Traffic Sign Detection Based on Camera Imaging Apriority

Weiwei Wu, Yanyun Qu*, Xiaoqing Yang, and Li Lin

Computer Science Department, Xiamen University, 361005, P.R. China
{wuww0528,quyanyun}@gmail.com

Abstract. Traffic sign detection is very important to the vehicle intelligent auxiliary driving system and the driverless system. However, traffic sign detection is still a challenging problem, and there is not a satisfactory solution until now. In this paper, we aim at improving the speed and accuracy of traffic sign detection. In order to improve the detection speed, we use the image-forming principle to select the scale of sliding windows instead of the standard sliding window scheme. This operation will reduce the computational complexity from $O(N^4)$ to $O(N^2)$. In order to improve the detection accuracy, we adopt the hierarchical detection scheme. In the first stage, we use the cascade GentleAdaBoost classifier combined with the Haar-like features; in the second stage, we use the GentleAdaboost classifier combined with the multiple features fusing the color cues. The hierarchical detection scheme greatly reduces the false positive rate. We implement our approach on the Swedish Traffic Signs Dataset, the experimental results demonstrate our approach is effective and our approach could greatly reduce the false positive rate while keeping the detection rate.

Keywords: traffic sign detection, camera imaging apriority, multiple features, cascade classifier, hierarchical detection, sliding window scheme.

1 Introduction

With the quick increase of the amount of vehicles, the study of intelligent auxiliary driving system and driverless vehicles attracts more and more attentions. As we know, traffic signs give important information of road condition. In this paper, we aim at improving the speed and accuracy of traffic sign detection. The driver could be reminded to pay attention to the road condition by the traffic-sign-detection system and avoid the dangers.

However, traffic sign detection is still a challenging problem. There are two difficulties: 1) the searching scheme which can achieve real-time detection; 2) the design of the detector, which can accurately detect traffic signs. The sliding window scheme is widely used in object detection. However, it is time-consuming, because it needs to search the object at all locations and scales. Its computational complexity is $O(N^4)$, where N is the width of an image. In the aspect

* Yanyun Qu is the corresponding author.

B. Huet et al. (Eds.): PCM 2013, LNCS 8294, pp. 720–729, 2013.

of the detector design, it is very difficult to design a perfect detector, because the appearances of traffic signs vary with the changes of illumination, weather condition, and viewpoint and so on.

In order to overcome the two difficulties, many researchers discussed the problem of traffic sign detection. Karla [1] gave an overview of the methods of traffic sign detection, and he divided these methods into three classes: the color-based methods, the shape-based methods and the machine learning based methods. In this paper, we are concerned about machine learning based methods. As we mentioned above, two critical factors greatly influence the performance of traffic sign detection: the detector design and the searching scheme. Xie [2]used HOG [3] to represent traffic signs, and design a SVM classifier to distinguish traffic signs from non-traffic-sign object. Karla [1] used Haar-like features [4] to represent traffic signs and designed an Adaboost classifier to determine the traffic signs. In the aspect of the searching scheme, the methods can be divided into two classes: the sliding window methods and the saliency methods. As we know, the sliding window strategy is time consuming. The saliency methods solve the problem by mimicking visual attention. Thresholding based methods [5,6,7] can be regarded as belonging to saliency methods. Kamada [5] proposed using thresholding method in RGB space, and Ohta [6] improved this method. Gavrila [7] used shape information to detect traffic signs. However, saliency methods are not reliable.

In this paper, motivated by Hoiem [8], who proposed to put objects into perspective and model the relationship between the scale and location variances in an image, we use the camera imaging principle to determine the scale of the sliding windows at a given position so as to reduce the computational cost of the sliding window strategy. We derive the interplay among the traffic signs in the real world, the camera and the projection of the traffic signs in an image. Furthermore, we design a hierarchical detection scheme to filter the non-traffic-sign regions. In the first stage, we use the cascade GentleAdaBoost classifier combined with the Haar-like features to find the candidate traffic-sign regions; in the second stage, we use the GentleAdaboost classifier combined with the multiple features fusing color cues to filter the false positive regions. The hierarchical detection scheme could greatly reduce the false positive rate. Therefore, the contributions of our work are as follows: 1) using camera imaging apriority to select the scale of sliding windows; 2) designing a hierarchical detection strategy to distinguish traffic-sign regions from non-trafffic-sign regions.

The rest of this paper is organized as follows: Section 2 briefly introduces camera image apriority. Section 3 introduces the framework of our method, and Section 4 details the implementation of our approach. The experimental results are shown in Sections 5. In section 6 we give the conclusions.

2 Camera Imaging Apriority of Traffic Signs

Hoiem [8] proposed putting local objects in the context of the overall $3D$ scene. He modeled the interdependence of objects, surface orientations and camera

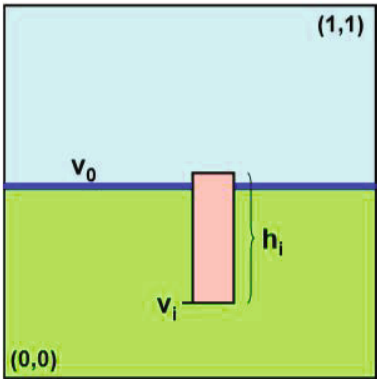


Fig. 1. The illustration of the parameters used in camera imaging principle

viewpoints. In this paper, we would use his method to derive the project information of traffic signs in an image in order to determine its scale in an image.

Now we give a brief introduction of the camera imaging principle. We assume that objects stand upright on the ground plane. In our approach, we regard a traffic sign and its stick as a whole. If we know the horizon line and the camera height, we will derive the relationship between the object's world height y and the object image height using the following notation: world coordinates (x, y, z) with y being height and z being depth which is the distance between the object and the camera; camera tilt θ_x and focal length f; camera height y_c; pixel coordinate (u, v) in an image ranging from $(0, 0)$ at the bottom-left and $(1, 1)$ at the top-right; camera optical center (u_c, v_c); the horizon position v_0 as the vanishing line of the ground plane in image coordinate and we assume that the ground plane is at $y = 0$ and its location in an image is v_0. According to the camera imaging principle, the camera tilt angle can be approximately computed as $\theta_x \approx 2arctan\frac{v_c - v_0}{2f}$. Furthermore, the relationship between the pixel coordinates (u, v) and the object's world coordinates (x, y, z) is given

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{z} \begin{bmatrix} f & 0 & u_c \\ 0 & f & v_c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & cos\theta_x & -sin\theta_x & y_c \\ 0 & sin\theta_x & cos\theta_x & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ x \\ 1 \end{bmatrix} \tag{1}$$

Inferring Equ. (1), we can obtain the object's height y:

$$y = \frac{z(fsin\theta_x - (v_c - v)cos\theta_x) - fy_c}{(v_c - v)sin\theta_x + fsin\theta_x} \tag{2}$$

We let v_t and v_b denote the heights of the top and bottom of the object in an image. When the height of the object is zero, that is, $y = 0$, from Equ. (2), we can solve the object depth z as

$$z = \frac{fy_c}{fsin\theta_x - (v_c - v_b cos\theta_x)}.$$

We assume that the camera tilt is small, that is,

$$cos\theta_x = 1, sin\theta_x \approx \theta_x, \theta_x \approx \frac{v_c - v_c}{f};$$

the object's world height y can be inferred as

$$y = \frac{y_c \frac{v_t - v_b}{v_0 - v_b}}{1 + \frac{(v_c - v_0)(v_c - v_i)}{f^2}}$$

As we assumed before, the horizon position is near the camera optical center, that is

$$\frac{(v_c - v_0)(v_c - v_i)}{f^2} \approx 0,$$

so we can simplify Equ. (2) as follows,

$$y = y_c \frac{v_t - v_b}{v_0 - v_b}. \tag{3}$$

In this paper, we modify Equ. (3) as follows,

$$y_i = k y_c \frac{h_i}{v_0 - v_b} \tag{4}$$

where the h_i, y_i, v_i are the object's image height, world height, and bottom position respectively of object i. Thus, we can solve for object image height h_i

$$h_i = \frac{(v_0 - v_b) y_i}{k y_c} \tag{5}$$

where the threshold $k = 1.5$. The explanation of the parameters used in this section is shown in Figure 1. As mentioned before, we regard a traffic sign and its stick as a whole, so h_i represents the height of the traffic-sign object. In the real world, the size of traffic signs is about $0.5m \times 0.5m$ and the possible height of the whole is $1.0m, 2.5m, 4.0m$. So from Equ. (5), we can limit the size of the sliding windows as follows,

$$\alpha \frac{(v_0 - v_b) y_i}{k y_c} - \varepsilon < h_i < \alpha \frac{(v_0 - v_b) y_i}{k y_c} + \varepsilon \tag{6}$$

3 Framework

Figure 2 shows the framework of our approach. For a query image, we firstly sample the patches by the sliding window scheme which selects the scale of sliding windows based on the camera imaging apriority. After that, we determine if a sampled patch belongs to the traffic-sign class by a cascade binary classifier in which each classifier is an Adaboost classifier with Haar-like features. Furthermore, we filter the candidate regions by the second-level Adaboost classifier with multiple features fusing the color cues. The second level operation can effectively reduce the false positive rate.

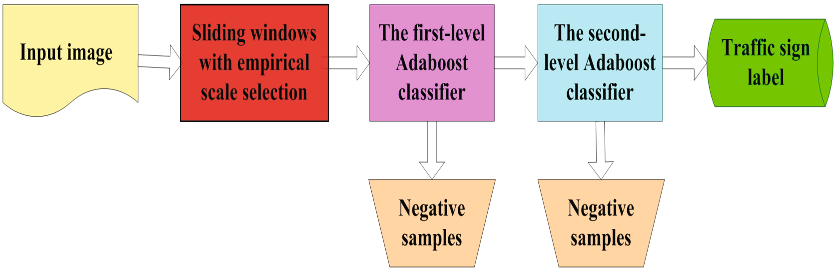


Fig. 2. The framework of our approach

3.1 Implementation Details

In our approach, we first select the scale of the sliding windows via camera imaging apriority. As shown in Figure 3, we show an example how the sliding windows vary with the location. The camera imaging apriori knowledge implies that the size of traffic-sign objects is among a range. In our experiment, the width of sliding windows is selected between 20 pixels and 300 pixels.

In the first stage, we use the cascade GentleAdaBoost classifier [9] combined with the Haar-like features [4]. Figure 4 shows the flowchart of the cascade classifier. This classifier has a high detection rate and the lower false detection rate. Consider that the results obtained from the first level detection contain many false positive patches, we implement the second level classifier to refine the traffic sign regions. In the second stage, we use the GentleAdaboost classifier combined with the multiple features fusing the color cues and it was shown effective in traffic sign detection and recognition in our previous work [10].

In this section, we design the color fused feature which is concatenated by HOG, LBP and Hue histogram. The HOG (Histogram of Oriented Gradients) was proposed by Dalal [3] for pedestrian detection. In this paper, we normalize

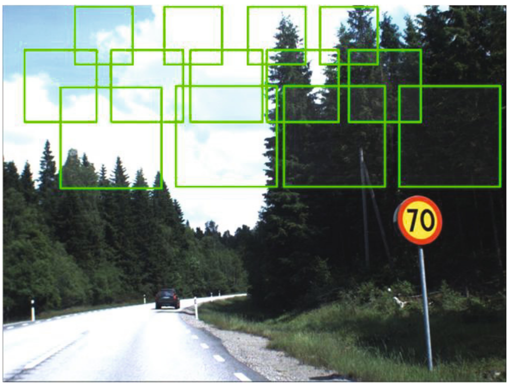


Fig. 3. The scales of sliding windows based on camera imaging apriority

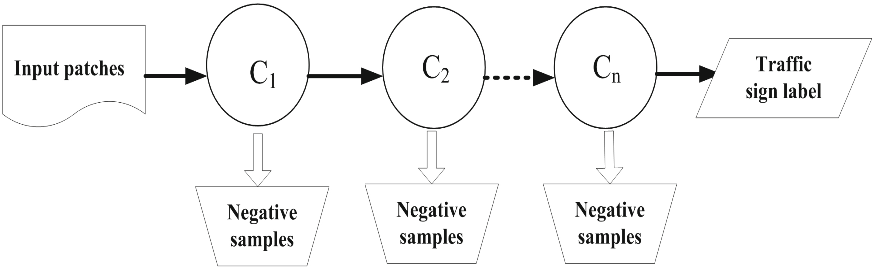


Fig. 4. The cascade classifier

a traffic-sign window to 40×40. Then, it is divided into some block and each block is divided into some cells. In our experiment, we use 49 blocks, each block includes 2×2 cells, the size of a cell is 5×5 , in which a gradient histogram of 9 bins is obtained. Thus,we can obtain a HOG feature vector with 1764 dimensions.

LBP feature is proposed by Ojala [11], and it is a popular texture descriptor. The basic operator assigns a label to every pixel of an image by thresholding its 3×3-neighborhood with the center pixel value and considering the result as a binary number. Then the histogram of the labels can be used as a descriptor. In our approach, we normalized the labeled patch to 40×40. The normalized patch is tiled by a 4×4 grid. The descriptors are then concatenated to form a global description of the labeled region.

Both of HOG features and LBP features neglect the color cues, but the color cues are important to traffic sign detection. So we use the HOG and LBP features fusing the color cues. Our previous work has shown that the concatenation of HOG features, LBP features and Hue histogram can achieve better detection performance. Figure 5 shows the multiple features which fuse the hue histogram. In our approach, we convert images from RGB to HSV, and the Hue histogram is obtained by dividing Hue to 256 bins. Thus, the color descriptor is a vector of 256 dimensions.

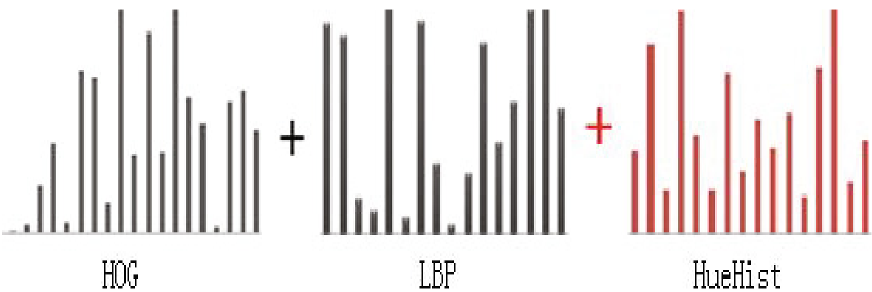
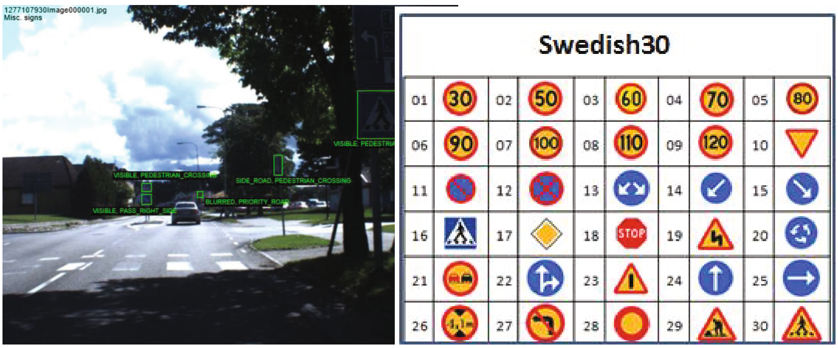


Fig. 5. The color cue fused multiple features

4 Experimental Results

In this section, we implement our approach on the STSD dataset (Swedish Traffic Signs Dataset). The STSD contains more than 20000 pictures and the size of each picture is 1986 × 960. In our experiments, we use 1970 labeled pictures from the STSD as our testing set. From the remaining pictures, we construct the Swedish30 set as our training set which contains 3192 positive samples belonging to 30 classes of traffic signs. Figure 6 a) shows a typical labeled picture in the STSD, and Figure 6 b) shows the 30 classes of traffic signs in Swedish30.



a) b)

Fig. 6. An example and the traffic signs of STSD dataset. a) An example of STSD dataset, b) 30 classes of traffic signs in STSD dataset.

In the first experiment, we evaluate how the camera imaging apriority is reasonable to traffic sign detection. We select 100 pictures from the testing set randomly. The 100 pictures contain 159 traffic signs. And we use the camera imaging apriority to sample. In the samples, there are 156 traffic signs. The result demonstrates that the camera imaging apriori information could keep about 98% of traffic signs detected. Figure 7 shows some regions generated by camera imaging apriority.

In the second experiment, we evaluate how the camera imaging apriority reduce the computational complexity. In our experiments, the size of sliding windows ranges from 20 to 300 pixels, the size of a sliding window increases by 10 pixels per iterating step and a sliding window moves at 5 pixels per time. We compare the standard sliding window scheme with our searching scheme. Both of them use the GentleAdaboost classifier combined the multiple features fused with the color cues to filter the sampled patches. Table 1 shows the comparison results in a query image. It demonstrates that our approach could reduce the computational cost theoretically from $O(N^4)$ to $O(N^2)$. The real consuming time could be reduced from 40000 seconds to 50 seconds in a query image.

In the last experiment, we evaluate the performance of our hierarchical detection method. We implement our detection approach in the testing set and

a) b)

Fig. 7. The reasonableness of scale selection via the camera imaging apriority. a) A query image; b) The sampled regions based on camera imaging apriority.

Table 1. The comparison of two searching schemes

	Standard sliding window scheme	Our approach
Computation complexity	$O(N^4)$	$O(N^2)$
Time(s)	40000	50

Table 2. The comparison in terms of the false detection rate (%)between the first level and second level detections

Scales of windows	10 × 10	20 × 20	30 × 30	40 × 40	50 × 50	70 × 70	100 × 100
The first level	99.23	97.73	92.79	89.46	87.47	89.36	81.03
The second level	88.18	73.34	53.31	43.39	36.94	37.42	26.66

Table 3. The comparison in terms of the detection accuracy (%)between the first level and second level detections

Scales of windows	10 × 10	20 × 20	30 × 30	40 × 40	50 × 50	70 × 70	100 × 100
The first level	89.4	90.69	90.47	88.66	86.96	81.52	69.23
The second level	89.3	90.66	90.46	88.62	86.95	81.52	69.23

we use the detection accuracy and the false detection rate as our criterion. The detection accuracy is the ratio of the number of correctly detected regions and the number of ground truth regions. A region is called to be correctly detected if the ratio of the intersection area and the union area between the detected region and the ground truth region is greater than 50%. The false detection rate is the ratio of the number of false positive samples and the total number of detected regions. Considering that the pictures in STSD are complex and the scales of traffic signs change greatly, we count our results according to scales of traffic signs in the testing set. As shown in Tables 2 and 3, the results in the column of 10×10 (the false positive rate or the detection accuracy) are obtained on all the samples whose size is greater than 10×10 and the results in other columns are explained in the same way. Table 2 shows the false detection rates. We can see that the second level detection can effectively reduce the false positive rate. In the aspect of the detected samples, the false detection rate is only 26.6%, which reduces about 55% compared with the first level detection. Moreover, Table 3 shows that the detection accuracy after the second level detection is almost unchangeable compared with the first level detection. Figure 8 shows the visual results of the first level and second level detections. We can see that most of the obtained ROI from the first level detection can be filtered in the second level detection. Thus, the proposed hierarchical detection method could effectively reduce the false detection rate while keeping the detection rate.

a) b)

Fig. 8. The comparison of the first level detection and the second level detection. a) The result obtained by the first level detection, b) The result obtained by the second level detection.

5 Conclusions

In this paper, we design a prototype system of traffic sign detection. The experimental results have demonstrated that the camera imaging apriority is effective for reducing the computational complexity of the sliding window searching, and the hierarchical detection scheme has a good performance in traffic sign detection, which greatly reduces the false positive rate. In the future, we would optimize the algorithm in order to implement it on real videos in real time.

Acknowledgments. This research work was supported by the National Natural Science Foundation of China Under Grant No. 61373077, the Natural Science Foundation of Fujian Province of China Under Grant No. 2013J01257, Research Funds for the Central Universities Under Grant No.2010121067 and the 2013 national college students' innovative and entrepreneurial training project.

References

1. Brkic, K.: An overview of traffic sign detection methods (2010)
2. Xie, Y., Liu, L., Li, C., Qu, Y.: Unifying Visual Saliency with HOG Feature Learning for Traffic Sign Detection. In: IVS 2009, pp. 24–29 (2009)
3. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 886–893 (2005)
4. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on CVPR 2001, vol. 1, pp. I-511–I-518 (2001)
5. Kamada, H., Naoi, S., Gotoh, T.: A compact navigation system using image processing and fuzzy control. In: Proceedings of the Southeastcon 1990, pp. 337–342. IEEE (1990)
6. Ohta, Y.I., Kanade, T., Sakai, T.: Color information for region segmentation. Computer Graphics and Image Processing 13, 222–241 (1980)
7. Gavrila, D.M., Philomin, V.: Real-time object detection for "smart" vehicles. In: The Proceedings of the Seventh IEEE International Conference on Computer Vision, pp. 87–93 (1999)
8. Hoiem, D., Efros, A.A., Hebert, M.: Putting objects in perspective. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 2137–2144 (2006)
9. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). The Annals of Statistics 28, 337–407 (2000)
10. Yang, X., Qu, Y., Fang, S.: Color fused multiple features for traffic sign recognition. In: Proceedings of the 4th International Conference on Internet Multimedia Computing and Service, pp. 84–87 (2012)
11. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern Recognition 29, 51–59 (1996)