3.2. Experiment results

Figure 3 is the mean loss of training set results to the two datasets; Table 1 is the results of ten-folders crossing validation; Table 2 is the test results of severe circumstances samples with proposed method

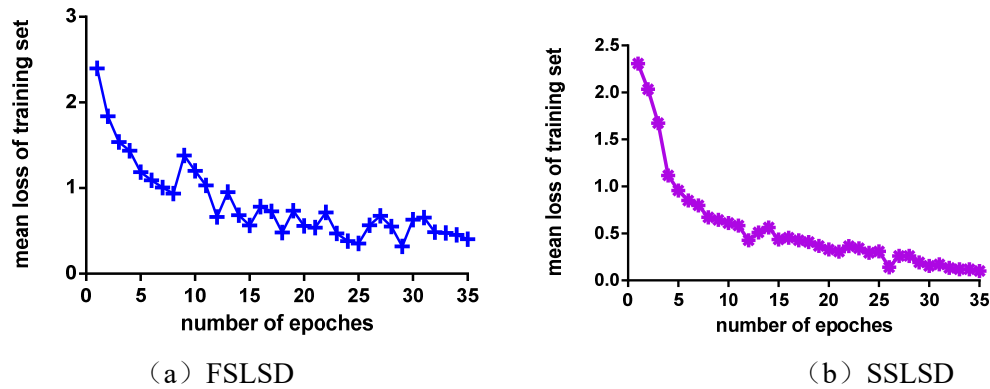


Figure 3 The mean loss curve of training set for ten-fold training

Table 1 Ten-folder crossing validation test results of the two Datasets

Test turns	1	2	3	4	5	6	7	8	9	10	mean
FSLSD(%)	86.44	93.22	88.14	91.53	93.22	86.44	86.44	89.83	89.83	91.53	89.66
SSLSD(%)	98.30	100	100	98.30	96.61	100	98.30	100	100	98.30	98.98

Table 2 Special circumstances sample test results based on proposed method

Image conditions	Different size	Light-blurred	Positions tilted
Recognition results	59/59	50/52	45/46
Accuracy(%)	100	96.15	97.83

3.3. Discussion

It can be seen from Figure 3 (a) and (b) that the baseline model used in this paper is convergent during the training process with the iteration epochs increases, and the convergence trend of figure (b) is better than that of figure (a), it illustrated that the black-and-white traffic speed limited sign, which is obtained by proposed method in this paper has better effect to neural network learning.

From Table 1, it can be obviously seen that the recognition rate of SSLSD is better than that of FSLSD. This is because that the traffic speed limit signs as a saliency targets are ensured by the transfer nodes, then the recognition of a speed limit sign is completed by the SPP-CNN.

From 587 original images, some with different sizes, blurred lights or titled positions in uncertain environments are selected, respectively. The proposed approach is applied on these subsets. Table 2 shows the results. With the proposed approach, BAMC-SPP-CNN, the recognition rate on the 59 images with different sizes is 100%, the recognition rate on the 52 image with blurred lights is 96.15%, and the recognition rate on the 46 images with tiled positions is 97.83%. Due to the introduction of SPP, the CNN is robust for different sizes of speed signs.

4. CONCLUSIONS

The contribution of this paper is summarized as follows: (1) for salient target segmentation based on Markov chain, first three boundary nodes are used as the absorption nodes, the absorption time of all the super-pixel blocks in an image is calculated, and the foreground is distinguished from the background based on the specified threshold. (2) a new database of images without background is produced to be the input of CNN for classification, and the ten-folder crossing validation is used for the performance assessment. (3) The recognition rate of SPP-CNN, fed with SSLSD is up to 9.32% better than that of the SPP-CNN, fed with FSLSD.

Acknowledgements

This research was financially supported by the Key project of natural science in universities in Anhui Province (KJ2018A0111). Anhui Key Laboratory open project of Detection Technology and Energy Saving Devices, Anhui Polytechnic University, China (2017070503B026-A01), (1506c085002). Anhui Polytechnic University student Research Project (KC00418025).

References

- [1] M.-Y. Fu, Y.-S. Huang. (2010) A survey of traffic sign recognition. in Proceedings of International Conference on Wavelet Analysis and Pattern Recognition, July, 119–124.
- [2] H. Fleyeh, M. Dougherty. (2005) Road and traffic sign detection and recognition,” in Proceedings 10th EWGT Meet/ 16th Mini-EURO Conference, pp.644 – 653.
- [3] Girshick, R., Donahue, J., Darrell, T., et al. (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 580–587.
- [4] He, K., Zhang, X., Ren, S., et al. (2014) Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 1–14.
- [5] Lee, H. S., Kim, K. (2018) Simultaneous Traffic Sign Detection and Boundary Estimation using Convolutional Neural Network. Transactions on Intelligent Transportation Systems. 1–12.
- [6] Zang, D., Wei, Z., Bao, M., et al. (2018) Deep learning-based traffic sign recognition for unmanned autonomous vehicles. Proceedings of the Institution of Mechanical Engineers, Journal of Systems and Control Engineering. 232(5) 497–505.
- [7] Zhu, Z., Liang, D., Zhang, S., et al. (2016) Traffic-Sign Detection and Classification in the Wild. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2110–2118.
- [8] Jiang B W, Zhang L H, Lu H C, et al. (2013) Saliency detection via absorbing Markov chain. Proceedings of IEEE International Conference on Computer Vision. Sydney, NSW, Australia. 1665-1672.
- [9] Jiang F L, Zhang H T, Yang J, et al. (2018) Image saliency detection based on background absorbing Markov chain [J]. Journal of Image and Graphics. 23(6) 0857-0865.
- [10] C. Liu, F. Chang, Z. Chen, et al. (2016) Fast traffic sign recognition via high-contrast region extraction and extended sparse representation. IEEE Transactions on Intelligent Transportation Systems. 17(1) 79–92.
- [11] T. Chen, S. Lu. (2016) Accurate and efficient traffic sign detection using discriminative AdaBoost and support vector regression. IEEE Transactions on Vehicular Technology. 65(6) 4006–4015.
- [12] Zhao, H., Shi, J., Qi, X., et al. (2017) Pyramid scene parsing network. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017–January, 6230–6239.



DSPACE

<https://dspace.org/>

Recognition of speed signs in uncertain and dynamic environments

Zhu, Zhilong; Xu, Gang; He, Hongmei; Jiang, Juanjuan; Wang, Tao

2019-05-08

IOP Publishing

<https://doi.org/10.1088/1742-6596/1187/4/042066>

Zhu Z, Xu G, He H, et al., (2019) Recognition of speed signs in uncertain and dynamic environments. Journal of Physics: Conference Series, Volume 1187, Issue 4, 2019, Article Number 042066