Detection and Recognition of Traffic Planar Objects Using Colorized Laser Scan and Perspective Distortion Rectification

Zhidong Deng, *Member, IEEE*, and Lipu Zhou

***Abstract*— Reliable detection and recognition of planar objects including traffic sign, street sign, and road surface in dynamic cluttered natural scenes are a big challenge for self-driving cars. In this paper, we propose a comprehensive method for planar object detection and recognition. First, the data association of LIDAR and camera is set up to acquire colorized laser scans, which simultaneously contain both color and geometrical information. Second, we combine three color spaces of RGB, HSV, and CIE L*a*b* with laser reflectivity as an aggregation-based feature vector. Third, the 3-D geometrical characteristics of planar objects that contain planarity, size, and aspect ratio are exploited to further reduce false alarm. Fourth, in order to increase robustness to any viewpoint variation, we present a new virtual camera-based rectification method to synthesize fronto-parallel views of refined object descriptors in 3-D space. Finally, experimental results achieved under a variety of challenging conditions show that integration of color space aggregation and laser reflectivity is superior to individuals. Specifically, the proposed perspective distortion rectification method remarkably eliminates false recognition error by 45.5%. Overall, the detection rate of our comprehensive method has up to 95.87% and the recognition rate even reaches 95.07% for traffic signs ranging within 100 m, with about 33.25 ms average running time per frame.**

***Index Terms*— Autonomous vehicle, colorized laser scan, color space aggregation, perspective distortion rectification, planar object detection and recognition.**

I. Introduction

PLANAR objects like traffic sign, street sign, and road surface are designed to inform or guide human drivers and pedestrians of road status, traffic rules, and geographical information. Reliable detection and recognition of such planar objects are crucial for self-driving car to safely navigate in real urban environments. Currently, self-driving car is equipped with a diversity of sensors to perceive surroundings and vehicle itself. Camera and LIDAR are considered as two popular sensing devices. Camera is able to seize appearance of planar objects like color, texture, and shape. Substantial camera-based algorithms have been proposed to detect and recognize planar objects. A lot of problems, however, are still open and challenging in pattern recognition, due to vulnerability and incompleteness of camera data. Typically, camera data are quite sensitive to changes in outdoor illumination, color temperature, and viewpoint. But 3D geometric information may be lost, and hard to be quickly and reliably recovered from images acquired by commonly-used perspective camera [1]. Actually, utilization of 3D *a prior* knowledge of planar objects will facilitate elimination of false detection error. Fortunately, LIDAR, as one of major active visions, can provide such geometric information. But it cannot capture visual information about planar objects. Hence data association of measurements from both sensors gives promising direction for detection and recognition of planar objects in dynamic natural scenes. To the best of our knowledge, however, few related work has been reported so far.

This paper proposes a reliable detection and recognition method for planar objects through combination of colorized laser scan, color space aggregation (CSA), which consists of RGB, HSV, and CIE L*a*b* color spaces, and perspective distortion rectification. First, colorized laser scans are acquired through data association of both LIDAR and camera. Second, planar objects including traffic sign, street sign, and road surface are detected directly in 3D natural scenes on the basis of aggregation-based feature vector, which concatenates CSA and laser reflectivity (LR) from colorized laser scans. In order to lower the false alarm, *a prior* knowledge of 3D planar geometry is further utilized. Third, for each of object descriptors or bounding boxes, we correct possible perspective deformation to produce their fronto-parallel views with *fixed* size. Recognition is then performed on such rectified fronto-parallel views using linear SVM with HOG features [2]. Finally, the experimental results achieved under a variety of challenging conditions demonstrate that our comprehensive method outperforms state-of-the-art results.

The main contributions of this paper include:

1) A framework to detect planar objects in dynamic cluttered traffic scenes by data association of LIDAR and camera. Laser points are virtually projected onto image plane to seek for corresponding colors, which are

Manuscript received June 19, 2016; revised November 28, 2016, April 21, 2017, and June 25, 2017; accepted July 2, 2017. Date of publication August 21, 2017; date of current version May 2, 2018. This work was supported in part by the National Science Foundation of China under Grant 91420106, Grant 90820305, and Grant 60775040, and in part by the National High-Tech Research and Development Program of China under Grant 2012AA041402. The Associate Editor for this paper was L. M. Bergasa. *(Corresponding author: Zhidong Deng.)*

Z. Deng is with the State Key Laboratory of Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science, Tsinghua University, Beijing 100084, China (e-mail: michael@tsinghua.edu.cn).

L. Zhou is with the Department of Computer Science, Tsinghua University, Beijing 100084, China (e-mail: zlp09@mails.tsinghua.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2017.2723902

associated to generate colorized laser scans. We then use color, LR, 3D planar geometry of colorized laser scans to jointly estimate refined object descriptors in 3D space. CSA is presented to entirely express different planar object colors and incorporates with LR as aggregation-based feature vector. We further exploit *a prior* knowledge about geometric characteristics of planar objects to reduce false alarm rate.

2) A new virtual camera-based rectification method for perspective distortion of planar object images. Perspective distortion rectification is critical to significantly improve recognition performance using naïve classifiers [3], [4]. In the proposed rectification method, we introduce a virtual camera to synthesize fronto-parallel views of object descriptors in 3D space, which does not require reliable shape classification and feature association like [3], [4].

Different from our published preliminary results in [5] and [6], this paper proposes a novel generic detection method for planar objects instead of traffic signs alone on the basis of colorized laser scans. Specifically, aggregation-based feature extraction is accompanied by incorporation of*a prior* knowledge of 3D planar geometry and especially, a virtual camera-based distortion rectification is presented. The paper is organized as follows: In Section II, the related work is reviewed. Section III proposes a planar object detection method. On the basis of perspective distortion rectification, a reliable planar object recognition method is presented in Section IV. Section V illustrates the experimental results yielded using our comprehensive method. Finally, Section VI concludes the paper with a brief summary.

II. Related Work

A. Planar Object Detection

Normally, positions of traffic sign posts are constrained. Some algorithms use this to produce ROI for planar objects in 2D image plane. In [7], typical positions of traffic signs were modeled in 3D space through three segments of hollow cylinders. ROI of traffic signs was generated by projecting 3D points into image plane. For real time application, ROI for 15 driving situations was computed in advance and stored as look up tables. In [8], the fuzzy set was employed to obtain adaptive rectangular ROI, whose position and size were determined by current speed and steering wheel angle of vehicle. These methods can significantly reduce search space for traffic sign. However, the ROIs obtained from both algorithms are rough. They are unable to eliminate the background, such as sky and ground, where there are not planar objects. Additionally, the ROI yielded in [8] ignores the margin of image. This may miss some nearby traffic signs. In this paper, we use LIDAR data to get much tighter and more accurate ROI or bounding box in 3D space.

Color and shape are two distinguished features of planar objects and have been extensively studied [3], [9]–[12]. A large amount of researchers jointly employ them in detection phase of planar objects for color image. Basically, color feature is used to find ROI and shape feature is then extracted to screen and categorize candidates. For instance, in [9], hue and saturation components of HSI color space were exploited to detect red, blue, and yellow colors, and achromatic decomposition was used to search for white color. Heuristic rules over size and aspect ratio were used to eliminate false alarm. Sequentially, distance to borders (DtBs) as input features were fed to SVM to classify shape of candidate regions. In addition, there also has the previous work that either color or shape is utilized in a sole or major manner. Shape-based algorithms analyze the edge of image obtained from some derivative operators, such as Canny operator [13]. They take advantage of the fact that traffic signs are circle or regular polygons which can be detected by Hough-based algorithms, such as the radial symmetry detector [14], and its derivatives [15], [16].

Appearance like color and shape is another important cue for traffic sign detection. In [8], two adaboost classifiers [17] were trained to detect prohibitory and warning signs in the region determined by color and driving status. In [18], traffic sign was detected by embedding dissociated dipoles feature in a cascade classifier trained by the evolution version of adaboost algorithm. Histogram of oriented gradients (HOG) [2] is a powerful descriptor for object detection and recognition. In traffic sign detection, HOG has been widely used and frequently achieved excellent performance. For instance, Creusen *et al.* [19] made further improvement of HOG-based detection algorithm by introducing a color transformation. Overett and Petersson [20] presented HOG variants and compared their performance in planar object detection.

Although all those algorithms are highly competitive for traffic sign detection and recognition, the comparison of them seems difficult. Actually, it is no clear which one outperforms others. This is because most of the algorithms are trained and tested on their own dataset. In recent years, such situations have been improved. Mogelmose *et al.* [21] analyzed performance of existing detection algorithms and then recommended some public traffic sign databases. Specifically, the competition on German Traffic Sign Detection Benchmark (GTSDB) [22] strongly promotes comparative studies of different algorithms. For GTSDB, some algorithms reported excellent results. But their running time is not suitable for real time application. Liang *et al.* [23] detected traffic signs by using the HOG feature combined with the color histogram to refine the ROIs provided by the color and shape information. The running time of the method is about 0.4-1.0 second. Mathias *et al.* [24] adopted integral channel features classifier (ChnFtrs) [25] to detect traffic signs. To deal with perspective distortion, the detector was applied in 50 scales and 5 ratios of the image. The speed of the detector is about 0.35 Hz with the help of GPU. Wang *et al.* [26] employed the HOG feature and the coarse-to-fine sliding window scheme, which processing time reached up to several seconds.

As mentioned above, LIDAR also provides important geometric information about planar objects. But there are few previously published approaches on this problem. Considering that planar objects used in traffic scenes, e.g., traffic sign, street sign, and road surface, are usually painted with highly retroreflective materials, strong LR measurements can be yielded.

This characteristic is adopted by [27]–[29] so as to detect planar objects. LR, however, is not only associated with materials that laser beam contacts, but also affected by object poses.

B. Planar Object Recognition

Some images captured by camera for planar objects, including traffic sign, street sign, and road surface, are subject to serious perspective distortion. In general, such planar object images need to be first rectified in the recognition phase, in order to significantly improve recognition accuracy.

First, the perspective distortion of traffic sign is inevitable, even if being perfectly perpendicular to a roadway. Actually, maneuver of self-driving cars may cause captured traffic sign images to suffer from significant perspective distortion. Consequently, the recovery of the fronto-parallel view of distorted traffic sign images is important for its reliable recognition, and has been studied by many researchers [4], [18], [30], [31]. Owing to the fact that there are specific shapes in traffic sign, the crux of these algorithms is to recognize the shape of ROI and then establish point correspondences between the reference shape and the distorted one. In [4], the shape is classified by analyzing the FFT of ROI contour. Soheilian *et al.* [30] described a RANSAC-based algorithm to recognize the shape of ROI. Baró *et al.* [18] used different cascade classifiers that are trained by an evolutionary version of adaboost, in order to detect traffic signs with different shapes. In [31], the shape information is obtained by SVM classifier with HOG feature [2]. Apparently, the algorithms are all dependent on identification of traffic sign shapes and reference points, which may get worse in complex scenarios. This is because either shape recognition or feature point detection is not robust against large perspective distortion and occlusion. Once traffic sign shapes are misclassified or control points are occluded or misestimated, these algorithms will yield false results. On the other hand, the homography is generally approximated by the affine transformation for triangular sign [4], [18], [30], [31] and circular sign [30], because the number of point correspondences is not adequate to evaluate the homography in the cases.

Second, text contained in street sign has valuable semantic information to explain scenes. Similarly, perspective recovery before text recognition is also one of the main challenges [32]–[34]. In traffic scenes, text along roadside captured by onboard camera is generally subject to serious perspective distortion, which leads to reduction of performance in optical character recognition (OCR) [35]. Clark and Mirmehdi [36] presented an algorithm to remove perspective distortion for the text image by estimating the horizontal and vertical vanishing points. Such an algorithm requires the text to have multiple lines, rather than single line case indicated in street sign. In addition, the computational load of this algorithm is too heavy to be suitable to real-time applications. Cambra and Murillo [37] focused on rectifying text that is enclosed by a rectangular box. But it is not applicable to other shapes. Merino-Gracia *et al.* [35] proposed a perspective correction algorithm based on the geometry of characters themselves instead of the layout or border of text. As stated in [35], the accuracy of the algorithm relies on image quality, and false text segmentation or low resolution may result in bad performance.

Third, lane markings in road surface are important for self-driving cars and advanced driver assistant system. Due to constrained data acquisition conditions, there always exists perspective distortion in a road surface image. As a result, the construction of a bird's-eye view of road surface image is a crucial step for substantial approaches on lane markings detection and tracking [38]–[41]. In the literatures, the adoption of static rectifying homography, which is often calculated off-line, is essentially based on assumption that the relative pose between camera and road surface is kept unchanged in a period of autonomous driving of self-driving cars. Bertozzi and Broggi [38] presented an inverse perspective mapping approach to achieve a bird's-eye view of a road region in front of vehicle. In [39] and [40], the rectifying homography is estimated by four reference points. Since static homography is exploited in those approaches, their performance degrades as vehicle jolts or upcoming roadway is uphill or downhill. For this problem, Yao *et al.* [41] proposed an algorithm that attempts to dynamically evaluate camera pitch angle through combination of LIDAR and camera data. But the accuracy of homography estimation of such the algorithm heavily relies on to what degree the road surface parallels the x-y plane of LIDAR, since the z-component of laser points on a roadway is ignored.

Basically, those algorithms depend on classification of planar object shapes and estimation of control points or curve parameters. This is main drawback of such algorithms. In fact, it probably generates false results in complicated scenarios, once planar object shapes are misclassified or control points are occluded or misestimated.

Substantial approaches have been devoted to improve performance of planar object recognitions. In [9], SVM with Gaussian kernel was trained for traffic sign recognition. The approach proposed by Greenhalgh and Mirmehdi [42] employed linear SVM and HOG feature. Carrasco *et al.* [43] compared template matching and neural networks with different image preprocessing techniques. Their experimental results showed that cascade neural networks using the P-Tile preprocessing gave the best performance. In addition, a couple of excellent classifiers, e.g., the k-d tree, the random forest, and the linear SVM, and a few of sophisticated features, i.e., HOG and DT, were comparatively studied by Zaklouta and Stanciulescu [44]. They reported that the random forest with the HOG feature could have the best performance. Sermanet and LeCun [45] presented an algorithm based on multi-scale convolutional neural networks. In the study by Ciresan *et al.* [46], multi-column deep neural network was trained for traffic sign classification. The recognition rate of [45]–[47] achieved 99.46%, 98.31%, and 99.65% in German traffic sign recognition benchmark (GTSRB) [48], respectively. Lately, Haloi [49] proposed a new classification approach for traffic sign using deep inception based convolutional networks and further achieved state-of-the-art performance of 99.81% on GTSRB dataset. But the computational load for deep

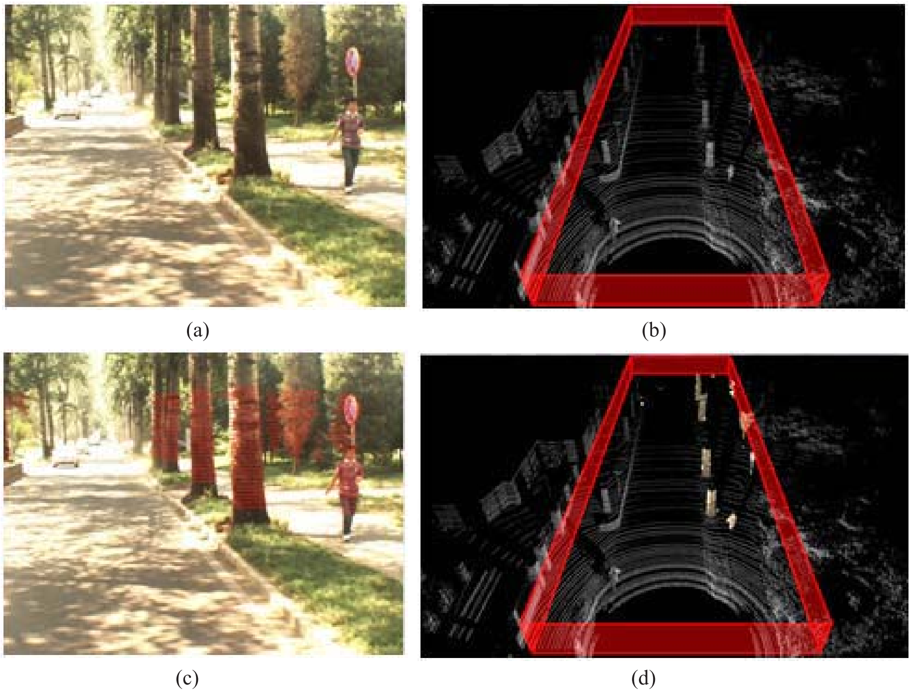


Fig. 1. The data association of LIDAR and camera. (a) Original image. (b) Search space for planar objects. (c) Projection from laser scans in the search space to image pixels. (d) Colorized laser scans achieved.

convolutional neural networks is likely beyond capabilities of mainstream host, which has to be practically realized on high-power GPUs.

III. Planar Object Detection Using Incorporation of Aggregation-Based Feature Vector Into a Prior Knowledge of 3D Geometry

A. Colorized Laser Scan

Let us consider LIDAR and camera pair that is rigidly fixed. In order to associate a laser point with its corresponding image pixel, extrinsic parameters of both LIDAR and camera, together with intrinsic parameters of camera itself, must be precomputed as a prerequisite. In general, extrinsic parameters can be obtained through [50]. According to [51], it is easy to estimate camera intrinsic parameters.

In 3D space, relative position of planar objects provides useful knowledge to narrow down search space for planar object detection. But it is intractable in 2D image plane due to the fact that geometric relationship between objects is lost. In this paper, a much tighter bounding box for planar objects is presented. Thanks to geometrical information acquired from LIDAR, search space for planar objects can be roughly determined in the range of surroundings along roadside. Owing to search spaces being greatly decreased, more sophisticated algorithms can be adopted here.

1) In general, relative positions of planar objects like traffic/street signs are physically constrained by roadway. Precise road model can be built using both high-definition 3D grid maps and on-line camera-based lane markings detection. Considering that a paved road is usually of straight line or smoothly curved, the search space for planar objects can be roughly restrained in a driving cuboid aligned along roadway boundaries like that in Fig. 1(b). Hence we can explore and locate planar objects only in such a cuboid. Corresponding ROIs in image plane is found by means of transformation matrix from laser scans to image pixels, as shown in Fig. 1(c).
2) Owing to the fact that data association between laser scans and image pixels is completed, the laser scans falling into the camera field of view can gain colors. In the meantime, the corresponding pixels are also able to yield depths. This results in generation of colorized laser scans, as shown in Fig. 1(d). In contrast to the previous work of detecting planar objects in 2D image plane, the planar object detection in our method is all done in 3D space based on colorized laser scans that contain both color and geometrical information. Apparently, it is much easier to deal with since 3D geometric model is invariant in terms of rigid-body transformation.

B. Aggregation-Based Feature Vector

Planar objects are often designed to have specific colors so as to distinguish them from the surroundings. For color detection, there are many approaches, such as [3] and [9]–[12], to empirically set threshold over color space. In fact, either outdoor lighting conditions or color temperatures are not controllable. Illumination, weather conditions (e.g., fog), and sign aging or shade all have impact on captured colors. In addition, similar colors from cluttered background are not easily discriminable and may result in high false alarm rates.

On the other hand, planar objects like traffic and street signs are broadly painted with highly retro-reflective materials. Fortunately, laser reflectivity (LR) acquired from LIDAR perfectly characterizes such property [27]. The planar surface attribute and specific materials make their LR values

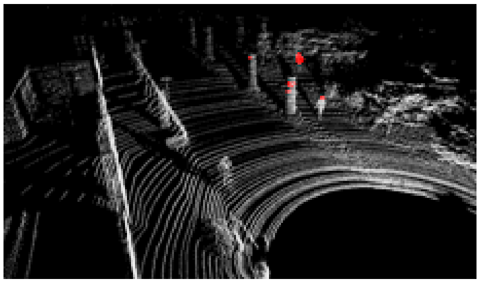


Fig. 2. Segmentation of the laser points obtained using linear SVM with aggregation-based feature vector (CSA+LR). The red points represent the laser points on traffic signs.

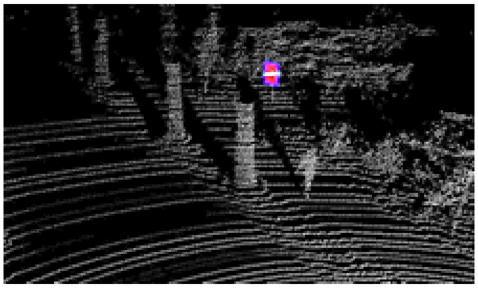


Fig. 3. The planar surface fitting and the bounding box of red points in Fig. 2.

very different from surrounding objects (e.g., trees and pedestrians) [6]. But LR is dependent on both materials and object poses. If orientation of planar objects has large angle deviation from laser direction, LR of planar objects becomes low and even loses discriminability [6].

The effectiveness of LR is straightforward. But it is hard to describe color features in a sense. This is because there are several color spaces and colors painting on planar objects often contain red, green, yellow, blue, white, and black. Roughly, the previous work separately classified the colors in individual color space. It probably leads to the use of many thresholds or multiple classifiers. Consequently, we attempt to treat them uniformly. This is different from that proposed by Tsai *et al.* [3], where two colors of red and green instead of six colors in our study are handled as one class. Actually, focusing them to be one class and classifying them in individual color space will result in bad performance. If single color space is well suited to expressing single or a few colors, several color spaces can then be integrated to classify all the colors. For this reason, we present a color space aggregation (CSA), consisting of RGB, HSV, and CIE L*a*b* color spaces, to detect colors. Such three color spaces were also used for object recognition in [52]. Correspondingly, the aggregation-based feature vector for each laser point has 10 elements, i.e., 3×3 CSA components and one LR. Furthermore, linear SVMs are trained to classify/detect colorized laser scans. An example of segmentation of colorized laser scans is illustrated in Fig. 2. The red points indicate laser points on traffic sign. It is readily observed that the detection results produce some false alarms. But it is much better than that given in [6].

C. 3D Geometric Characteristics of Planar Objects

In order to further eliminate false detection error, we proceed to exploit *a prior* knowledge about 3D geometric characteristics of planar objects, including planarity, physical size, and aspect ratio of objects in 3D space. Assume that N colorized laser points are classified as belonging to planar objects. Let us denote such point set as S. Before further analysis, we make segmentation of S according to Euclidean distance between laser points. Suppose S is divided into M partitions, i.e.,

$$S = \bigcup_{i=1}^{M} S_i = \bigcup_{i=1}^{M} \{\boldsymbol{p}_{ij}\}_{j=1,\cdots,N_i}, \tag{1}$$

where S_i is the i-th partition of S, which contains N_i laser points p_{ij}.

We should check to what degree S_i can be approximated by a plane. One of straightforward methods is to fit plane for each S_i. In this case, the residual of the fitting is checked to determine whether S_i is on a plane. But S_i may have points from other objects such as leaves shading planar objects. Planar objects themselves may also be partially damaged. Fitting plane using entire S_i may result in eliminating correct regions. In fact, those points with large noise have impact on accuracy of plane parameters, even if all points in S_i are exactly from a planar object. Accurate plane parameters are essential for subsequent region verification and planar object rectification. Based on the above-mentioned reasons, we should reduce non-planar part of S_i and have removal of laser points that are corrupted by noise. It implies that we need to robustly estimate optimal planar part P_i of S_i. The RANSAC algorithm [53] is better suited to such problem. After P_i is obtained, the ratio between the sizes of P_i and S_i is a fair indicator to evaluate planarity of S_i. Partitions S_i are discarded if the ratio is less than a threshold $r_p (r_p = 0.6$ in our experiment). The planar surface fitting of red laser points in Fig. 2 is shown in Fig. 3.

3D physical size and aspect ratio of planar objects are additional two cues to decrease false alarm rates, which is extensively used in the previous work such as [9], [10], [21], and [42]. Stallkamp *et al.* [54] analyzed size of planar object images in GTSRB. In such benchmark, size of planar object images varies from 15 × 15 to 222 × 193 pixels. Meanwhile, aspect ratio of planar object images depends on pose of planar objects relative to camera, which may lead to big aspect ratio changes. Physical size of planar objects in 3D space, however, is much more restrained. Detecting them directly in 3D space is more reasonable. Although colorized laser points become sparse as planar objects being far away LIDAR, 3D geometric characteristics such as size and aspect ratio of planar object are still roughly maintained.

After the foregoing steps, non-planar laser point sets are eliminated. For the remaining regions, those points in optimal planar set P_i are orthogonally projected into estimated planar surface $\boldsymbol{\Pi}_i$. Let Q_i denote corresponding projection of P_i. We then calculate a bounding box B_i of Q_i, as indicated by blue in Fig. 3. The size and aspect ratio of B_i are analyzed to further reduce false alarms. Assume the width and height of B_i are expressed by w_i and h_i, respectively. In our experiment, we kept record of B_i, where $\max(w_i, h_i)$ 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 (or 4.0), if S_i is falling into the LIDAR's field of view.

IV. Reliable Planar Object Recognition Using Perspective Distortion Rectification

A. Virtual Camera-Based Rectification

Some of planar object images captured are subject to serious projection distortion, as shown in Fig. 1(a). In this paper, we propose a novel rectification algorithm based on data association of LIDAR and camera. The basic idea of the proposed algorithm is to set a virtual camera C_v so as to synthesize desired fronto-parallel view of planar object images. For the purpose of this, virtual camera C_v is placed to look orthogonally at planar objects. Suppose that centroid of all laser points projected onto planar object $\boldsymbol{\Pi}$ is denoted as $\bar{\boldsymbol{P}}^l$, planar surface $\boldsymbol{\Pi}$ in LIDAR coordinate system as $\boldsymbol{\pi}^l = \left[(\boldsymbol{n}^l)^T,\ d^l\right]^{\mathrm{T}}$, and translation and orientation of C_v relative to LIDAR L as $\boldsymbol{t}_v^l$ and $\boldsymbol{R}_v^l$, respectively. We put such virtual camera at k meter in front of planar objects, along the line segment that passes centroid of planar objects $\bar{\boldsymbol{P}}^l$ and is oriented as the same as planar normal $\boldsymbol{n}^l$, which implies that z-axis of C_v is onto direction $-\boldsymbol{n}^l$.

If the y axis $\boldsymbol{r}_y$ of C_v is determined, $\boldsymbol{t}_v^l$ and $\boldsymbol{R}_v^l$ can then be given by

$$\boldsymbol{t}_v^l = \bar{\boldsymbol{P}}^l + k\boldsymbol{n}^l, \boldsymbol{R}_v^l = \left[{}^1\boldsymbol{r}_v^l,\ {}^2\boldsymbol{r}_v^l,\ {}^3\boldsymbol{r}_v^l\right], \tag{2.a}$$

where

$${}^1\boldsymbol{r}_v^l = \frac{\boldsymbol{r}_y \times \left(-\boldsymbol{n}^l\right)}{||\boldsymbol{r}_y \times \left(-\boldsymbol{n}^l\right)||_2},\ \ {}^2\boldsymbol{r}_v^l = \frac{-\boldsymbol{n}^l \times {}^1\boldsymbol{r}_v^l}{||-\boldsymbol{n}^l \times {}^1\boldsymbol{r}_v^l||_2},\ \ {}^3\boldsymbol{r}_v^l = -\boldsymbol{n}^l. \tag{2.b}$$

Meanwhile, if the x axis $\boldsymbol{r}_x$ of C_v is evaluated, we have

$${}^1\boldsymbol{r}_v^l = \frac{{}^2\boldsymbol{r}_v^l \times \left(-\boldsymbol{n}^l\right)}{||{}^2\boldsymbol{r}_v^l \times \left(-\boldsymbol{n}^l\right)||_2},\ \ {}^2\boldsymbol{r}_v^l = \frac{-\boldsymbol{n}^l \times \boldsymbol{r}_x}{||-\boldsymbol{n}^l \times \boldsymbol{r}_x||_2},{}^3\boldsymbol{r}_v^l = -\boldsymbol{n}^l. \tag{2.c}$$

Let us represent the planar surface $\boldsymbol{\Pi}$ in the C_v coordinate system as $\boldsymbol{\pi}^v = \left[(\boldsymbol{n}^v)^T,\ d^v\right]^{\mathrm{T}}$. The relationship between $\boldsymbol{\pi}^v$ and $\boldsymbol{\pi}^l$ can be described by

$$\boldsymbol{\pi}^v = \boldsymbol{T}\boldsymbol{\pi}^l, \quad \boldsymbol{T} = \begin{bmatrix} \left(\boldsymbol{R}_v^l\right)^T & \boldsymbol{0} \\ \left(\boldsymbol{t}_v^l\right)^T & 1 \end{bmatrix}. \tag{3}$$

Assume that translation and orientation of C_v relative to C_r are expressed as $\boldsymbol{t}_v^r$ and $\boldsymbol{R}_v^r$, respectively. It has

$$\boldsymbol{t}_v^r = \boldsymbol{R}_l^r\boldsymbol{t}_v^l + \boldsymbol{t}_l^r, \boldsymbol{R}_v^r = \boldsymbol{R}_l^r\boldsymbol{R}_v^l. \tag{4}$$

Let $\boldsymbol{P}$ be a point on planar object. Suppose that $\boldsymbol{q}_r$ and $\boldsymbol{q}_v$ indicate corresponding pixel image on C_r and C_v, respectively. Based on [1], the relationship between $\boldsymbol{q}_r$ and $\boldsymbol{q}_v$ can be expressed below,

$$\tilde{\boldsymbol{q}}_r = \boldsymbol{H}\tilde{\boldsymbol{q}}_v, \quad \boldsymbol{H} = \boldsymbol{K}\left(\boldsymbol{R}_v^r - \boldsymbol{t}_v^r\left(\boldsymbol{n}^v\right)^T/d^v\right)\boldsymbol{K}^{-1}, \tag{5}$$

where $\tilde{\boldsymbol{q}}_r$ and $\tilde{\boldsymbol{q}}_v$ denote homogeneous coordinates of $\boldsymbol{q}_r$ and $\boldsymbol{q}_v$, respectively, and $\boldsymbol{K}$ represents intrinsic parameter of C_r. According to (5), we can generate fronto-parallel view of planar objects with *fixed* size, as shown in Fig. 4.

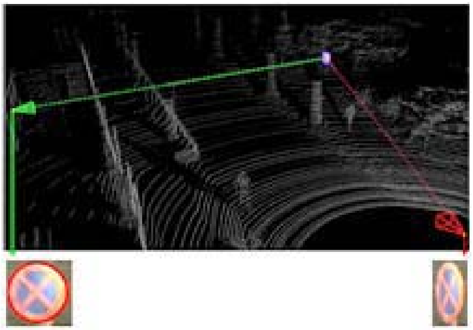


Fig. 4. The desired fronto-parallel view (left) achieved by the perspective distortion rectification.

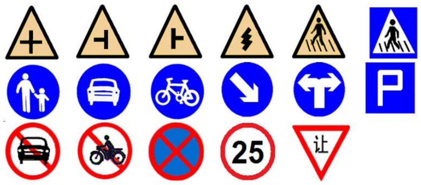


Fig. 5. The 17 classes of traffic signs in our dataset.

B. Feature Selection and Classifier

HOG was at first designed for human detection [2]. At present, it becomes one of the most successful image features in computer vision and is further applied in traffic sign recognition [17]. This paper adopts HOG as feature vectors for classification or recognition of rectified fronto-parallel views of planar objects. The parameters of HOG features are selected below: 16×16 block, each containing four 8×8 cells; 8 pixels block spacing stride; 9 orientation bins; L2-Hys block normalization. This produces 1,764 dimensional feature vectors for a size of 64×64 sliding window. Furthermore, linear SVM is trained. Since there exist calibration errors and bounding box detected or object descriptor may contain multiple planar objects, the sliding window is conducted around each object descriptor.

V. Experimental Results

In this section, we evaluated performance of our method through extensive experiments. The classes of traffic signs in our dataset are shown in Fig. 5, all of which were ranged within 100m in the field of view of both camera and LIDAR. Actually, the image of planar objects became tiny and illegible as planar objects were too far away. For example, the front view of 0.5m.×0.5m real traffic sign of being 100m away from the camera was shrunk to the size of about 11.×11pixels in camera image plane. Fig. 6 shows some of challenging samples in our dataset.

A. Detection Results

In our experiments on data collection, laser point clouds were captured by Velodyne HDL-64E S2 LIDAR installed,

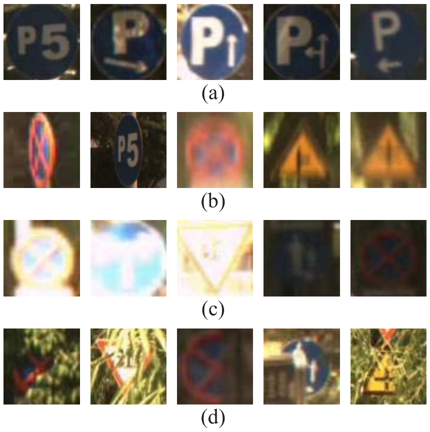


Fig. 6. Some of challenging samples included in our dataset. (a) Wide variability in appearance. (b) Large perspective distortion and serious blur. (c) Bad illumination condition. (d) Partial occlusion.

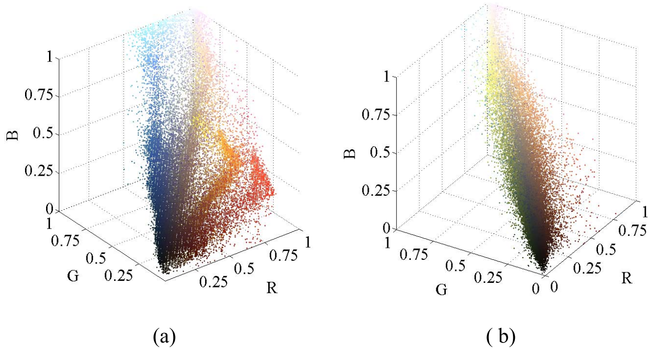
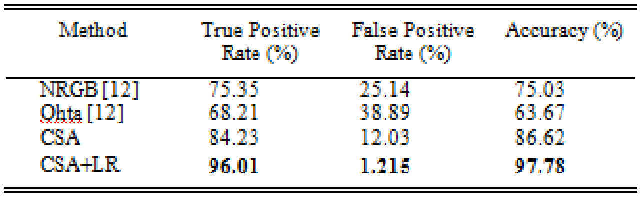


Fig. 7. The color distributions. (a) Positive samples. (b) Negative samples.

while corresponding images were acquired by Basler digital camera with resolution of 1292x964. In the detection phase, we selected 48,134 positive samples related to planar objects and 85,426 negative ones concerning other objects, where both color and LR were kept and 50 percent of all the positive and negative samples were randomly chosen as training datasets. Fig. 7 gives color distributions. For each of features except LR, linear SVM classifier was trained on those samples. Fig. 8(a) shows the ROC curves with different color spaces, i.e., individual RGB, HSV, L*a*b*, and CSA (RGB + HSV + L*a*b*). It was obvious that CSA outperformed other individual color space. Fig. 8(b) shows the ROC curves of CSA, LR, CSA+LR, and different color spaces with LR. It is easy to see that CSA+LR were superior to the other features. In [12], Gómez-Moreno *et al.* compared different image-based segmentation methods. They observed that the best result came from thresholding normalized RGB and Ohta color spaces. Table I lists the results of the Gómez-Moreno's methods, CSA, and CSA+LR. It is very clear that CSA+LR achieved the best result, and CSA surpassed Gómez-Moreno's method [12].

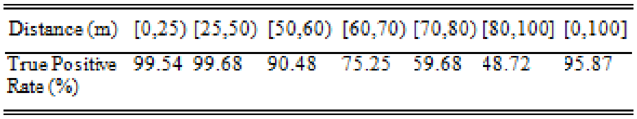
The true positive rate of the whole dataset had about 95.87%. Table II lists the true positive rate within different ranges. We found that our method yielded very high true positive rate, if planar objects were ranged within 50 m. As planar objects were far away from the self-driving car, the true positive rate decreased. This was because data quality of LIDAR and camera reduced with planar objects being far from sensors. Our algorithm achieved 33 false positives for a total of 1,028 images. Hence the false positive rate per frame was only about 3.21%, ranging within 100 m.

TABLE I
THE COMPARISON OF GÓMEZ-MORENO'S METHOD . [12] AND OUR METHOD



Method	True Positive Rate (%)	False Positive Rate (%)	Accuracy (%)
NRGB [12]	75.35	25.14	75.03
Ohta [12]	68.21	38.89	63.67
CSA	84.23	12.03	86.62
CSA+LR	**96.01**	**1.215**	**97.78**

TABLE II
THE TRUE POSITIVE RATE FOR DIFFERENT RANGES



Distance (m)	[0,25)	[25,50)	[50,60)	[60,70)	[70,80)	[80,100]	[0,100]
True Positive Rate (%)	99.54	99.68	90.48	75.25	59.68	48.72	95.87

B. Rectification Results

In the following, we consider the virtual camera-based rectification for perspective distortion of planar objects, including traffic sign, street sign, and road surface, as shown in Fig. 9.

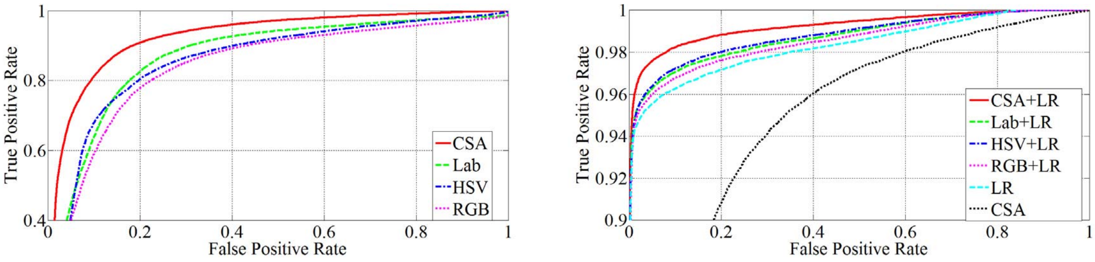
As shown in Fig. 9(a) and (b), the traffic sign and the street sign are erected in the shoulders. But the normal of them may not be oriented vertically to the camera image plane. Thus we can assign the y-axis of virtual camera C_v parallel the normal of the road surface. In our self-driving car, the z-axis of LIDAR L is approximately perpendicular to the road surface. Accordingly, we set $\boldsymbol{r_y} = [0,\ \ 0,\ \ -1]^{\mathrm{T}}$. Fig. 10 shows the rectification results for the traffic sign and the street sign given in Fig. 9(a) and (b), respectively. It is clear that our rectification method correctly recovers the fronto-parallel views of the distorted images.

Let us consider the perspective rectification for road surface. Similarly, we can use z-coordinates of the laser points to roughly identify the road surface. Owing to the fact that the y-axis of L is consistent with the roll axis of our self-driving car, we choose the y-axis of C_v to be the opposite direction of the y-axis of L, i.e., $\boldsymbol{r_y} = [0,\ \ -1,\ \ 0]^{\mathrm{T}}$. Fig. 11(a) shows the poses of C_r and C_v for the road surface given in Fig. 9(c). Fig. 11(b) gives the fronto-parallel view. It is obvious that the parallelism of the lane markings is perfectly recovered.

For other planar objects with irregular shape or unknown aspect ratio, which are intractable for camera-based approaches, our generic virtual camera-based rectification method still achieves distinguished performance, as show in Fig. 12.

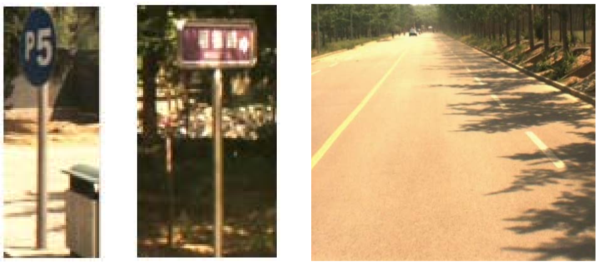
C. Recognition Results

In the recognition phase, we collected 1,028 sample images with 1,292.×964 and corresponding LIDAR data as well.



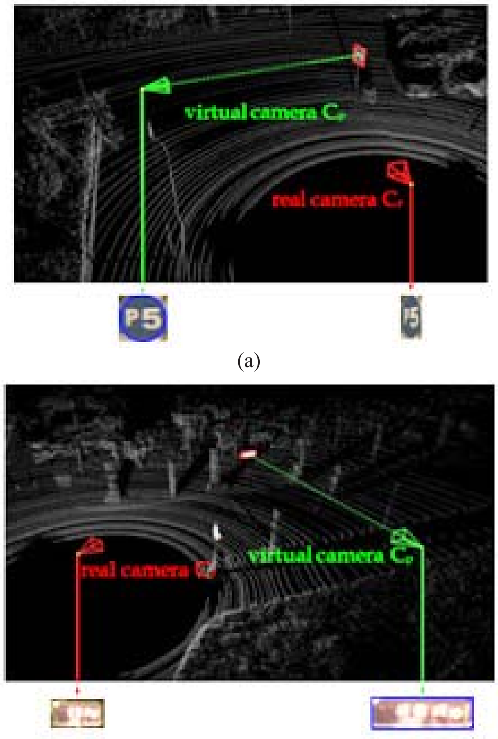
(a) (b)

Fig. 8. The ROC curves with different features. (a) The ROC curves of CSA and three individual color spaces. (b) The ROC curves of CSA, LR, CSA+LR, and different color spaces with LR.



(a) (b) (c)

Fig. 9. Some samples of planar object images that are subject to serious perspective distortion.(a) Traffic sign. (b) Street sign. (c) Road surface.

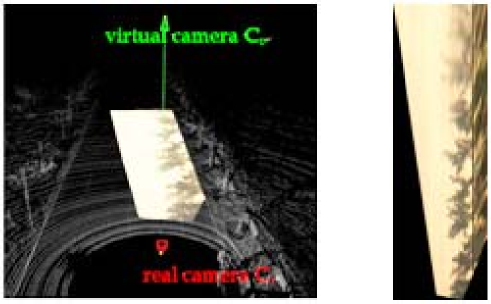


(a)

(b)

Fig. 10. The rectification results for the traffic sign and the street sign given in Fig. 9(a) and (b), respectively.

It belonged to 17 classes (Fig. 5) and 75 percent of the samples of each traffic sign class were randomly selected to train linear SVM classifiers. As shown in Fig. 13, 41 traffic signs were too illegible to identify their classes due to severe occlusion and very bad illumination condition. Those traffic signs were then ignored in the recognition phase for planar objects. Since there



(a) (b)

Fig. 11. The rectification results for the road surface in Fig. 9(c). (a) A schematic diagram of the road surface rectification. The virtual camera C_v looks orthogonally toward the road surface so as to generate the bird's-eye view. (b) In the rectified road surface image, the parallelism between the lane markings is correctly recovered.

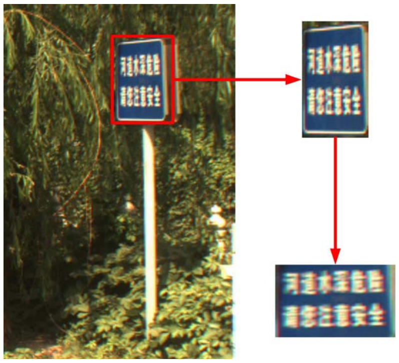


Fig. 12. The rectification results for other planar objects with unknown aspect ratio.

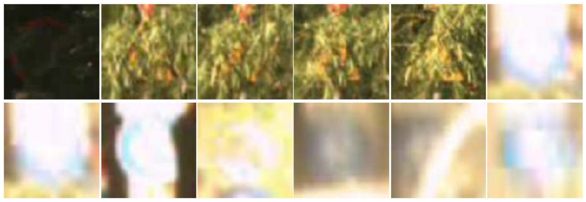


Fig. 13. Some samples of unidentified traffic signs. These traffic signs are too illegible to recognize.

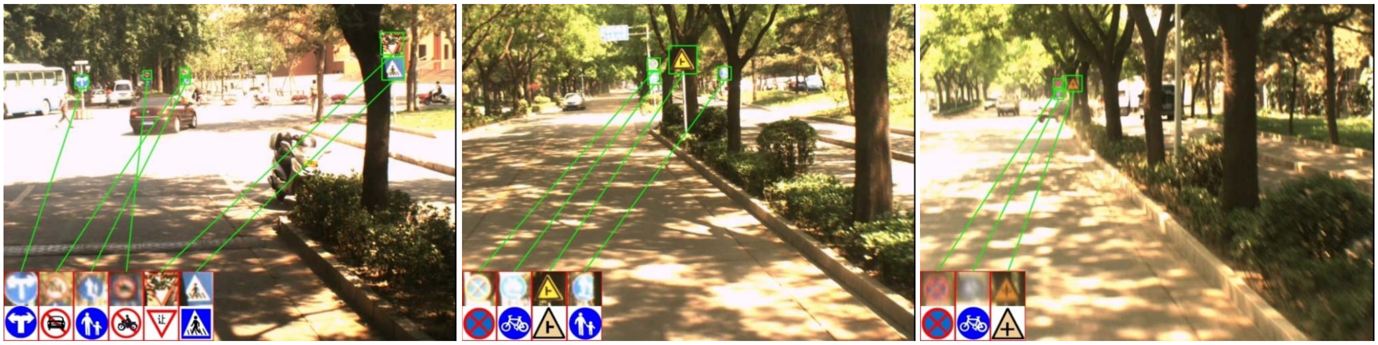
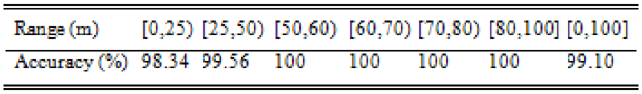


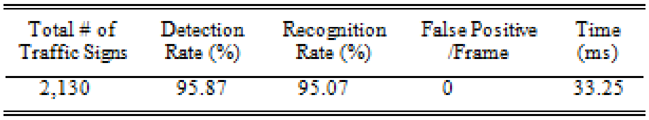
Fig. 14. Some results of our method in challenging conditions.

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



Range (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
PERFORMANCE OF OUR METHOD



Total # of Traffic Signs	Detection Rate (%)	Recognition Rate (%)	False Positive /Frame	Time (ms)
2,130	95.87	95.07	0	33.25

were some false positives in the above-mentioned detection, we treated them as the negative class for 17 classes of planar objects in our training dataset.

Let us consider the effect of the perspective distortion rectification. With rectification, the recognition error of the detected traffic signs was 0.898%. If no rectification, this recognition error increased to 1.65%. As a result, our rectification method remarkably decreased the recognition error by 45.5%. Table III lists the recognition accuracy of the traffic signs detected within different ranges. We observed an interesting result. The recognition accuracy did not decrease as planar objects were away from the self-driving car. The reason was that we only examined the recognition accuracy of traffic signs detected. Some traffic signs with poor image quality and low resolution were missed in the detection phase and did not take into account in the recognition phase. Table IV lists performance of our method. The false positive for traffic signs was successfully eliminated. The average running time of our method was about 33.25 ms per frame on Intel Core i7. Apparently, it had real-time performance. Fig. 14 shows some results of our method in different challenging conditions. It is readily observed from Fig. 14 that traffic signs were correctly detected and recognized, even under bad illumination condition, partial occlusion, and low resolution.

VI. CONCLUSION

In this paper, we propose a novel planar object detection and recognition method based on data association of LIDAR and camera. Planar objects are detected directly in 3D space, instead of usual 2D image plane, through jointly employing relative position, color, LR, and several 3D geometric characteristics of planar objects. For planar object recognition, we present a generic virtual camera-based rectification method so as to synthesize fronto-parallel view of object descriptors in 3D space. Our experimental results show that aggregation of CSA and LR surpasses individual color spaces. The perspective distortion rectification substantially decreases the false recognition error by 45.5%. For a total of 2,130 traffic signs ranging within 100 m under challenging conditions in dynamic cluttered natural scenes, rather than those simply adopted from GTSRB benchmark, the detection rate of the proposed method has up to 95.87%, and the recognition rate even reaches 95.07%. Meanwhile, our comprehensive method achieves real time performance. The average computing time is about 33.25 ms per frame. This suggests that association-based exploitation of multi-modal sensing data in the field of object detection and recognition should be paid more attention. Expectedly, the proposed generic method could also be extended to other applications such as the mapping of traffic signs like [28].

ACKNOWLEDGMENT

The authors are grateful to the anonymous reviewers for their valuable comments, which considerably contributed to improve this paper.

REFERENCES

[1] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 886–893.

[3] L. W. Tsai, J. W. Hsieh, C. H. Chuang, Y. J. Tseng, K. C. Fan, and C. C. Lee, "Road sign detection using eigen colour," *IET Comput. Vis.*, vol. 2, no. 3, pp. 164–177, Sep. 2008.

[4] P. G. 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 Process.*, vol. 88, no. 12, pp. 2943–2955, 2008.

[5] L. Zhou and Z. Deng, "Perspective distortion rectification for planar object based on LIDAR and camera data fusion," in *Proc. IEEE 17th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2014, pp. 270–275.

[6] L. Zhou and Z. Deng, "LIDAR and vision-based real-time traffic sign detection and recognition algorithm for intelligent vehicle," in *Proc. IEEE 17th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2014, pp. 578–583.

[7] C. Nunn, A. Kummert, and S. Müller-Schneiders, "A novel region of interest selection approach for traffic sign recognition based on 3D modelling," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2008, pp. 654–659.

[8] C.-C. Lin and M.-S. Wang, "Road sign recognition with fuzzy adaptive pre-processing models," *Sensors*, vol. 12, no. 5, pp. 6415–6433, 2012.

[9] 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. Transp. Syst.*, vol. 8, no. 2, pp. 264–278, Jun. 2007.

[10] J. F. Khan, S. M. A. Bhuiyan, and R. R. Adhami, "Image segmentation and shape analysis for road-sign detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 83–96, Mar. 2011.

[11] S. Xu, "Robust traffic sign shape recognition using geometric matching," *IET Intell. Transp. Syst.*, vol. 3, no. 1, pp. 10–18, Mar. 2009.

[12] 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, no. 4, pp. 917–930, Dec. 2010.

[13] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, Nov. 1986.

[14] G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points of interest," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 959–973, Aug. 2003.

[15] N. Barnes, A. Zelinsky, and L. S. Fletcher, "Real-time speed sign detection using the radial symmetry detector," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 2, pp. 322–332, Jun. 2008.

[16] N. Barnes, G. Loy, and D. Shaw, "The regular polygon detector," *Pattern Recognit.*, vol. 43, no. 3, pp. 592–602, 2010.

[17] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

[18] X. Baro, S. Escalera, J. Vitria, O. Pujol, and P. Radeva, "Traffic sign recognition using evolutionary adaboost detection and forest-ECOC classification," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 1, pp. 113–126, Mar. 2009.

[19] I. M. Creusen, L. Hazelhoff, and P. H. N. de With, "Color transformation for improved traffic sign detection," in *Proc. IEEE Int. Conf. Image Process.*, Sep./Oct. 2012, pp. 461–464.

[20] G. Overett and L. Petersson, "Large scale sign detection using HOG feature variants," in *Proc. IEEE Intell. Veh. Symp. (IV)*, Jun. 2011, pp. 326–331.

[21] 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, no. 4, pp. 1484–1497, Dec. 2012.

[22] 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 *Proc. Int. Joint Conf. Neural Netw.*, Aug. 2013, pp. 1–8.

[23] M. Liang, M. Yuan, X. Hu, J. Li, and H. Liu, "Traffic sign detection by ROI extraction and histogram features-based recognition," in *Proc. Int. Joint Conf. Neural Netw.*, Aug. 2013, pp. 1–8.

[24] M. Mathias, R. Timofte, R. Benenson, and L. Van Gool, "Traffic sign recognition—How far are we from the solution?" in *Proc. Int. Joint Conf. Neural Netw.*, Aug. 2013, pp. 1–8.

[25] P. Dollár, Z. Tu, P. Perona, and S. Belongie, "Integral channel features," in *Proc. BMVC*, 2009, pp. 91-1–91-11.

[26] G. Wang, G. Ren, Z. Wu, Y. Zhao, and L. Jiang, "A robust, coarse-to-fine traffic sign detection method," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Aug. 2013, pp. 1–5.

[27] H. González-Jorge, B. Riveiro, J. Armesto, and P. Arias, "Geometric evaluation of road signs using radiometric information from laser scanning data," in *Proc. 28th Int. Symp. Autom. Robot. Construct.*, 2011, pp. 1007–1012.

[28] A. Vu, Q. Yang, J. A. Farrell, and M. Barth, "Traffic sign detection, state estimation, and identification using onboard sensors," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, Oct. 2013, pp. 875–880.

[29] T. Li and D. Zhidong, "A new 3D LIDAR-based lane markings recognition approach," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2013, pp. 2197–2202.

[30] B. Soheilian, N. Paparoditis, and B. Vallet, "Detection and 3D reconstruction of traffic signs from multiple view color images," *ISPRS J. Photogramm. Remote Sens.*, vol. 77, pp. 1–20, Mar. 2013.

[31] G. Wang, G. Ren, Z. Wu, Y. Zhao, and L. Jiang, "A hierarchical method for traffic sign classification with support vector machines," in *Proc. Int. Joint Conf. Neural Netw.*, Aug. 2013, pp. 1–6.

[32] Á. González, L. M. Bergasa, and J. J. Yebes, "Text detection and recognition on traffic panels from street-level imagery using visual appearance," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 228–238, Feb. 2014.

[33] J. J. Weinman, Z. Butler, D. Knoll, and J. Feild, "Toward integrated scene text reading," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 375–387, Feb. 2014.

[34] X.-C. Yin, K. Huang, X. Yin, and H.-W. Hao, "Robust text detection in natural scene images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 5, pp. 970–983, May 2014.

[35] C. Merino-Gracia, M. Mirmehdi, J. Sigut, and J. L. González-Mora, "Fast perspective recovery of text in natural scenes," *Image Vis. Comput.*, vol. 31, no. 10, pp. 714–724, 2013.

[36] P. Clark and M. Mirmehdi, "Rectifying perspective views of text in 3D scenes using vanishing points," *Pattern Recognit.*, vol. 36, no. 11, pp. 2673–2686, 2003.

[37] A. B. Cambra and A. C. Murillo, "Towards robust and efficient text sign reading from a mobile phone," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Nov. 2011, pp. 64–71.

[38] M. Bertozzi and A. Broggi, "GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE Trans. Image Process.*, vol. 7, no. 1, pp. 62–81, Jan. 1998.

[39] J. Ruyi, K. Reinhard, V. Tobi, and W. Shigang, "Lane detection and tracking using a new lane model and distance transform," *Mach. Vis. Appl.*, vol. 22, no. 4, pp. 721–737, 2011.

[40] Z. Kim, "Robust lane detection and tracking in challenging scenarios," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 1, pp. 16–26, Mar. 2008.

[41] W. Yao, Z. Deng, and Z. Chen, "A global and local condensation for lane tracking," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, Sep. 2012, pp. 276–281.

[42] J. Greenhalgh and M. Mirmehdi, "Real-time detection and recognition of road traffic signs," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1498–1506, Dec. 2012.

[43] J.-P. Carrasco, A. de la Escalera, and J. M. Armingol, "Recognition stage for a speed supervisor based on road sign detection," *Sensors*, vol. 12, no. 9, pp. 12153–12168, 2012.

[44] F. Zaklouta and B. Stanciulescu, "Real-time traffic-sign recognition using tree classifiers," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1507–1514, Dec. 2012.

[45] P. Sermanet and Y. LeCun, "Traffic sign recognition with multi-scale convolutional networks," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul./Aug. 2011, pp. 2809–2813.

[46] D. Ciresan, U. Meier, J. Masci, and J. Schmidhuber, "Multi-column deep neural network for traffic sign classification," *Neural Netw.*, vol. 32, pp. 333–338, Aug. 2012.

[47] J. Jin, K. Fu, and C. Zhang, "Traffic sign recognition with hinge loss trained convolutional neural networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 1991–2000, Oct. 2014.

[48] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The German traffic sign recognition benchmark: A multi-class classification competition," in *Proc. Int. Joint Conf. Neural Netw.*, Jul./Aug. 2011, pp. 1453–1460.

[49] M. Haloi. (Nov. 2015). "Traffic sign classification using deep inception based convolutional networks." [Online]. Available: https://arxiv.org/abs/1511.02992

[50] R. Unnikrishnan and M. Hebert, "Fast extrinsic calibration of a laser rangefinder to a camera," Robot. Inst., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-RI-TR-05-09, 2005.

[51] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

[52] J. Carreira, R. Caseiro, J. Batista, and C. Sminchisescu, "Semantic segmentation with second-order pooling," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2012, pp. 430–443.

[53] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

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

Zhidong Deng received the B.S. degree in computer science and automation from Sichuan University, Chengdu, China, in 1986 and the Ph.D. degree in computer science and automation from Harbin Institute of Technology, Harbin, China, in 1991.

From 1992 to 1994, he was a Post-Doctoral Researcher with the Computer Science Department, Tsinghua University, Beijing, China, where he became an Associate Professor in 1994. From 1996 to 1997, he served as a Research Associate with The Hong Kong Polytechnic University, Kowloon, Hong Kong. From 2001 to 2003, he was a Visiting Professor with Washington University, St. Louis, MO, USA. He has been a Full Professor with Tsinghua University since 2000. His research interests include artificial intelligence, deep learning, computational neuroscience, computational biology, self-driving car, robotics, wireless sensor network, and virtual reality.

Lipu Zhou received the B.S. degree in computer science from Tianjin University, Tianjin, China, in 2005 and the M.S. degree in computer science from Peking University, Beijing, China, in 2009. He is currently working toward the Ph.D. degree with the Department of Computer Science, Tsinghua University, Beijing. His research interests include pattern recognition, computer vision, and robotics.