A robust, coarse-to-fine traffic sign detection method

Gangyi Wang, Guanghui Ren, Zhilu Wu, Yaqin Zhao and Lihui Jiang

Abstract— **We present a traffic sign detection method which has won the first place for the prohibitory and mandatory signs and the third place for the danger signs in the GTSDB competition. The method uses the histogram of oriented gradient (HOG) and a coarse-to-fine sliding window scheme. Candidate ROIs are first roughly detected within a small-sized window, and then further verified within a large-sized window for higher accuracy. Experimental results show that the proposed method achieves high recall and precision ratios, and is robust to various adverse situations including bad lighting condition, partial occlusion, low quality and small projective deformation.**

I. INTRODUCTION

TRAFFIC sign recognition is an important function for driver assistance systems. Although it has been studied for many years, existing methods are still far from mature. The major difficulties include bad lighting condition, similar background color, partial occlusion, low quality, etc. Some difficult examples are shown in Fig. 1.

Fig. 1. Difficult cases for traffic sign recognition.

There are usually two steps for most traffic sign recognition methods: detection and classification. The detection step finds out the region of interests (ROI), each of which contains a traffic sign; and the classification step determines the classes of the traffic signs. As methods for classification have shown great success in recent researches [1], we only focus on traffic sign detection method in this paper.

Color-based detection methods are very popular because traffic signs are usually in fixed colors. These methods are usually fast and robust to projective deformation, but sensitive to the lighting condition. To solve this problem, some methods [2], [3] use the normalized RGB color space, while some other methods [4], [5], [6] convert the color space from RGB to HSI (hue-saturation-intensity), which is based on human color perception. Both the two ways are effective in a certain extent, but fail on some extreme cases such as serious backlight or color cast.

Gangyi Wang and Guanghui Ren are with the School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, Hei longjiang, China (email: wogan1@gmail.com, rgh@hit.edu.cn).

This research was supported in part by the National Natural Science Foundation of China under Grant No. 61102085, and Natural Scientific Research Innovation oundation in Harbin Institute of Technology under Project No. HIT.NSRIF 201151.

TABLE I

SHAPES AND COLORS OF THE THREE CATEGORIES

Category	Color	Shape
Prohibitory	Red, white	Circular
Danger	Red, white	Triangular
Mandatory	Blue, white	Circular

Another popular class of detection methods is shape-based since traffic signs are usually with regular shapes. Hough transform is a commonly used method in detecting regular shapes such as line and circle. Some methods utilize Hough transform [7], [8] to detect circles or triangles in the extracted edge bitmap to locate potential traffic signs, but the memory and computational requirement is quite high for large images. Some other methods adopt genetic algorithm to detect circles and achieve high robustness to projective deformation [9], [10]. However, these methods are too slow for practical use.

In recent years, the sliding window scheme is popular in the object detection field. Two representatives are HOG (histogram of oriented gradient) [11] and VJ (Viola-Jones) [12]. Both the two approaches have been adopted in traffic sign detection and reported good results [13], [14], [15], [16].

However, as pointed out in [17], different methods are hard to compare because they all use different dataset. The GTSDB (German traffic sign detection benchmark) [18] announced recently provides an open platform for testing various traffic sign detection methods. We propose a coarse-to-fine method based on HOG, and have won the first place for the prohibitory and mandatory category, and the third place for the danger category. In this paper, we first give a description of the proposed method, then describe the training strategy and present experimental results. Finally, we make a conclusion of the proposed method.

II. THE COARSE-TO-FINE TRAFFIC SIGN DETECTION METHOD

The traffic signs included in the GTSDB dataset are classified into three categories: prohibitory, danger, and mandatory. Traffic signs within each category share the same color and shape, as listed in Table I.

The proposed method derives from the HOG algorithm [11], which has been proved to be very effective in detecting objects with specific shapes. The original HOG algorithm first extract HOG feature from a fix-sized window, then classify the feature with a linear SVM (support vector

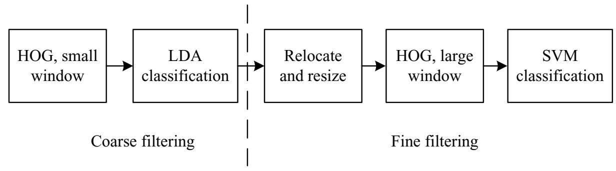

Fig. 2. Block diagram of the proposed method.

machine). The window slides throughout many scales of the original image to detect objects with different sizes. The size of the window is usually equal to or slightly larger than the minimum size of the object. However, traffic signs in some images are very small (as small as 16×16 pixels), and such size of window is hard to be classified accurately. To solve this problem, we propose a coarse-to-fine method, in which two sizes of windows are used, as shown in Fig. 2. The method first roughly finds out all the candidate region of interests (ROI) with a small-sized sliding window, which we call the coarse filtering; then further verifies the candidates with a large-sized window, which we call the fine filtering; finally, non-maximal suppression (NMS) is performed to suppress multiple nearby ROIs. As the original image is downscaled to many different sizes, we first relocate the ROIs output by the coarse filtering to the original image and then resize the ROIs to large-sized windows. In this way, more information from the original image can be preserved than directly resizing from the small-sized windows, which improves the classification accuracy of the large-sized windows. HOG feature is utilized by both the two filtering. The coarse filtering uses LDA for its efficiency, while the fine filtering uses SVM for its better accuracy.

Compared with the original HOG algorithm which uses a fix-sized sliding window, the proposed coarse-to-fine method has the following two advantages:

First, for the original HOG algorithm, the size of the window has to be very small to detect the smallest traffic signs, which on the other hand leads to low detection accuracy. In the proposed method, the small-sized window used in coarse filtering makes sure that the smallest traffic signs can be detected, and the large-sized window used in the fine filtering ensures high recall and precision.

Second, the coarse filtering is more efficient and the fine filtering is more accurate, combining the two filtering makes the proposed method more efficient than the original HOG algorithm while not losing accuracy.

A. The color HOG feature

As each category of traffic signs has a specific color, we use color information to increase the detection accuracy. The original HOG algorithm uses color information by calculating separate gradients for each color channel and taking the one with the largest gradient [11]. Another way to use color information is to concatenate the HOG feature for each color channel [14]. We present a third way which outperforms the former two methods. We calculate the histograms for each color channel and normalize the histograms of all the color channels together. In this way, we ensure that different color channels share the same normalization factor, which is the key improvement to the second way. Comparison of the three methods is given in section III.

B. Detection of the danger and mandatory signs

In the experiments, we get high recall ratios with the proposed method for all the three categories of traffic signs, but only achieve high precision ratio for the prohibitory signs. Therefore, we add some extra steps for the danger and mandatory signs.

The block diagram of danger sign detection is shown in Fig. 3. First, red pixels are extracted from a detected ROI and generate a red bitmap, then Hough transform is performed to find out the best equilateral triangle in the red bitmap, which is supposed to be the red rim of the danger sign. After that, the ROI is transformed to make the detected triangle after transformation equilateral, upright and in the center of the new ROI. Finally, the HOG and SVM classification is performed on the new ROI to determine whether it contains a danger sign. After this step, most false positives without an equilateral triangular red rim are filtered out, as shown in the last column of Fig. 4, leading to higher precision ratio. For more details about the equilateral triangle detection and the perspective adjustment methods, please refer to [19].

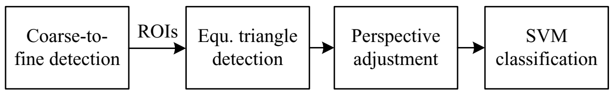

Fig. 3. Block diagram of danger sign detection.

Fig. 4. Examples of perspective adjustment. The three images in each column are the input ROI, extracted red bitmap and adjusted ROI, respectively. The last column is a false positive which can be filtered out after adjustment and reclassification.

For the mandatory signs, perspective adjustment is not applicable because the rotation degree cannot be estimated. Therefore, we train eight class-specific SVMs following the fine filtering step, as shown in Fig. 5. Each SVM corresponds to one class of mandatory sign, and determines whether a ROI belongs to the class or not. For each candidate ROI, HOG feature is extracted and sent to all of the eight SVMs. If any of the SVMs outputs positive response, the ROI is determined to be a true positive. In this way, most false positives are filtered out, including the blue and circular signs of similar appearance with mandatory signs.

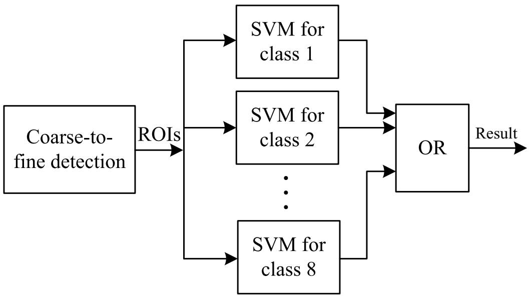

Fig. 5. Block diagram of mandatory sign detection.

III. EXPERIMENTS

We test the proposed method with the GTSDB dataset, and achieve good results for all the three categories of traffic signs. Details about the experiments are described in the following subsections.

A. Parameter selection

The GTSDB dataset consists of a training dataset and a test dataset, containing 600 and 300 images respectively. The sizes of the traffic signs in the images vary from 16×16 to 128×128. Therefore, we choose 20×20 as the size of the small-sized window, which is able to detect even the smallest traffic signs; and choose 40×40 as the size of the large-sized window, which is large enough to distinguish most of the traffic signs from false positives. To detect traffic signs with different sizes, the input image is downscaled in 22 steps using a scaling factor of 1.1. Although the scaling factor of 1.05 is more popular, we find that 1.1 is enough to achieve similar accuracy and is more efficient.

We have tried several sets of parameters for HOG, and choose the parameters giving the best recall ratio, which are $cell\_size = 4$, $num\_bins = 8$, $directional$ for the small-sized window, and $cell\_size = 8$, $num\_bins = 8$, $directional$ for the large-sized window. The HOG features of the small-sized windows are extracted in grayscale for efficiency, while those of the large-sized windows use color information. We compare the three methods of utilizing color information described in section II by classifying the extracted HOG features with IK-SVM (intersection kernel SVM), and find that the method in [14] and the proposed method achieve similar classification accuracy, both higher than the original method. However, the trained SVM model of the proposed method utilize significant less support vectors, which indicates the HOG feature extracted by the proposed method is more representative. The detailed results are listed in Table II.

We also compare the performance of different kernels of SVM, and find the intersection kernel gives the best accuracy. In addition, there is fast implementation for IK-SVM [20], which makes it nearly as fast as the linear SVM. Therefore, we choose the IK-SVM as our classifier.

TABLE II
COMPARISON OF THE THREE COLOR HOG METHODS (ACCURACY/NUMBER OF SVS)

Category	Original	Method in [14]	Proposed
Prohibitory	99.61%/876	99.82%/821	99.88%/748
Danger	99.79%/636	99.88%/1122	99.85%/565
Mandatory	99.27%/1089	99.81%/802	99.85%/684

B. The training strategy

To train the LDA classifier, we generate positive samples by randomly selecting traffic signs in the GTSDB training images and deforming them with randomly selected rotations and translations, which significantly improves the detection accuracy. Negative samples are randomly selected square windows which do not overlap with traffic signs. With this strategy, the trained LDA classifier has a high recall ratio and a low precision ratio, which is acceptable because the following fine filtering step is in charge of increasing the precision ratio.

The training process of SVM consists of multiple iterations: the first iteration is the same as the training process for the LDA classifier; then the trained classifier is applied to the training images, and the detected false positives together with the training samples of the first iteration are used to train a new SVM, which is the second iteration; the following iterations repeat the second interation until the trained SVM achieves the demanded recall and precision ratios or no more false positives are generated with training dataset. The whole process usually stops after four or five iterations.

For the mandatory signs, there are eight class-specified SVM to be trained. For each SVM, we use the mandatory signs belongs to the corresponding class in the training images as positive samples, and other classes of mandatory signs together with some hard false positives as negative samples. The SVM are easily trained to achieve the classification accuracy of 100%.

C. Detection results

Detection results of the proposed method for prohibitory signs are shown in Fig. 6. Different precision-recall pairs are obtained with different decision threshold of SVM. It can be seen that the proposed method achieves perfect detection result with both recall and precision equals to 1.

Fig. 7 shows the detection results for danger signs. Two curves are generated, representing the results of the methods with and without perspective adjustment. The method with adjustment achieves nearly perfect result, which is better than that without adjustment.

Fig. 8 shows the detection results for mandatory signs. There are also two curves, representing the results with and without class-specific SVMs. With the help of the class-specific SVMs, the proposed method achieves perfect result with both recall and precision equals to 1.

The method is robust to many kinds of adverse situations including bad lighting condition, partial occlusion, low qual-

Fig. 9. Some detection results by the proposed method.

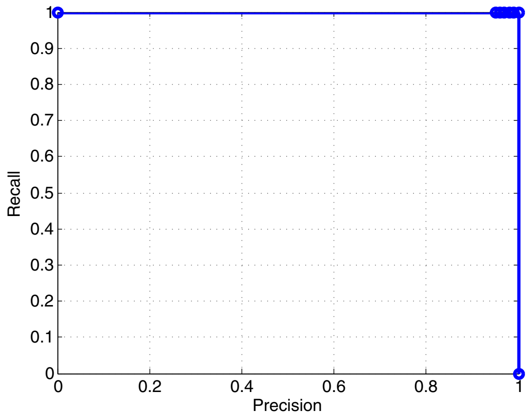

Fig. 6. Precision-recall curve for prohibitory sign detection by the proposed method.

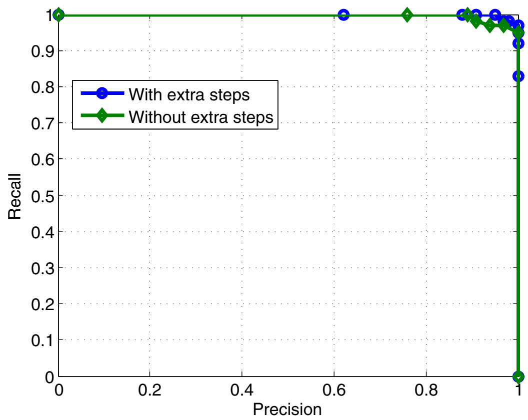

Fig. 7. Precision-recall curves for danger sign detection by the proposed method.

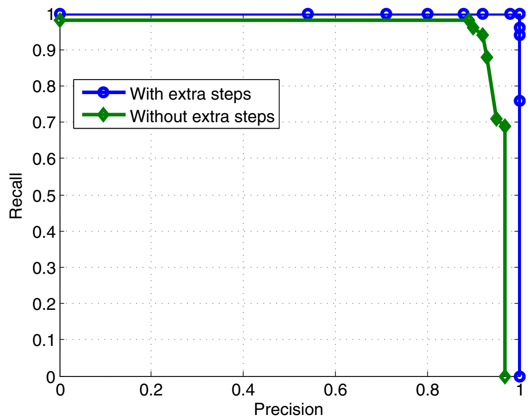

Fig. 8. Precision-recall curves for mandatory sign detection by the proposed method.

TABLE III

MEAN TIME REQUIRED FOR DETECTING THE THREE CATEGORIES

Category	Baseline (s)	With extra steps (s)
Prohibitory	1.122	-
Danger	1.148	1.179
Mandatory	1.140	1.232

ity and small projective deformation. Some examples are shown in Fig. 9.

D. Speed analysis

We test the proposed method on a hardware/software platform of Core I3 3.3 GHz, 4 GB DDR3, 64-bit Linux, and MATLAB 2011b. The matlab toolbox by Piotr Dollár [21] is employed for fast HOG feature extraction, and the toolbox by Subhransu Maji [20] is employed for fast IK-SVM classification. The mean time for detecting one image in each category is listed in Table III.

There is little difference between the processing time for the three categories with the baseline of the proposed method. After adding extra steps for the danger and mandatory signs, the processing time only increases by dozens of milliseconds, which indicates that the extra steps are not time consuming compared with the baseline method.

We have to say the speed is not enough for real-time application. However, as the code is written in MATLAB, we believe the speed could be improved if it is rewritten in C/C++. In addition, both the HOG and SVM could be accelerated with GPU, which could further improve the speed.

IV. Conclusions

In this paper, we present a state-of-the-art traffic sign detection method based on a coarse-to-fine sliding window scheme. The method uses HOG as feature descriptor, LDA and SVM as classifiers for coarse and fine filtering respectively. The method is robust to various lighting conditions, partial occlusion, low quality and small projective deformation. Experiments on the GTSDB dataset indicate that the proposed method achieves perfect results for the prohibitory and mandatory signs, and nearly perfect result for the danger signs.

In the future, we will focus on improving the efficiency of the proposed method, and make it suitable for real-time applications such as driver assistance.

References

[1] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition," *Neural Networks*, vol. 32, pp. 323–332, 2012.

[2] R. Janssen, W. Ritter, F. Stein, and S. Ott, "Hybrid approach for traffic sign recognition," in *Proceedings of the Intelligent Vehicles Symposium*, 1993, pp. 390–395.

[3] A. de la Escalera, L. E. Moreno, M. A. Salichs, and J. M. Armingol, "Road traffic sign detection and classification," *IEEE Transactions on Industrial Electronics*, vol. 44, no. 6, pp. 848–859, 1997.

[4] C. Fang, S. Chen, and C. Fuh, "Road-sign detection and tracking," *IEEE Transactions on Vehicular Technology*, vol. 52, no. 5, pp. 1329–1341, 2003.

[5] S. Maldonado-Bascon, S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gomez-Moreno, and F. Lopez-Ferreras, "Road-sign detection and recognition based on support vector machines," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 264–278, 2007.

[6] Y. Y. Nguwi and S. Y. Cho, "Emergent self-organizing feature map for recognizing road sign images," *Neural Computing & Applications*, vol. 19, no. 4, pp. 601–615, 2010.

[7] M. A. Garcia-Garrido, M. A. Sotelo, and E. Martin-Gorostiza, *Fast road sign detection using hough transform for assisted driving of road vehicles*, ser. Lecture Notes in Computer Science. Springer Berlin, Heidelberg, 2005, vol. 3643, pp. 543–548.

[8] N. Barnes, G. Loy, D. Shaw, and A. Robles-Kelly, "Regular polygon detection," in *Proceedings of the Tenth IEEE International Conference on Computer Vision*, 2005, pp. 778–785.

[9] Y. Aoyagi and T. Asakura, "A study on traffic sign recognition in scene image using genetic algorithms and neural networks," in *Proceedings of the 1996 IEEE IECON 22nd International Conference on Industrial Electronics, Control, and Instrumentation*, vol. 3, 1996, pp. 1838–1843.

[10] A. d. l. Escalera, J. Armingol, and M. Mata, "Traffic sign recognition and analysis for intelligent vehicles," *Image and Vision Computing*, vol. 21, no. 3, pp. 247–258, 2003.

[11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.

[12] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2001, pp. 511–518.

[13] Y. Xie, L.-f. Liu, C.-h. Li, and Y.-y. Qu, "Unifying visual saliency with hog feature learning for traffic sign detection," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, 2009, pp. 24–29.

[14] I. Creusen, R. Wijnhoven, E. Herbschleb, and P. d. With, "Color exploitation in hog-based traffic sign detection," in *Proceedings of the IEEE International Conference on Image Processing*, 2010, pp. 2669–2672.

[15] X. Baro, S. Escalera, J. Vitria, O. Pujol, and P. Radeva, "Traffic sign recognition using evolutionary adaboost detection and forest-ecoc classification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 1, pp. 113–126, 2009.

[16] C. G. Keller, C. Sprunk, C. Bahlmann, J. Giebel, and G. Baratoff, "Real-time recognition of u.s. speed signs," in *Proceedings of the Intelligent Vehicles Symposium*, 2008, pp. 518–523.

[17] A. Mogelmose, M. M. Trivedi, and T. B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: perspectives and survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1484–1497, 2012.

[18] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, and C. Igel, "Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark," in *International Joint Conference on Neural Networks (submitted)*, 2013.

[19] G. Wang, G. Ren, Z. Wu, Y. Zhao, and L. Jiang, "A hierarchical method for trafc sign classication with support vector machines," in *International Joint Conference on Neural Networks (submitted)*, 2013.

[20] S. Maji, A. C. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[21] P. Dollár, "Piotr's Image and Video Matlab Toolbox (PMT)," http://vision.ucsd.edu/ pdollar/toolbox/doc/index.html.