2014 IEEE 17th International Conference on
Intelligent Transportation Systems (ITSC)
October 8-11, 2014. Qingdao, China

LIDAR and Vision-Based Real-Time Traffic Sign Detection and Recognition Algorithm for Intelligent Vehicle

Lipu Zhou, Zhidong Deng*

Abstract—**Real-time traffic sign detection and recognition are essential and challenging tasks for intelligent vehicle. The previous works mainly focus on detecting and recognizing traffic sign based on images captured by onboard camera. Visual features of traffic sign such as color, shape, and appearance, however, are often sensitive to illumination condition, angle of view, etc. Except for camera, LIDAR also provides important and alternative features of traffic sign. The fusion of complementary data acquired from both sensors can improve the robustness of the algorithm, especially when data from either of them are of low quality. For this reason, we propose a new traffic sign detection and recognition algorithm based on the fusion of camera and LIDAR data. Specifically, position prior, color, laser reflectivity, and 3D geometric features are integrated to detect traffic sign in a 3D space. In most of previous works, different colors of traffic signs are individually handled in a specific color space, which generally results in the use of many thresholds or multiple classifiers. In this paper, we use combined color spaces (CCS) such that traffic sign colors can be entirely treated as one class. For traffic sign recognition, in order to improve the robustness to any viewpoint variation, regions of interest (ROIs), which suffer from perspective deformation, are rectified first by fusing LIDAR and camera data. Then the histogram of oriented gradient (HOG) feature and linear support vector machines (SVMs) are used to classify traffic signs. Finally, extensive experimental results in challenging conditions show that our algorithm is real-time and robust.**

I. Introduction

Traffic signs are designed to inform drivers of the current road condition and the traffic rules that must be complied with. Real-time traffic sign detection and recognition are important for the intelligent vehicle to safely navigate in the urban environment. Currently, intelligent vehicles are equipped with multiple sensors to perceive the environment. Two common categories of them are camera and LIDAR. Camera captures color, shape and appearance of the traffic sign. Substantial vision-based algorithms have been proposed to detect and recognize the traffic sign. However, this problem is still open and has many challenges. Some challenges are caused by the vulnerability and incompleteness of the camera data. In general, the camera data are sensitive to illumination condition, angle of view, etc. Furthermore, the 3D geometric information is lost, and hard to be fast and robustly recovered from the images acquired by a common perspective camera [1]. 3D information of the traffic sign is needed for the intelligent vehicle to decide which traffic signs should be responded to, and how to respond to. Fortunately, LIDAR is capable of providing such information. But it cannot acquire the visual information of the traffic sign. Therefore, fusing the measurements of both sensors gives a highly promising direction for traffic sign detection and recognition in complex environments. To the best of our knowledge, however, few related works have been reported so far.

In this paper, we propose a novel LIDAR and vision-based traffic sign detection and recognition algorithm. For the detection, unlike most of pervious works which search traffic signs in the image, we detect traffic signs in the 3D space. The position of the traffic sign is restricted. Some vision-based algorithms [2] use this fact to generate the ROI of the traffic sign in the 2D image plane. Since the camera data lose 3D information, the ROIs obtained from [2] are generally too rough to effectively eliminate the background where the traffic sign will not appear. In this paper, we employ the LIDAR data to obtain a much tighter ROI. Laser points in the ROI are projected into the image plane to obtain the corresponding colors. In the literature, color-based methods [3] generally choose a certain color pace to separately identify different traffic sign colors, which results in many thresholds or multiple classifiers. For this problem, we use combined color spaces (CCS) that consist of RGB, HSV, and CIE $L^*a^*b^*$ color spaces to represent color, and treat different traffic sign colors as one class. Then color is combined with laser reflectivity, planarity, physical size and aspect ratio to detect the traffic sign in the 3D space. Once all potential candidates are located, the recognition module is triggered. To make the recognition algorithm robust to viewpoint variation, perspective distortion of the candidate is rectified first. Perspective rectification has been studied by many researchers [4, 5]. In these works, the rectification is performed by two steps, i.e., analyzing the shape of the traffic sign, and then establishing the association between the detected feature points and the control points in the reference shape. The drawback of these algorithms is that both shape recognition and feature point detection are not robust to large perspective distortion and occlusion. For this reason, we synthesize the fronto-parallel view of the ROI by viewing transformation based on fusing LIDAR and camera data. This algorithm does not need to identify the shape of the ROI, and not need to detect the feature point either. After rectification, traffic sign classification is performed through linear SVMs trained by the HOG feature [6]. The overview of our algorithm is summarized in Fig. 1. By evaluating the proposed algorithm in challenging environments, we show that our algorithm is robust to poor illumination condition, motion blur, significant perspective deformation and partial occlusion. Besides, the average runtime of the whole system is only 33.25 *ms* per frame on a mainstream computer.

This work was supported in part by the National Science Foundation of China (NSFC) under Grant Nos. 90820305 and 60775040 and by the National High-Tech R & D Program of China under Grant No. 2012AA041402.

The authors are with State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China. (email: zlp09@mails.tsinghua.edu.cn, michael@tsinghua.edu.cn)

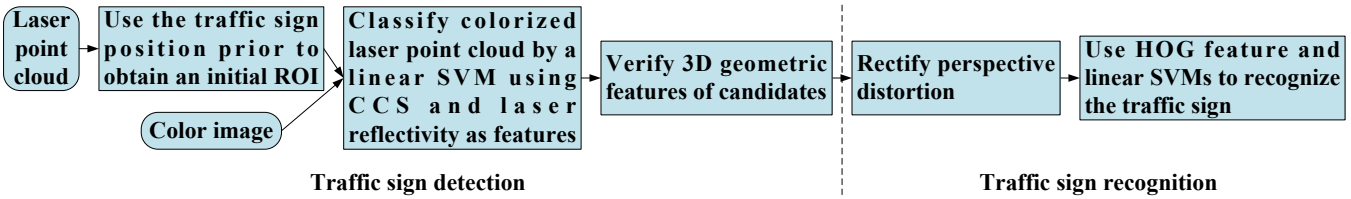


Fig. 1. Overview of the proposed algorithm.

II. Related Works

In the literature, vision-based algorithm for traffic sign detection and recognition is predominant. Extensive studies have been devoted to this field. For traffic sign detection, Møgelmose *et al.* [7] gave a detailed review of the algorithms before 2012. Recently, the competition on the German Traffic Sign Detection Benchmark (GTSDB) [8] enables the comparison of different traffic sign detection algorithms. For the GTSDB, several algorithms achieve high accuracy [9-11], but their computational costs are expensive. For example, Mathias *et al.* [10] used the integral channel features to train three cascade detectors, each for one traffic sign subcategory in the GTSDB. The speed of this algorithm is about 0.35 Hz, even with the help of GPU. The running time of this algorithm makes it unsuitable for the real-time application.

For traffic sign recognition, German Traffic Sign Recognition Benchmark (GTSRB) [12] is a widely used dataset. So far, convolutional neural network (CNN) based methods have achieved excellent results on this dataset. In [13], multi-column deep neural networks were exploited for traffic sign classification. The recognition rate of [13] achieves 99.46%. Recently, Jin *et al.* [14] proposed a hinge loss stochastic gradient descent method to train CNNs, and further improved the record on the GTSRB to 99.65%. However, these algorithms are very computationally demanding.

As mentioned above, LIDAR also provides important information of the traffic sign. However, the number of published works on this aspect is relatively small. Traffic signs are painted with highly reflective material which can generate high laser reflectivity measurement. This characteristic is used by [15, 16] to detect the traffic sign. However, laser reflectivity is not only related to the material, but also affected by the object pose. Therefore, it is not robust that only using this feature to determine whether a laser point belongs to a traffic sign or not.

III. Traffic Sign Detection

A. Fusing Camera and LIDAR Data

Let us consider a rigidly fixed LIDAR and camera pair. According to the common pinhole model, the relationship between a laser point $\mathbf{P}_l$ and its corresponding image $\mathbf{p}_u$ can be described below [1]:

$$\tilde{\mathbf{p}}_u = s\mathbf{K}\left(\mathbf{R}\mathbf{P}_l + \mathbf{t}\right), \tilde{\mathbf{p}}_u = \begin{bmatrix} \mathbf{p}_u \\ 1 \end{bmatrix}, \quad (1)$$

where $\mathbf{K}$ is the camera intrinsic matrix, $\mathbf{R}$ and $\mathbf{t}$ are the rotation matrix and the translation vector from the LIDAR coordinate system to the camera coordinate system, s is a scalar.

The position of the traffic sign provides a powerful prior to narrow down the search region of the traffic sign. In the 3D space, the traffic sign position with respect to the road is well-established. Normally, they are aligned along the roadside or above the road. The accurate road model can be obtained from a vision-based lane marking detection system. However, as the road is generally a straight line or a smooth curve with low curvature, the search region of the traffic sign can be simply constrained in a cuboid aligned along the road. Laser points in the cuboid are projected into the image plane by (1). After this process, each laser point obtains the corresponding color.

B. Classification of the Colorized Laser Point Cloud

Traffic signs have specific colors. But the illumination condition, weather such as fog, shadow, and sign aging all impact on the captured color. In addition, the similar color from other objects may result in false alarms. Fig. 2 (a) gives an example in which color is seriously distorted by the poor illumination condition. We use the thresholds with respect to the normalized RGB color space given by [3] to seek the red and blue pixels in Fig. 2 (a). The result is shown in Fig. 2 (b). As we can see, color is not effective to segment the traffic signs in this case. Besides, many false alarms are introduced. On the other hand, traffic signs are painted with highly reflective material which can make its laser reflectivity value larger than the surrounding objects [15, 16], as demonstrated in Fig. 2 (c). In the above case, although color is distorted, laser reflectivity provides an effective cue to find the traffic sign. However, laser reflectivity is related not only to the material, but also to the object pose. When the traffic sign orientation has large angle with respect to the incoming laser beam, the laser reflectivity value may decease so as to lose discriminability. Fig. 3 (c) gives such an example. But, in this case, color provides a valid cue to extract the traffic sign, as illustrated in Fig. 3 (b). Although either color or laser reflectivity has its limitation alone, the fusion of both can yield robust and discriminative features to detect traffic signs.

The utility of the reflectivity is straightforward. However, it is a bit difficult to describe the color feature. This is because there are several color spaces, and traffic signs have different colors including red, green, yellow, blue, white and black. Basically, the previous works separately handle these colors in a specific color space [3, 7, 11]. This will result in many thresholds or multiple classifiers. Therefore, it is more preferable to identify them by a single classifier. However, these colors are very different. Forcing them to be one class and representing them in a single color space will result in bad classification performance. Intuitively, if one color space can be used to identify one or a few of the traffic sign colors, then several color spaces can be combined to identify all of them. For this reason, we use Combined Color Spaces (CCS), consisting of RGB, HSV, and CIE $L^*a^*b^*$ color spaces, to represent the color. Thus the resulting feature vector for each laser point has 10 components, i.e. 1 laser reflectivity component and 9 color components. A linear SVM is trained to

classify the colorized laser points. The classification results of the laser points in Fig. 2 and Fig. 3 are illustrated in Fig. 4. The red points represent laser points classified as belonging to the traffic sign. We can see that the results include some false alarms. But compared to the false alarms in Fig. 2 (b) and Fig. 3 (b), our results are much better. In the following section, we will discuss how to eliminate such errors by the 3D geometric features of the traffic sign.

(a) (b) (c)

Fig. 2. An example to demonstrate the situation that color is seriously distorted, but laser reflectivity provides useful cue to detect the traffic sign. (a) The color image. (b) The results of red and blue pixels detection. (c) Laser point cloud rendered by the laser reflectivity. In (b), the red and blue pixels are detected by the threshold with respect to the normalized RGB color space given by [3].

(a) (b) (c)

Fig. 3. An example to demonstrate the situation that laser reflectivity loses discriminability, but color is effective to identify the traffic sign. (a) The color image. (b) The results of red and blue pixels detection. (c) Laser point cloud rendered by the laser reflectivity. In (b), the red and blue pixels are detected by the same algorithm used in Fig. 2 (b).

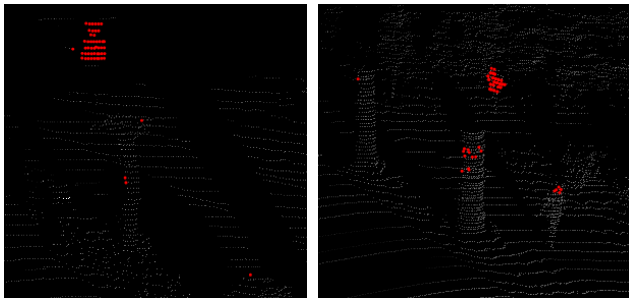


Fig. 4. The classification results of the laser points in Fig. 2 (c) and Fig. 3 (c).

C. 3D Geometric Feature Verification

Assume N laser points are classified to belong to the traffic sign. Let S denote this set. Before further analysis, we segment S according to the Euclidean distance between laser points. Suppose S can be divided into M partitions such that

$$S = \bigcup_{i=1}^{M} S_i = \bigcup_{i=1}^{M} \left\{ \mathbf{p}_{ij}^{l} \right\}_{j=1\ldots N_i}, \tag{2}$$

where S_i is the i-th partition of S, which consists of N_i points. Since the surface of the traffic sign is a plane, we should check to what degree each S_i can be approximated by a plane. One straightforward way is to fit a plane to each S_i, then check the residual of the fitting to determine whether S_i is a plane or not. However, as S_i may be damaged or contain points from other objects, such as leaves overlaying on the traffic sign, simply fitting a plane to the entire S_i may result in eliminating the correct region. Furthermore, even all the points in S_i are exactly from a traffic sign, points with large noises will impact on the accuracy of the plane parameters. The accurate plane parameters are essential for the subsequent 3D geometric feature verification and perspective rectification. Based on the foregoing reasons, we need to robustly determine the optimal planar part P_i of S_i. The RANSAC algorithm [17] is applied for this purpose. Assume P_i contains M_i points. The ratio between M_i and N_i is a fair indicator to evaluate the planarity of S_i. A region will be discarded if this ratio is less than a threshold (0.6 in our experiment) or N_i is small. Fig. 5 shows the result of Fig. 4 after this step. We can see that false alarms and laser points on traffic signs with large noises are successfully eliminated.

After the foregoing step, non-planar point sets are discarded. But the planar object that has the similar color and laser reflectivity value to the traffic sign cannot be removed, as the bus surface shown in Fig. 6. Size and aspect ratio of the traffic sign are two import cues to filter such kind of false alarm. Both quantities are widely used in the previous vision-based algorithms [5, 18]. However, the ranges of the possible size and aspect ratio of the traffic sign image are significantly broad. In GTSRB, this parameter ranges from 15×15 to 222×193 pixels [19]. On the other hand, the aspect ratio of the traffic sign image is affected by the pose of the traffic sign relative to the camera, thus it can also change widely. However, size and aspect ratio of the traffic sign are factually much more restricted in the 3D space. Therefore, inspecting them in the 3D space is more reasonable.

In order to check size and aspect ratio of the remaining regions, points in P_i are orthogonally projected into the estimated plane. Denote Q_i the corresponding projection of P_i. Then we calculate the bounding box B_i of Q_i, as illustrated by the blue box in Fig. 6 (b). The size and the aspect ratio of B_i are used to further eliminate false alarms. The limits of both parameters should be specified to ensure the detection algorithm has high recall even in extreme conditions, such as the serious occlusion shown in Fig. 9(c). Denote the width and the height of B_i by w_i and h_i. In our experiment, we keep the B_i whose $\max(w_i, h_i)$ is delimited between $0.12m$ and $1.2m$, and $\max(w_i, h_i)/\min(w_i, h_i)$ is not greater than 3.2 when S_i is fully within the LIDAR's field of view or 4 when S_i is partially within the LIDAR's field of view. After this step, the false alarm caused by the bus surface in Fig. 6 (b) is successfully eliminated, as shown in Fig. 6 (c). Regions passing the 3D geometric feature verification will be analyzed by the following recognition module.

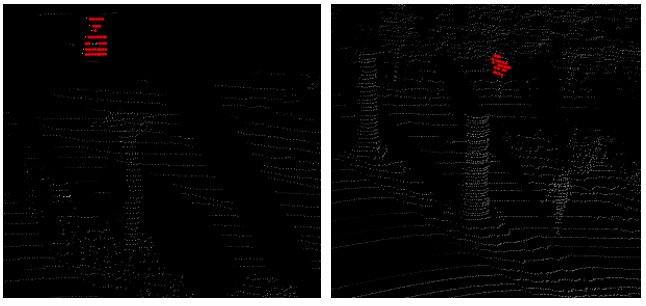
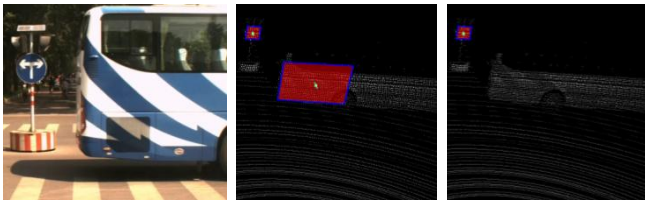


Fig. 5. The results of eliminating the false alarms in Fig. 4 by planarity.



(a) (b) (c)

Fig. 6. An example of eliminating the planar object that has similar color and laser reflectivity to the traffic sign through verifying size and aspect ratio of the ROI. (a) The color image. (b) The bounding boxes of the ROIs. (c) The left ROI after size and aspect ratio verification

IV. Traffic Sign Recognition

A. Perspective Distortion Rectification

Some captured traffic sign images are subjected to serious perspective distortion, as shown in Fig. 3 (a). In this paper, we correct such distortion based on the algorithm [20], which is a generic algorithm to rectify the perspective distortion of the planar object. For our problem, we present a specialized version of such generic algorithm.

In order to rectify the perspective distortion of the traffic sign, we introduce a virtual camera C_v that look orthogonally at the sign. Assume the centroid of the laser points on the traffic sign is $\bar{\mathbf{P}}^l$, the parameters of the traffic sign plane $\mathbf{\Pi}$ in the LIDAR coordinate system are $\boldsymbol{\pi}^l = [(\mathbf{n}^l)^T, d^l]^T$, and the orientation and translation of C_v relative to the LIDAR are $\mathbf{R}_v^l$ and $\mathbf{t}_v^l$, respectively. We place the virtual camera k meters in front of the traffic sign, along the ray passing $\bar{\mathbf{P}}^l$ with the direction the same as the traffic sign unitary normal $\mathbf{n}^l$, i.e. the z-axis of C_v is in the direction $-\mathbf{n}^l$. Generally, traffic signs are approximately upright along the road. So we can use the normal vector of the ground to define the up-vector of C_v. Since the xy-plane of our 3D LIDAR approximately parallels the ground, we chose $\mathbf{r}_y = [0,0,-1]^T$ as the up-vector of C_v. Based on the above description, $\mathbf{R}_v^l$ and $\mathbf{t}_v^l$ have the form

$$\mathbf{R}_v^l = \left[{}^1\mathbf{r}_v^l, {}^2\mathbf{r}_v^l, {}^3\mathbf{r}_v^l \right], \quad \mathbf{t}_v^l = \bar{\mathbf{P}}^l + k\mathbf{n}^l,$$
$$ {}^1\mathbf{r}_v^l = \frac{\mathbf{r}_y \times \left(-\mathbf{n}^l\right)}{\| \mathbf{r}_y \times \left(-\mathbf{n}^l\right) \|_2}, \; {}^2\mathbf{r}_v^l = \frac{-\mathbf{n}^l \times {}^1\mathbf{r}_v^l}{\| -\mathbf{n}^l \times {}^1\mathbf{r}_v^l \|_2}, \; {}^3\mathbf{r}_v^l = -\mathbf{n}^l . \tag{3}$$

Fig. 7 demonstrates the geometric relationship between the virtual camera C_v and the real camera C_r. Naturally, C_v can generate the desired fronto-parallel view of the traffic sign.

It is well known that the traffic sign images yielded by C_v and C_r are related by a homography $\mathbf{H}$ [1]. Suppose $\mathbf{P}$ is a 3D point on the traffic sign, and $\mathbf{q}_r$ and $\mathbf{q}_v$ are its images in C_r and C_v, respectively. Then the relationship between $\mathbf{q}_r$ and $\mathbf{q}_v$ can be expressed by

$$\tilde{\mathbf{q}}_r = \mathbf{H}\tilde{\mathbf{q}}_v \tag{4}$$

where $\tilde{\mathbf{q}}_r$and $\tilde{\mathbf{q}}_v$are the homogeneous coordinates of $\mathbf{q}_r$ and $\mathbf{q}_v$. Consequently, given $\mathbf{H}$, the fronto-parallel view of the traffic sign can be recovered from the distorted one. Now, we derive the formula of $\mathbf{H}$ for our problem. Denote $\mathbf{R}_v^r$ and $\mathbf{t}_v^r$ the orientation and translation of C_v relative to C_r, and $\mathbf{R}_l^r$ and $\mathbf{t}_l^r$ the orientation and translation of the LIDAR relative to C_r. Based on the foregoing description, we have

$$\mathbf{R}_v^r = \mathbf{R}_l^r \mathbf{R}_v^l, \quad \mathbf{t}_v^r = \mathbf{R}_l^r \mathbf{t}_v^l + \mathbf{t}_l^r . \tag{5}$$

Represent the parameters of the traffic sign plane $\mathbf{\Pi}$ in the C_v coordinate system by $\boldsymbol{\pi}^v = [(\mathbf{n}^v)^T, d^l]^T$. The relationship between $\boldsymbol{\pi}^v$ and $\boldsymbol{\pi}^l$ can be described by [1]

$$\boldsymbol{\pi}^v = \mathbf{T}\boldsymbol{\pi}^l, \mathbf{T} = \begin{bmatrix} \left(\mathbf{R}_v^l\right)^T & \mathbf{0} \\ \left(\mathbf{t}_v^l\right)^T & 1 \end{bmatrix}. \tag{6}$$

Finally, according to [1], $\mathbf{H}$ has the form

$$\mathbf{H} = \mathbf{K}\left(\mathbf{R}_v^r - \mathbf{t}_v^r \left(\mathbf{n}^v\right)^T / d^v\right)\mathbf{K}^{-1}, \tag{7}$$

where $\mathbf{K}$ is the intrinsic matrix of C_r defined in (1). Using (7), we can generate the fronto-parallel view of the traffic sign, as shown in Fig. 7.

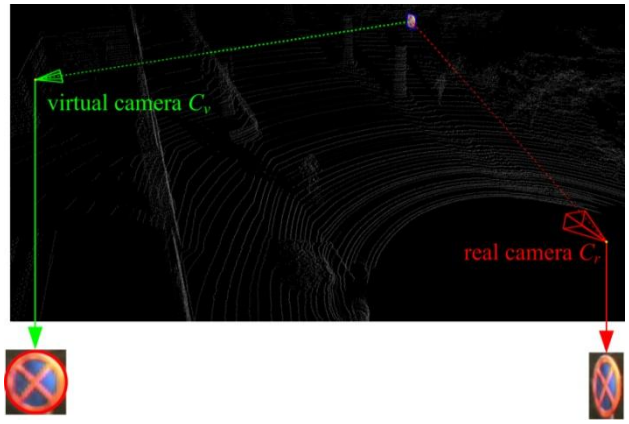


Fig. 7. An example of the perspective distortion rectification. The overlaid red circle illustrates that our rectification algorithm is effective.

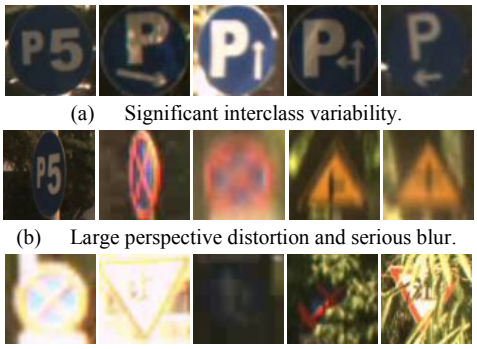
B. Feature and Classifier

HOG [6] was first designed for human detection. Now it is one of the most popular features and widely used in the traffic sign recognition [17]. In this paper, we adopt HOG as feature. To calculation the HOG feature, an image is divided into overlapping blocks. Each block consists of several cells. The gradient of each pixel is calculated first. Histogram of gradient orientation weighted by the gradient magnitude is computed for each cell. Then these histograms are locally normalized over the corresponding block. The final descriptor is a vector constructed by concatenating the normalized histograms of all the cells. The HOG feature used in this paper adopts the following parameters: 16×16 block containing four 8×8 cells; 8 pixels block spacing stride; 9 orientation bins; L2-Hys block normalization. This results in a 1764 dimension feature vector for a size of 64×64 window. For classification, linear SVMs are trained. Since calibration errors exist and the detected ROIs may contain multiple traffic signs, sliding window is performed around each ROI.

V. Experimental Results

In this section, we test the performance of our algorithm for the traffic signs within 100 meters in front of the vehicle, under challenging conditions. When the traffic sign is too far away, its image is very small and illegible. Furthermore, the LIDAR data beyond 100m are very sparse and hard to provide effective information. So we only consider traffic signs which are within 100m in the common fields of view of the camera and LIDAR. We collected 1,028 images with resolution 1,292×964 and the corresponding LIDAR data using our intelligent vehicle. A total of 2130 traffic signs were labeled. They belong to 17 classes illustrated in Fig. 8. Some challenging cases in our dataset are demonstrated in Fig. 9.

Fig. 8. The 17 classes of traffic signs in our dataset.



(a) Significant interclass variability.

(b) Large perspective distortion and serious blur.

(c) Poor illumination condition and serious occlusion.

Fig. 9. Some of challenging samples in our dataset.

A. *Traffic Sign Detection Results*

We first compare the performance of different point cloud segmentation algorithms. For this purpose, we collect 48,134 laser points from the traffic signs and 85,426 negative samples from other objects. Fig. 10 gives the distribution of the color. 50% of the positive and negative samples are randomly selected as the training samples. Linear SVM classifiers are trained on these samples for CCS alone and laser reflectivity integrated with different color spaces including RGB, HSV, $L^*a^*b^*$ and CCS. Fig. 11 shows the performance of the above features and the individual laser reflectivity (denoted by LR). We can see that CCS integrated with laser reflectivity is superior to other features. Gómez-Moreno *et al* [3] compared different color-based segmentation methods. They found that the best results come from thresholding the normalized RGB and Ohta color spaces. Table I lists the results of Gómez-Moreno's method [3], CCS, and CCS with laser reflectivity. We find that CCS outperform Gómez-Moreno's method [3], and CCS with laser reflectivity get the best result.

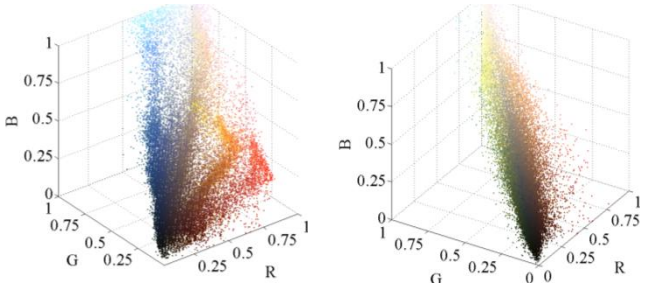
For the traffic sign detection, the true positive rate of the whole dataset achieves about 95.87%. Table II summarizes the true positive rate of our algorithm within different ranges. We find that our method gets very high true positive rate when the traffic sign is within 50*m*. As the traffic sign is far away from the vehicle, the true positive rate decreases. This is because the quality of both LIDAR and camera data reduces when the traffic sign is distant. On the other hand, the false positive rate of our algorithm is low. Our algorithm gives 33 false positives for the total 1,028 images. The false positives per frame are only about 3.21×10^{-2}. Due to the limited space, we do not present our detection results alone, but give several selected results of the whole system in Fig. 13 and Fig. 14.

TABLE I. THE RESULTS OF THE ALGORITHM IN [3] AND OUR ALGORITHM.

Method	True Positive Rate(%)	False Positive Rate(%)	Accuracy(%)
NRGB [3]	75.35	25.14	75.03
Ohta [3]	68.21	38.89	63.67
CCS	84.23	12.03	86.62
CCS+LR	**96.01**	**1.215**	**97.78**

TABLE II. THE TRUE POSITIVE RATE WITHIN DIFFERENT RANGES.

Distance(*m*)	[0,25)	[25,50)	[50,60)	[60,70)	[70,80)	[80,100]
True Positive Rate(%)	99.54	99.68	90.48	75.25	59.68	48.72



(a) Positive samples. (b) Negative samples.

Fig. 10. The color distributions of positive and negative samples.

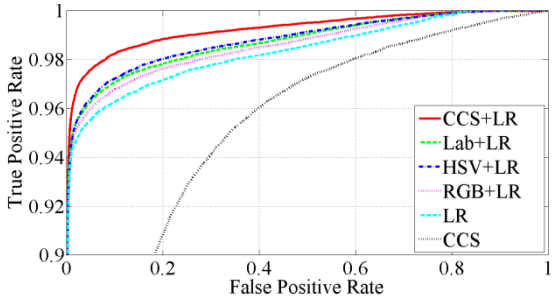


Fig. 11. The ROC curves of CCS, laser reflectivity, and laser reflectivity with different color spaces. LR denotes the laser reflectivity.

B. *Traffic Sign Recognition Results*

For the traffic sign recognition, 75% of samples of each traffic sign class are randomly selected to train linear SVMs. Because of serious occlusion and very bad illumination condition, 41 traffic signs are too illegible to determine their classes, as shown in Fig. 12. According to the scheme used in [21], these traffic signs are ignored in the traffic sign recognition stage to fairly evaluate the performance of our algorithm. Since our detection algorithm generates some false alarms, we use them and some background images as the negative class for the 17 traffic sign classes in our dataset.

We first consider the effect of the perspective distortion rectification. With rectification, the recognition error of the detected traffic signs is 0.898%. Without rectification, this error increases to 1.65%. Thus the rectification reduces the recognition error by 45.5%. Table III presents the recognition accuracy of the detected traffic signs within different ranges. We find an interesting result. The recognition accuracy does not decrease as the traffic sign is away from the intelligent vehicle. This is because we only evaluate the recognition performance on the detected traffic signs. Some traffic sign images with poor quality and low resolution are missed in the detection stage, so are not taken into account in the recognition stage. Table IV presents the final performance of our system. The false alarms yielded in the traffic sign detection stage are successfully eliminated. Additionally, our algorithm achieves real time. The average execution time of our algorithm is about 33.25 *ms* per frame on an Intel Core i7 computer. Fig. 13 demonstrates some results of our system in the challenging conditions. We can see that the proposed algorithm is able to correctly detect and recognize traffic signs that suffer from bad illumination condition, large perspective distortion, significant occlusion, and serious blur. Fig. 14 shows several samples containing incorrect results in our system. We observe that extreme illumination condition and large occlusion may result in the failure of our algorithm.

Fig. 12. Some samples of the ignored traffic signs. Such traffic signs are too illegible to recognize.

TABLE III. THE RECOGNITION ACCURACY OF THE DETECTED TRAFFIC SIGNS WITHIN DIFFERENT RANGES.

Distance(*m*)	[0,25)	[25,50)	[50,60)	[60,70)	[70,80)	[80,100]	[0,100]
Accuracy(%)	98.34	99.56	100	100	100	100	99.10

TABLE IV. OUR SYSTEM PERFORMANCE.

Total traffic signs	**Detection rate (%)**	**Recognition rate (%)**	**False Alarms /frame**	**Time (*ms*)**
2130	95.87	95.07	0	33.25

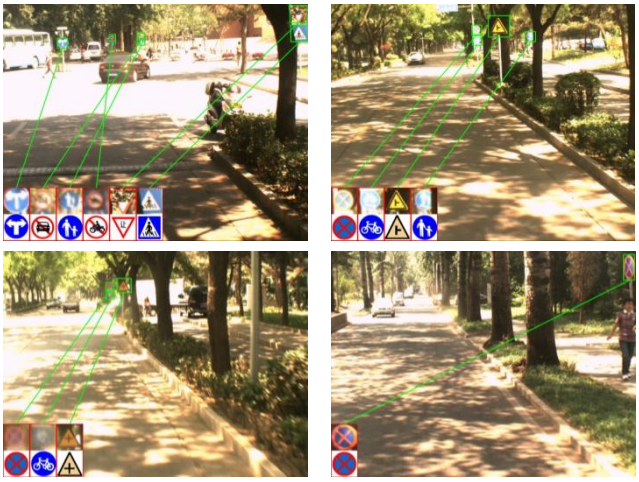


Fig. 13. Some results of our system in challenging conditions including bad lighting condition, large perspective distortion, serious occlusion and blur.

Fig. 14. Some samples that contain incorrect results. Extreme illumination condition and large occlusion make the traffic signs in the red rectangles (manually labeled data) missed or incorrectly recognized by our algorithm.

VI. CONCLUSION

In this paper, we propose a new traffic sign detection and recognition algorithm based on fusing LIDAR and camera data. Traffic signs are detected in the 3D space by employing the traffic sign position prior, color, laser reflectivity, and 3D geometric features. For the color, we introduce the CCS to represent the color, and treat different traffic sign colors as one class. The CCS and laser reflectivity are integrated to train a linear SVM to identify laser points belonging to the traffic sign. The 3D geometric features of the traffic sign are used to verify the ROIs. For traffic sign recognition, the ROIs suffering from perspective deformation are rectified first. Then, traffic signs are represented by the HOG descriptor and classified by linear SVMs. Our experimental results show that CCS outperform a single color space; fusing laser reflectivity and CCS outperforms the individual ones; the perspective distortion rectification reduces the false recognition error by 45.5%. For traffic signs within 100*m* under challenging conditions, the detection rate of our algorithm is 95.87%, and the recognition rate is 95.07%. In addition, our algorithm achieves real time performance. The average computing time is about 33.25 *ms* per frame. Hopefully, this algorithm can be extended to build the traffic sign map like that in [16].

REFERENCES

[1] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*: Cambridge university press, 2003.

[2] C. Nunny, A. Kummert, and S. Muller-Schneiders, "A novel region of interest selection approach for traffic sign recognition based on 3d modelling," in *IEEE Intelligent Vehicles Symposium*, Eindhoven, The Netherlands, 2008, pp. 654-659.

[3] 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, pp. 917-930, 2010.

[4] P. Gil Jiménez, S. M. Bascón, H. G. Moreno, S. L. Arroyo, and F. L. Ferreras, "Traffic sign shape classification and localization based on the normalized FFT of the signature of blobs and 2D homographies," *Signal Processing,* vol. 88, pp. 2943-2955, 2008.

[5] B. Soheilian, N. Paparoditis, and B. Vallet, "Detection and 3D reconstruction of traffic signs from multiple view color images," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 77, pp. 1-20, 2013.

[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, 2005, pp. 886-893.

[7] 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 Trans. Intell. Transp. Syst.,* vol. 13, pp. 1484-1497, 2012.

[8] 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 (IJCNN)*, Dallas, TX, 2013, pp. 1-8.

[9] G. Wang, G. Ren, Z. Wu, Y. Zhao, and L. Jiang, "A robust, coarse-to-fine traffic sign detection method," in *International Joint Conference on Neural Networks (IJCNN)*, Dallas, TX, 2013, pp. 1-5.

[10] M. Mathias, R. Timofte, R. Benenson, and L. Van Gool, "Traffic sign recognition—How far are we from the solution?," in *International Joint Conference on Neural Networks (IJCNN)*, Dallas, TX, 2013, pp. 1-8.

[11] M. Liang, M. Yuan, X. Hu, J. Li, and H. Liu, "Traffic sign detection by ROI extraction and histogram features-based recognition," in *International Joint Conference on Neural Networks (IJCNN)*, Dallas, TX, 2013, pp. 1-8.

[12] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The German traffic sign recognition benchmark: a multi-class classification competition," in *International Joint Conference on Neural Networks*, San Jose, California, USA, 2011, pp. 1453-1460.

[13] D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber, "Multi-column deep neural network for traffic sign classification," *Neural Networks,* vol. 32, pp. 333-338, 2012.

[14] J. Jin, K. Fu, and C. Zhang, "Traffic Sign Recognition With Hinge Loss Trained Convolutional Neural Networks," *IEEE Trans. Intell. Transp. Syst.,* 2014.

[15] H. González-Jorge, B. Riveiro, J. Armesto, and P. Arias, "Geometric Evaluation of Road Signs using Radiometric Information from Laser Scanning Data," in *28th International Symposium on Automation and Robotics in Construction*, 2011, pp. 1007-1012.

[16] V. Anh, Y. Qichi, J. A. Farrell, and M. Barth, "Traffic Sign Detection, State Estimation, and Identification Using Onboard Sensors," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, The Hague, 2013, pp. 875-880.

[17] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM,* vol. 24, pp. 381-395, 1981.

[18] S. Maldonado-Bascon, S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gomez-Moreno, and F. López-Ferreras, "Road-sign detection and recognition based on support vector machines," *IEEE Trans. Intell. Trans. Syst.,* vol. 8, pp. 264-278, 2007.

[19] 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.

[20] L. Zhou and Z. Deng, "Perspective Distortion Rectification for Planar Object Based on LIDAR and Camera Data Fusion," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Qingdao, China, 2014, to be published.

[21] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 34, pp. 743-761, 2012.