Article:

Zhu, H., Yuen, K.-V., Mihaylova, L. orcid.org/0000-0001-5856-2223 et al. (1 more author) (2017) Overview of Environment Perception for Intelligent Vehicles. IEEE Transactions on Intelligent Transportation Systems. ISSN 1524-9050

10.1109/TITS.2017.2658662

Overview of Environment Perception for Intelligent Vehicles

Hao Zhu, Ka-Veng Yuen, Lyudmila Mihaylova and Henry Leung, Fellow, IEEE

***Abstract*—This paper presents a comprehensive literature review on environment perception for intelligent vehicles. The state-of-the-art algorithms and modeling methods for intelligent vehicles are given, with a summary of their pros and cons. A special attention is paid to methods for lane and road detection, traffic sign recognition, vehicle tracking, behavior analysis, and scene understanding. In addition, we provide information about datasets, common performance analysis, and perspectives on future research directions in this area.**

***Index Terms*—Intelligent vehicles, environment perception and modeling, lane and road detection, traffic sign recognition, vehicle tracking and behavior analysis, scene understanding.**

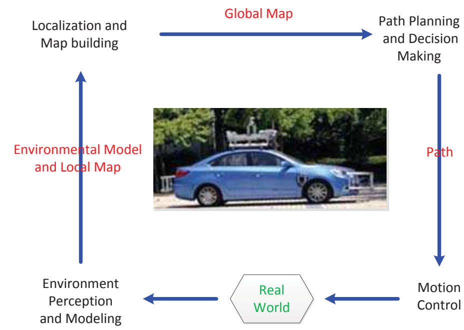


Fig. 1. Four fundamental technologies of intelligent vehicle [3].

I. Introduction

RESEARCH and development on environmental perception, advanced sensing, and intelligent driver assistance systems aim at saving human lives. A wealth of research has been dedicated to the development of driver assistance systems and intelligent vehicles for safety enhancement [1], [2]. For the purposes of safety, comfortability, and saving energy, the field of intelligent vehicles has become a major research and development topic in the world.

Many government agencies, academics, and industries invest great amount of resources on intelligent vehicles, such as Carnegie Mellon University, Stanford University, Cornell University, University of Pennsylvania, Oshkosh Truck Corporation, Peking University, Google, Baidu, and Audi. Furthermore, many challenges have been held to test the capability of intelligent vehicles in a real world environment, such as DARPA Grand Challenge, Future challenge, and European Land-Robot Trial.

This work is jointly supported by the National Natural Science Foundation of China (Grant No. 61301033), by Wenfeng Talents of Chongqing University of Posts and Telecommunications, by the Research Fund of young-backbone university teacher in Chongqing province, by Innovation Team Project of Chongqing Education Committee (Grant No. CXTDX201601019), by the Research Funds of Chongqing University of Posts and Telecommunications (Grant A2012-25 and Grant A2012-81), and by a Mobility grant with China, "Multi-vehicle tracking and classification for intelligent transportation systems" (Reference number E150823) from the UK Royal Society fund.

Hao Zhu is with Automotive Electronics and Embedded System Engineering Research Center, Department of Automation, Chongqing University of Posts and Telecommunications, Chongqing, 400065, P. R. China. (Email: haozhu1982@gmail.com)

Ka-Veng Yuen is with the faculty of Science and Technology, University of Macau, Macao. (Email: kvyuen@umac.mo)

Lyudmila Mihaylova is with the Department of Automatic Control and Systems Engineering, University of Sheffield, Mapping Street, Sheffield S1 3JD, United Kingdom. (Email: l.s.mihaylova@sheffield.ac.uk)

Henry Leung is with the Department of Electrical and Computer Engineering, University of Calgary, 2500 University Drive NW Calgary, Alberta, Canada T2N 1N4. (Email: leungh@ucalgary.ca)

Intelligent vehicles are also called autonomous vehicles, driverless vehicles, or self-driving vehicles. An intelligent vehicle enables a vehicle to operate autonomously by perceiving the environment and implementing a responsive action. It comprises four fundamental technologies: environment perception and modeling, localization and map building, path planning and decision-making, and motion control [3], as shown in Fig. 1.

One main requirement to intelligent vehicles is that they need to be able to perceive and understand their surroundings in real time. It also faces the challenge of processing large amount of data from multiple sensors, such as camera, radio detection and ranging (Radar), and light detection and ranging (LiDAR). A tremendous amount of research has been dedicated to environment perception and modeling over the last decade. For intelligent vehicles, data are usually collected by multiple sensors, such as camera, Radar, LiDAR, and infrared sensors. After pre-processing, various features of objects from the environment, such as roads, lanes, traffic signs, pedestrians and vehicles, are extracted. Both static and moving objects from the environment are being detected and tracked. Some inference can also be performed, such as vehicle behavior and scene understanding. The framework of environment perception and modeling is given in Fig. 2. The main functions of environment perception for intelligent vehicles are based on lane and road detection, traffic sign recognition, vehicle tracking and behavior analysis, and scene understanding. In this paper, we present a comprehensive survey of the state-of-the-art approaches and the popular techniques used in environment perception for intelligent vehicles.

This paper is organized as follows. Vehicular sensors for intelligent vehicles are presented in Section II. In Section III, a survey on lane and road detection is given. The technology on traffic sign recognition is summarized in Section IV. Then, the

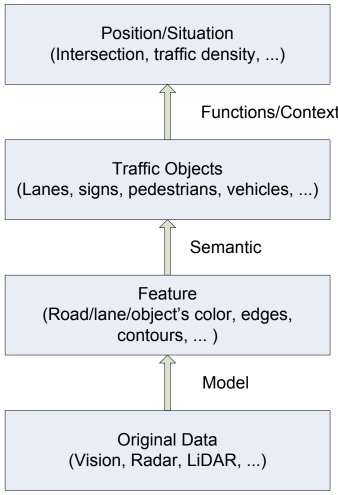


Fig. 2. The framework of environment perception and modeling [3].

survey of vehicle tracking and behavior analysis is presented in Section V. A review of scene understanding technologies is given in Section VI and discussions are presented in Section VII. Finally, conclusions and open questions for future work are presented in Section VIII.

II. Vehicular sensors

A significant progress has been made in the research of intelligent vehicles in recent years. Intelligent vehicles technologies are based on the information of the ego vehicle and its surroundings, such as the lanes, roads, and other vehicles, using the sensors of intelligent vehicles [4], [5]. The sensors in intelligent vehicles can be divided into internal and external sensors. The information of an ego vehicle can be obtained by internal sensors, such as engine temperature, oil pressure, battery and fuel levels. External sensors measure objects of the ego vehicle's surroundings, such as lanes, roads, other vehicles, and pedestrians. External sensors includes Radar, LiDAR, and Vision. In the Internet of vehicles, these sensors can communicate with other vehicles and road infrastructure. The communication among sensors, actuators and controllers is carried out by a controller area network (CAN). It is a serial bus communication protocol developed by Bosch in the early 80s [4], [6].

A. Global Positioning System

The Global Positioning System (GPS) is a space-based navigation system that provides time and location information. However, there is no GPS signal in an indoor environment. Other systems are also under development or in use. Typical examples are the Russian Global Navigation Satellite System, the Indian Regional Navigation Satellite System, the planned European Union Galileo positioning system, and the Chinese BeiDou Navigation Satellite System.

B. Inertial navigation system

The Inertial Navigation System (INS) is a self-contained navigation system. It can be used to track the position and orientation of an object without external references.

C. Radar

Radar is an object detection system. Using the signal of radio waves, it can be used to determine the range, angle, or velocity of objects. Radar is consistent in different illumination and weather conditions. However, measurements are usually noisy and need to be filtered extensively [7].

D. LiDAR

LiDAR has been applied extensively to detect obstacle in intelligent vehicles [8]. It utilizes laser light to detect the distance to objects in a similar fashion as Radar system. Compared with Radar, LiDAR provides a much wider field-of-view and cleaner measurements. However, LiDAR is more sensitive to precipitation [7].

E. Vision

Vision sensors are suitable for intelligent vehicle. Compared with Radar and LiDAR, the raw measurement of vision sensor is the light intensity [9]. Vision sensor can be grouped as camera, lowlight level night vision, infrared night vision, and stereo vision. It can provide a rich data source and a wide field of view.

III. Lane and road detection

Lane and road detection is an active field of research for intelligent vehicles. Some surveys on recent developments in lane and road detection can be found in [10], [11], [12]. We summarized some lane detection systems in Fig. 3. The characteristics of these systems are given as follows:

(1) Lane departure warning: By predicting the trajectory of the host vehicle, a lane departure warning system warns for near lane departure events.

(2) Adaptive cruise control: In the host lane, the adaptive cruise control follows the nearest vehicle with safe headway distance.

(3) Lane keeping or centering: The lane keeping or centering system keeps the host vehicles in the lane center.

(4) Lane change assist: The lane change assisting system requires the host vehicle to change the lane without danger of colliding with any object.

The difficulty of a lane and road detection system is condition diversity, such as lane and road appearance diversity, image clarity, and poor visibility. Therefore, in order to improve the performance of lane and road detect, various algorithms have been proposed according to different assumptions on the structured road. These assumptions are summarized as follows [11]:

(1) The lane/road texture is consistent.

(2) The lane/road width is locally constant.

(3) Road marking follows strict rules for appearance or placement.

(4) The road is a flat plane or follows a strict model for elevation change.

Existing algorithms apply one or more of these assumptions. Furthermore, the lane and road detection system usually consists of three components: pre-processing, feature extraction, and model fitting.

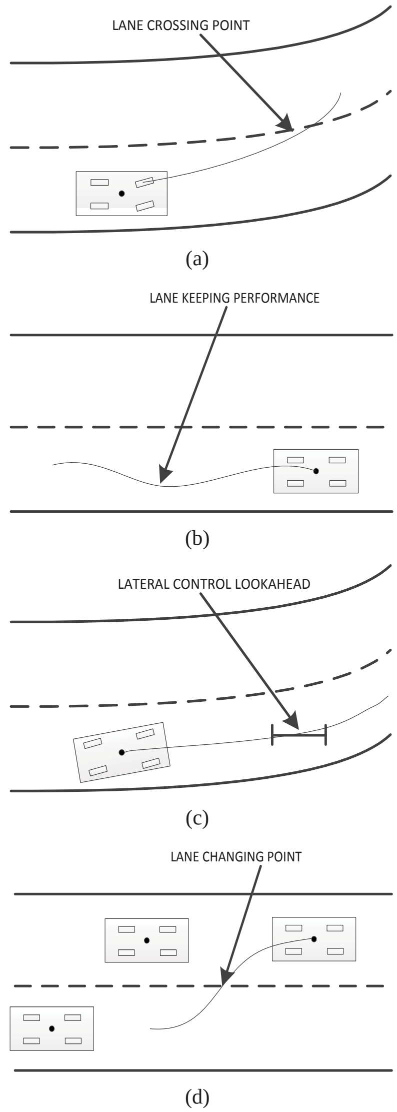


Fig. 3. Some lane detection systems [11]: (a) Lane departure warning (b) Adaptive cruise control (c) Lane keeping or centering (d) Lane change assist.

A. Pre-processing

Pre-processing is important for feature extraction in a lane and road detection system. The objective of pre-processing is to enhance feature of interest and reduce clutter. Pre-processing methods can be categorized into two classes: removing illumination-related effects and pruning irrelevant or misleading image parts [12].

Due to the effects of time of a day and weather conditions, vehicles face illumination-related problems. A robust lane and road detection system should be able to handle the illumination changes, from a sunny day to a rainy night. Information fusion methods from heterogeneous sensors are effective to solve this problem. Other weather-free methods have also been proposed. In [13], a perceptual fog density prediction model was proposed by using natural scene statistics and fog aware statistical features. Observations and modeling of fog were studied by cloud Radar and optical sensors in [14]. Furthermore, the cast shadow is another major illumination-related issue. In a sunny day, the shadow of trees can be casted on the road. Many color space transformations to Hue Saturation Lightness (HSL), Lightness and A and B for the color-opponent dimensions (LAB), and others, which are not affected by illumination changes, were proposed to eliminate the shadow effect [15], [16], [17]. In [18], three different shadow-free images (1D, 2D, and 3D) were investigated according to simple constraints on lighting and cameras.

Vehicles, pedestrians, and other objects can be treated as obstacles for the lane and road detection task. Many methods have been studied for pruning parts of the image. The traditional approach is Regions of Interest (ROI) and feature extraction is performed only on the ROI. A two-stage method, including ROI extraction and lane marker verification, was proposed for robust detection of pedestrian marked lanes at traffic crossings. ROI extraction was performed by using color and intensity information [19]. In [20], a set of Regions of Interests (ROIs) was detected by a Motion Stereo technique to improve the pedestrian detector's performance. Using dense stereo for both ROIs generation and pedestrian classification, a novel pedestrian detection system for intelligent vehicles was presented in [21].

B. Feature extraction

1) Lane feature: In general, a lane feature can be detected by appearance of shape or color [12]. The simplest approach of lane feature extraction assumes that the lane color is known. Using the median local threshold method and a morphological operation, lane markings can be extracted [22]. An adaptive threshold method was proposed to lane markings detection in [23].

Lane shape or color can be used to represent different types of lanes on the road, such as solid line, dashed line, segmented line, and circular reflector. Some colors can be used for lane detection, such as white, yellow, orange and cyan. Other lane feature extraction methods were based on one or more assumptions [11], [23].

The detection methods are based on differences in the appearance of lanes compared with the appearance of the whole road. With this assumption, gradient-based feature extraction methods can be applied. In [11], a steerable filter was developed by computing three separable convolutions to a lane tracking system for robust lane detection.

In [24], [25], [26], the lane marks were assumed to have narrower shape and brighter intensity than their surroundings. Compared with the steerable filter, a method with fixed vertical and horizontal kernels was proposed with the advantage of fast execution and disadvantage of low sensitivity to certain line orientations [24]. In [27], the scale of kernel can be adjusted.

Furthermore, some practical techniques ([28], [29], [30], [31]) were applied using mapping images to remove the perspective effect [12]. However, the inverse perspective mapping (IPM) assumes that the road should be free of obstacles. In order to resolve this problem, a robust method based on multimodal sensor fusion was proposed. Data from a laser range finder and the cameras were fused, so that the mapping was not computed in the regions with obstacles [32].

By zooming into the vanishing point of the lanes, the lane markings will only move on the same straight lines they are on [33]. Based on this fact, a lane feature extraction approach was presented [33], [34].

2) Road feature: Roads are more complicated than lanes as they are not bounded by man-made markings. Under different environments, different cues can be used for road boundaries. For example, curbs can be used for urban roads and barriers can be found in highway roads [12]. Different road features should be extracted in different environments based on different assumptions.

Roads are assumed to have an elevation gap with its surrounding [24], [35], [36], [37]. Stereo vision-based methods were applied to extract the scene structure [35]. In [24], [36], [38], a road markings extraction method is proposed based on three dimensional (3-D) data and a LiDAR system. In [37], a method was proposed to estimate the road region in images captured by vehicle-mounted monocular camera. Using an approach based on the alignment of two successive images, the road region was determined by calculating the differences between the previous and current warped images.

Another method for road feature extraction is based on road appearance and color, where it is assumed that the road has uniform appearance. In [17], a region growing method was applied to road segmentation. In [11], the road appearance constancy was assumed. Some methods based on road color features were considered in [39], [40]. A road-area detection algorithm based on color images was proposed. This algorithm is composed of two modules: boundaries were estimated using the intensity image and road areas were detected using the full color image [40].

Texture is also used as road feature [41], [42]. Using Gabor filters, texture orientations were computed. Then an edge detection technique was proposed for the detection of road boundaries [42]. In order to improve the performance of road detection, methods incorporating prior information have been proposed, such as temporal coherence [43] and shape restrictions [39]. Temporal coherence is averaging the results of consecutive frames. Shape restrictions mean the modeling of the road shape and restricting the possible road area [44]. Using geographical information systems, an algorithm was proposed to estimate the road profile online and prior to building a road map [44].

C. Model fitting

The lane and road model can be categorized into three classes: parametric models, semi-parametric models, and non-parametric models [12].

1) Parametric models: In the case of short range or highway, straight line is the simplest model for path boundaries. For curved roads, parabolic curves and generic circumference arcs were proposed in the bird's eye view. Hyperbolic polynomial curves and parabolic curves were applied to handle more general curved paths in the projective headway view [12].

Many methods were developed to fit the parametric models, such as random sampling consensus (RANSAC), Hough transform, vanishing point, and Kalman filter. RANSAC has the ability to detect outliers and to fit a model to inliers only. It has been investigated for all types of lane and road models. In [29], a Kalman filter-based RANSAC method was found to lane detection. In [45], a parabolic lane model was proposed and the parameters of the lane model were obtained by the randomized Hough transform and genetic algorithm. By assuming a constant path width, vanishing points can be applied as texture for linear boundaries. In [46], the Hough transform and a voting method were utilized to obtain the vanishing points and the road boundaries.

2) Semi-parametric models: In contrast to parametric models, semi-parametric models do not assume a specific global geometry of the path. Therefore, it is necessary to consider the problem of over-fitting and unrealistic path curvature [12].

Splines are piecewise-defined polynomial functions [12]. A cubic-spline curve enables fast fitting since the control points are on the curve. A lane-boundary hypothesis was represented by a constrained cubic-spline curve in [47]. A B-Spline can describe any arbitrary shape using control points. Using a B-Snake to perform lane marking detection, a lane tracking algorithm was proposed in [48]. Cubic Hermite splines ensure the continuity of the extracted features. In [24], the cubic Hermite spline was proposed to extract features, which represents the underlying lane markings. In all spline models [12], the curves were parameterized using a set of control points either on [47] or near [48] the curve. In [49], a Catmull-Rom spline was proposed for lane detection. One major advantage of splines is that small changes in the parameters lead to small changes in the appearance of the curves they model [12].

3) Non-parametric models: Non-parametric models require only continuity but not differentiability of the curve. In [50], an ant colony optimization method was proposed to solve the road-borders detection problem. In [51], a hierarchical Bayesian network method was used to detect off road drivable corridors for autonomous navigation. Considering only constrained relations among points on the left and right lane boundaries, a lane model was proposed in [28]. Table I summarizes various methods for lane and road detection.

D. Evaluation

In order to compare the performance of different methods, it is necessary to establish benchmark and evaluate algorithms for lane and road detection. Some datasets are available, such as the Caltech Lanes dataset and Road markings dataset. Caltech Lanes dataset[1] was built from streets in Pasadena, CA at different times of day. It includes 1225 individual frames. Road markings dataset[2] includes more than 100 original images of diverse road scenes.

We performed experiments on the Caltech Lanes dataset and 1224 frames have been used. Each frame has 640×480 pixels. A RANSAC line fitting-based method [56], a feature pattern-based method [58], a Hough transform-based method[3], and a B-Snake-based method [59] are compared. In the RANSAC

[1] The data set is publicly available at http://www.vision.caltech.edu/malaa/datasets/caltech-lanes/

[2] The data set is publicly available at http://www.lcpc.fr/english/products/image-databases/article/roma-road-markings-1817

[3] http://cn.mathworks.com/help/vision/examples/lane-departure-warning-system.html

TABLE I
VARIOUS METHODS IN LANE AND ROAD DETECTION

Research study	Feature extraction	Model fitting	Accuracy reported in the original paper	Processing time reported in the original paper	Comment
Li et al., 2003 [23]	color information	curve fitting	Not given the numerical performance	N/A	using an adaptive threshold method without being predefined, only consider the color information feature
Lee et al., 2001 [52]	transformation modules	piecewise linear	• Dataset: real expressway • Results: 58 lane changes of total 61 lane changes are successfully detected	Mean alarm triggering time: 0.643 seconds at 15Hz	lane departure warning. sensor fusion algorithm is proposed to estimate lane geometry
Jung et al., 2016 [53]	Hough transform	cubic curve	• Dataset: Borkar's dataset [54], Yoo's dataset [55], and Jung's dataset • Performance evaluate: Lane detection rate • Results: – Borkar's dataset: 95.72% – Yoo's dataset: 97.68% – Jung's dataset: 88.70%	0.117 seconds for a 1-second video	alignment of multiple consecutive scanlines
McCall et al., 2006 [11]	steerable filters	parabolic approximation	• Dataset: scenes from dawn daytime, dusk, and nighttime data • Performance evaluate: mean absolute error in position, standard deviation of error in position, and standard deviation of error in rate of change of lateral position • Results: – mean absolute error in position: 8.2481 cm (average) – standard deviation of error in position: 13.1377 cm (average) – standard deviation of error in rate of change of lateral position: 0.29595 cm/s (average)	N/A	robustness to complex environment
Aly et al., 2008 [56]	filtering with Gaussian kernels	RANSAC spline fitting	• Dataset: 1224 labeled frames containing 4172 marked lanes • Performance evaluate: the correct detection rate, the false positive rate, and the false positive/frame rate • Results: – 2-lanes mode: detecting only the two lane boundaries of the current lane ∗ the correct detection rate: 96.34% ∗ the false positive rate: 11.57% ∗ the false positive/frame rate: 0.191 – all lanes mode: detecting all visible lanes in the image ∗ the correct detection rate: 90.89% ∗ the false positive rate: 17.38% ∗ the false positive/frame rate: 0.592	N/A	generating a top view of the road using inverse perspective mapping
Wang et al., 2012 [33]	global shape information	parallel parabolas on the ground plane	• Dataset: several video sequences • Performance evaluate: ER value means parameter estimation error • Results: – scene 1 ER: 0.11 (parameter a), 0.27 (parameter b1), 0.38 (parameter b2), and 0.71 (parameter c) – scene 2 ER: 0.36 (parameter a), 0.99 (parameter b1), 0.27 (parameter b2), and 0.80 (parameter c) – scene 3 ER: 1.06 (parameter a), 0.41 (parameter b1), 0.19 (parameter b2), and 1.01 (parameter c)	N/A	based on the fact that by zooming into the vanishing point of the lanes, the lane markings will only move on the same straight lines they are on
Yamaguchi et al., 2009 [37]	shape information	road region boundary is represented by a single line	Not given the numerical performance	N/A	road have an elevation gap with its surrounding
Rasmussen et al., 2004 [42]	texture information	vanishing point	Not given the numerical performance	N/A	designed for ill-structured roads
Cui et al., 2016 [57]	color and shape information	white bars of a particular width against a darker background	• Dataset: different urban traffic scenes • Performance evaluate: the mean absolute error of the lane marking detection accuracy • Results: about 3 pixels	N/A	shape registration algorithm between the detected lane markings and a GPS-based road shape prior for localization

line fitting-based method [56], a top view of the image was obtained by IPM, then some selective oriented Gaussian filters were proposed, and a RANSAC spline fitting algorithm was used to detect lanes. In feature pattern-based method [58], lane features were detected by HSV color spaces, and the Hough transform algorithm was proposed to detect the lanes. In Hough transform-based method, the lane features was detected by color thresholding. Then, Hough transform algorithm was applied. In the B-Snake-based method [59], a B-Snake model was proposed to detect the lane boundaries. Lane detection samples by these four algorithms are given in Fig. 4. The measures of correct rate and false positive rate are utilized to evaluate these four algorithms. For simplicity, the results only focus on the two lane boundaries of the current lane. The detection results are given in Table. II. The RANSAC line fitting-based method has the highest correct and false positive rates.

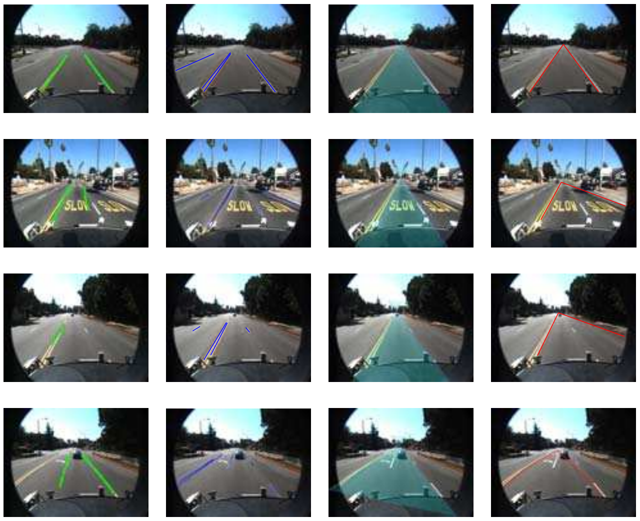


Fig. 4. Lane detection samples by these four algorithms. The first column is the RANSAC line fitting-based method results, the second column is the feature pattern-based method results, the third column is the Hough transform-based method results, and the fourth column is the B-Snake-based method results. The first row is the frame=65, the second row is the frame=67, the third row is the frame=201, and the fourth row is the frame=222.

TABLE II
THE RESULTS OF LANE DETECTION

Method	Correct rate (%)	False positive rate (%)
The RANSAC line fitting-based method	96.39	11.25
The Feature pattern-based method	95.33	6.44
The Hough transform-based method	95.49	8.40
The B-Snake-based method	94.22	10.63

IV. TRAFFIC SIGN RECOGNITION

Traffic signs are used on roads to represent different traffic situations. Most traffic signs are encoded as visual language, which can be quickly interpreted by drivers. Traffic signs can be divided into ideogram-based signs and text-based signs. Ideogram-based signs express the sign meaning through graphics. The text-based signs contain text or other symbols [60]. Therefore, the text-based signs usually contain more information than the ideogram-based signs, but recognition of the text-based signs is more time-consuming. As a result, most research works were dedicated to ideogram-based signs.

Traffic signs give warning to drivers, show the danger and special attention around them, and help them navigate [61]. Some traffic signs of China, United States and Europe are given in Table III. Even though the traffic signs of most countries are conformed to the Vienna Convention, some traffic signs are country-dependent.

The first paper on sign recognition was published in Japan in 1984 [61]. In the following decades, a large volume of research works on traffic sign recognition were conducted. However, as traffic signs are placed in complex outdoor conditions, traffic sign recognition is a nontrivial problem. The influencing factors are summarized as follows [60]:

TABLE III
SOME TRAFFIC SIGNS OF CHINA, AMERICA, AND EUROPE

Road Sign	Interpretation	Color	Shape	Region
	Prohibition	Red, White	Octagon	China
	Prohibition	Red, White	Octagon	Europe
	Prohibition	Red, White	Octagon	U.S.
	Prohibition	Red, White, Black	Circle	China
	Prohibition	Red, White, Black	Circle	Europe
	Prohibition	Red, White, Black	Circle	U.S.
	Warning	Red, White, Black	Circle	China
	Warning	Red, White, Black	Circle	Europe
	Warning	Red, White, Black	Circle	U.S.
	Obligation	Blue, White	Circle	China
	Obligation	Blue, White	Circle	Europe
	Obligation	Blue, White	Circle	U.S.

(1) lighting conditions: such as sunny, shady, rainy, cloudy, and windy;

(2) background: such as freeways, expressways, highways, boulevards, streets, and country roads;

(3) the presence of other objects in the scene: such as cars, pedestrians, and fog;

(4) varied appearance of traffic signs: the size, angle, and position of traffic signs may be different in different images;

(5) long exposure: under long exposure, the color of traffic sign may be faded;

(6) measurement: the image of traffic sign suffers from motion blur using a moving measured platform.

Other noteworthy works on recent developments in traffic sign recognition are given in [60], [62], [63], [64]. The system of traffic sign recognition can be partitioned into three steps: segmentation, feature selection and detection.

A. Segmentation

The purpose of the segmentation process is to obtain the location of the traffic sign. Some of the previous works bypassed this step and started directly with detection [65], [66], [67], [68]. Most of them were based on color and a threshold of images in some color space. The RGB color space is a popular representation modality for images, but the RGB space is sensitive to changes in lighting. As a result, Hue, Saturation, and Intensity (HSI) space and Hue, Saturation, and Value (HSV) space were proposed [69], [70], [71]. HSI color space is similar to human perception of colors [70]. A color appearance model CIECAM97 was used to measure color appearance under various viewing and weather conditions [72], [73]. In [74], the authors demonstrated that a good approach to image segmentation should be normalized with respect to illumination, such as RGB or Ohta normalization.

In addition to these color-based threshold methods, other methods were also proposed. In [75], a cascaded classifier trained with AdaBoost was proposed for traffic sign detection. In [76], a color-based search method was developed with a discrete-color image representation. In [77], it generated

ROIs with possible traffic sign candidates using a biologically motivated attention system.

B. Shape features

The methods of feature selection for traffic sign recognition can be divided into color-based approaches and shape-based approaches. The color of traffic sign includes yellow, black, red, blue, and white while the shape of traffic sign includes triangle, circle, and octagon. The color-based feature selection methods are the same as the methods for segmentation.

For the shape-based methods, the most popular one is based on edges. Using the Canny edge detection or other variations, the edges can be detected [78], [79], [80], [81]. The histogram of oriented gradients (HOG) method was also proposed. A traffic sign detection approach was developed by use of visual saliency and HOG feature learning [82]. Two new HOG features, namely the Single Bin HOG feature and Fishers discriminant analysis linearized HOG feature, were proposed in [83]. The HOG feature with color information was proposed to obtain a more robust feature [67]. A Haar wavelet-like feature for traffic sign is another alternative [68], [84]. Furthermore, some other methods were proposed, such as distance to bounding box [85], fast Fourier transform (FFT) of shape signatures [86], tangent functions [71], simple image patches [70], and combinations of various simple features [77]. In [64], integral channel features and aggregate channel features were proposed to detect U.S. traffic signs.

C. Detection

As the detection step is coupled with the feature extraction step, the choice of detection method depends on the features from the previous stage [63]. The Hough transform was proposed to process the edge features [87], [88]. A radial symmetry detector was proposed for sign detection [89], [90]. In [79], a fast radial symmetry transform was proposed to detect triangular, square or octagonal road signs.

A Support Vector Machine (SVM) was proposed to process the HOG feature [67], [82]. A cascaded classifier was also used to classify HOG features [83], [91]. A multilayer perceptron neural networks was also proposed to detect road sign in [70]. Genetic algorithms for sign road detection was presented in [92]. An overview of traffic sign detection methods is summarized in Table IV.

D. Evaluation

Public traffic sign datasets with ground truth are available at the German TSR Benchmark [4], KUL Belgium Traffic Signs Data set [5], Swedish Traffic Signs Data set [6], RUG Traffic Sign Image Database [7], Stereopolis Database, and LISA Dataset [8]. Results for detection with these data sets and tracking are presented in [63], [64]. The performance evaluation is performed with important measures such as detection rate and false positives per frame.

TABLE V
RESULTS OF COMPARISON TRAFFIC LIGHT RECOGNITION SYSTEMS

Segmentation	Feature	Classification	Precision (%)	Recall (%)
Color thresholding	HSV	Color thresholding	68	62.8
Color thresholding	RGB	Color thresholding	65	68
RGB Color thresholding	HOG	SVM	80.6	82
RGB Color thresholding	Local binary pattern	SVM	83	83.3
RGB Color thresholding	Gabor wavelet	Linear discriminant analysis	79	82.1

We performed experiments on traffic light recognition. The dataset of La Route Automotive at Mines ParisTech [9], Paris, was chosen. We used a set of 200 red lights, 200 green lights, 100 yellow lights for training and 200 red lights, 200 green lights, 100 yellow lights for testing. Each frame has 640×480 pixels. To evaluate traffic light recognition systems, the common measures precision and recall were chosen. They are defined as:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

where TP, FP, and FN denotes true positives, false positives, and false negatives, respectively. The results of the performance comparison of the considered methods are given in Table V. From this table, the method of SVM with local binary patterns is found to provide the the best performance.

V. VEHICLE DETECTION, TRACKING AND BEHAVIOR ANALYSIS

Although there have been substantial developments in the field of vehicle tracking and behavior analysis, the field is still at its infancy stage. The framework of vehicle tracking and behavior analysis is illustrated in Fig. 5. From this figure, some features can be used to perform vehicle detection from numerous vehicular sensors. Vehicles can be tracked by many multi-sensor multi-target tracking algorithms. Then, the behavior of vehicles can be inferred.

A. Vehicle detection

Research on vehicle detection faces the problem of outdoor complex environments, such as illumination and background changes and occlusions. Key developments on vehicle detection were summarized in [9], [101]. The vehicle detection methods can be categorized into appearance-based and motion-based [9].

[4] The data set is publicly available at http://benchmark.ini.rub.de/?section=gtsrb&subsection=dataset

[5] The data set is publicly available at http://www.vision.ee.ethz.ch/~timofter/

[6] The data set is publicly available at http://www.cvl.isy.liu.se/research/datasets/traffic-signs-dataset/

[7] The data set is publicly available at http://www.cs.rug.nl/~imaging/databases/traffic_sign_database/traffic_sign_database.html

[8] The data set is publicly available at http://cvrr.ucsd.edu/LISA/datasets.html

[9] http://www.lara.prd.fr/benchmarks/trafficlightsrecognition

TABLE IV
AN OVERVIEW OF TRAFFIC SIGN DETECTION METHODS

Research study	Segmentation	Features	Detection methods	Real-time implementation	Sign type	Accuracy reported in the original paper	Processing time reported in the original paper
Maldonado-Bascn et al., 2007 [85]	HSI thresholding	distance to bounding box	Shape classification based on linear SVMs	No	Circular, rectangular, triangular, and octagonal	the recognition success probabilities are 93.24%, 67.85%, and 44.90% for the small, medium-sized, and large masks, respectively	1.77s per frame on a 2.2 GHz Pentium 4M, where the frame dimensions are 720*576 pixels
Keller et al., 2008 [93]	Radial symmetry voting	Haar wavelet features	Cascaded classifier	Yes	Rectangular	• Dataset: a videos consists of 16826 frames with 80 different speed sign instances • Performance evaluate: Detection rate, classification rate, and recognition rate • Results: – Detection rate: 98.75% – Classification rate: 97.5% – Recognition rate: 96.25%	N/A
Gao et al., 2006 [72]	LCH thresholding	HOG	Comparison with template	No	Circular and Rectangular	• Dataset: several real road sign databases • Performance evaluate: Detection rate, classification rate, and recognition rate • Results: – Detection rate: 98.75% – Classification rate: 97.5% – Recognition rate: 96.25%	From 0.2 up to 0.7 seconds per image on a PC with Pentium.III
Barnes et al., 2008 [80]	None	Edges	Radial symmetry voting	Yes	Circular	• Dataset: real data collected on public roads around Canberra, Australian • Performance evaluate: Detection rate • Results: 93% successful detection with around 0.5 false positive per sequence	N/A
Gonzalez et al., 2011 [88]	None	Edges	Hough shape detection	No	Any sign	• Dataset: real data collected on Spanish road • Performance evaluate: detection signs ratio and valid signs ratio • Results: detection signs ratio: 98.10% and valid signs ratio: 99.52%	N/A
Xie et al., 2009 [82]	Saliency detection	HOG	SVM	N/A	Circular and square	• Dataset: real images • Results: – Detection ratio: 98.30% (average) – False positive rate: 4.71% (average)	10 seconds for a 400x300 image
Zaklouta et al., 2012 [94]	Color	HOG and distance transforms	K-D trees, random forests, and SVM	Yes	Red color sign	• Dataset: German TSR Benchmark • compare the performance of the k-d trees, the random forests, and the SVM classifiers	N/A
Yuan et al., 2014 [95]	None	Color global and local oriented edge magnitude pattern	SVM	N/A	Any sign	• Dataset: Spanish Traffic Sign Set, German TSR Benchmark, and Authors data set • Results: – Detection results ∗ Precision: 94.45% (Spanish Traffic Sign Set) 87.34% (Authors data set) ∗ Recall: 88.02% (Spanish Traffic Sign Set) 91.60% (Authors data set) – Classification results: accuracy rate ∗ Authors data set:97.24% ∗ German TSR Benchmark: 97.26%	4 frames per seconds, where the frame is 1360*1024
Yuan et al., 2015 [96]	Multithreshold segmentation	Color, saliency, spatial, and contextual	SVM	N/A	Any sign	• Dataset: Spanish Traffic Sign Set, German TSR Benchmark, and Swedish Traffic Signs Data set • Results: – Detection results ∗ Spanish Traffic Sign Set: Precision (91.46%), Recall (96.57%), and F-measure (93.95%) ∗ German TSR Benchmark: Precision (89.65%), Recall (87.84%), and F-measure (88.73%) ∗ Swedish Traffic Signs Data set: Precision (96.30%), Recall (96.21%), and F-measure (96.25%) – Classification results: accuracy rate ∗ Spanish Traffic Sign Set: 100% ∗ German TSR Benchmark: 97.63% ∗ Swedish Traffic Signs Data set: 95.49%	3 frames per seconds, where the frame is 1280*960
Yang et al., 2016 [97]	None	color probability model and HOG features	SVM and convolutional neural network	No	Any sign	• Dataset: German TSR Benchmark and Chinese Traffic Sign Dataset • Results: – Detection results: Recall ∗ German TSR Benchmark: 99.47% ∗ Chinese Traffic Sign Dataset: 99.51% – Classification results: accuracy rate ∗ German TSR Benchmark: 98.24% ∗ Chinese Traffic Sign Dataset: 98.77%	165 ms per image of 1360*800 pixel using a PC with a 4-core 3.7 GHz CPU
Liu et al., 2016 [98]	High contrast region extraction	Voting of neighboring features	Split-flow cascade tree detector and extended sparse representation classification	No	Any sign	• Dataset: German TSR Benchmark • Results: Classification results – Accuracy: 94.81% – False alarm rate: 4.10%	115 ms per image of 1360*800 pixel using a PC with a 4-core 3.19 GHz CPU
Chen et al., 2016 [99]	Saliency model	Color, shape, and spatial location information	AdaBoost and support vector regression	Yes	Any sign	• Dataset: German TSR Benchmark, Spanish Traffic Sign Set, and KUL Belgium Traffic Signs Data set • Performance evaluate: area under curve, precision, and recall • Results: – German TSR Benchmark: area under curve 99.96% – KUL Belgium Traffic Signs Data set: area under curve 97.04% – Spanish Traffic Sign Set: precision 94.52% recall 80.85%	0.05-0.5 seconds per image
Hu et al., 2016 [100]	None	Spatially pooled features and aggregated channel features	Normalized spectral cluster	N/A	Any sign	• Dataset: German TSR Benchmark and KITTI Dataset • Performance evaluate: precision-recall curve and average precision	about 0.6 seconds per image

1) Appearance-based methods: Many appearance features have been proposed to detect vehicles, such as: color, symmetry, edges, HOG features, and Haar-like features. Using color information, vehicles can be segmented from the background. In [102], multivariate decision trees for piecewise linear non-parametric function approximation was used to model the color of a target object from training samples. In [103], an adaptive color model was proposed to detect the color features of the objects around the vehicles. In [104], symmetry as a cue for vehicle detection was studied. In [105], a scheme of symmetry axis detection and filtering based on symmetry constraints was proposed.

More recently, simpler image features (e.g., color, symmetry, and edges) have been transformed to robust feature sets. In [106], vehicles were detected based on their edges of HOG features and symmetrical characteristics. In [107], HOG sym-

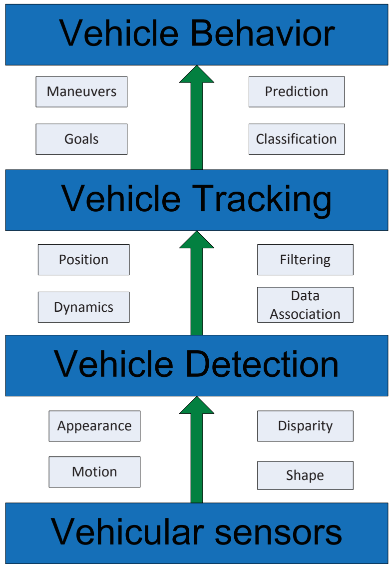


Fig. 5. The framework of vehicle tracking and behavior analysis [9].

metry vectors were proposed to detect vehicles. Haar features are sensitive to vertical, horizontal, and symmetric structures [9]. In [108], Haar and Triangle features were proposed for vehicle detection systems. HOG and Haar features were used to detect vehicle in [109].

After generating the hypothesis of locations of possible vehicles, verification is necessary for the presence of vehicles. SVM and AdaBoost methods are widely used for vehicle detection. A system of integrated HOG feature and SVM classification has been studied in [106], [110]. The combination of edge feature and SVM classification was given in [111]. AdaBoost was proposed to classify the symmetry feature and edge feature in [112] and [113], respectively. The Haar-like feature and AdaBoost classification has been applied to detect vehicles [114], [115].

2) Motion-based methods: In motion-based vehicle detection methods, optical flow and occupancy grids have been widely used. In [116], optical flow was proposed to detect any type of frontal collision. In [117], the optical flow method was applied to a scene descriptor for classifying urban traffic. The optical flow was also proposed to analyze road scenes [9], [118]. Occupancy grids are proposed for scene segmentation and understanding. In [119], occupancy grids were filtered both temporally and spatially. In [120], an occupancy grid tracking solution was proposed based on particles for tracking the dynamic driving environment.

B. Vehicle tracking

The aim of vehicle tracking is to reidentify and measure dynamics and motion characteristics and to predict and estimate the upcoming position of vehicles [9]. The major problems include: measurement error uncertainty, data association, and necessity to fuse efficiently data from multiple sensors.

1) Measurement uncertainty: In the tracking on platform of intelligent vehicles, the measurement noise is the main issue of measurement uncertainty. The Kalman filter is the optimal algorithm in a linear system under Gaussian noises. However, in Radar-based tracking non-Gaussian distributions are often observed [121]. Many methods have been proposed to deal with this non-Gaussian nature of the noises. They can be classified as recursive and batch approaches [122]. The recursive approaches are performed online [123], such as the Masreliez filter, multiple model (MM) approaches, Sequential Monte Carlo (SMC) approaches, and interacting multiple model (IMM) filters. The batch approaches are with offline implementations. In [124], an expectation maximization (EM) algorithm and an IMM algorithm were developed. In [125], a variational Bayesian (VB) algorithm was proposed to estimate the state and parameters in non-Gaussian noise systems.

2) Data association: Data association plays an important role in the multi-sensor multi-target systems. The algorithms of data association can be divided into explicit data association algorithms and implicit data association algorithms [126]. Methods for explicit data association tracking vary widely: from the nearest neighbor (NN) algorithm [127], the multi-hypothesis tracking (MHT) approach [128], the probabilistic data association (PDA) approach [129], to the joint probability data association (JPDA) algorithms [130], [131]. In contrast to explicit data association, implicit data association tracking approaches output a set of object hypotheses in an implicit way, such as particle filter approaches [132], probability hypothesis density (PHD) filtering [133], multi-target multi-Bernoulli (MeMBer) filtering [134], [135], and labeled multi-Bernoulli filtering [136].

3) Fusion: The architectures for sensor data fusion can be divided into centralized and decentralized fusion. Combining the overall system measurements, most of the data and information processing steps are performed at the fusion center in centralized fusion. In [137], a multitarget detection and tracking approach for the case of multiple measurements per target and for an unknown and varying number of targets was proposed. In [138], [139], a joint sensor registration and fusion approach was developed for cooperative driving in intelligent transportation systems. In [140], [141], a multi-sensor and multitarget surveillance system was developed based on solving jointly the registration, data association and data fusion problems.

For the decentralized fusion architecture, the fusion of tracks can be performed at the tracks level. In [142], based on equivalent measurements, a joint sensor registration and track-to-track fusion approach was proposed. In [143], using a pseudo-measurement approach, a joint registration, association and fusion method at distributed architecture was developed. In [144], using information matrix fusion, a track-to-track fusion approach was presented for automotive environment perception. Therefore, many heterogeneous sensor data can be fused for vehicle tracking [145].

4) Joint lane, vehicle tracking, and vehicle classification: The performance of vehicle tracking can be improved by utilizing the lane information and vehicle characteristics to enforce geometric constraints based on the road information. The lane tracking performance can be improved by exploiting vehicle tracking results and eliminating spurious lane marking

TABLE VII
THE RESULTS OF VEHICLE TRACKING

Method	Distance
The tracking by detection method	68.78
The scene flow-based method	18.12
The L1-based tracker method	0.95
The compressive tracking method	2.40

filter responses from the search space [146]. Vehicle characteristics can be used to enhance data association in multi-vehicle tracking. However, few works have explored simultaneous lane, vehicle tracking and classification. A joint lanes and vehicles tracking system was proposed by a PDA filter using camera in [147]. In [148], simultaneous lane and vehicle tracking method using camera was applied to improve vehicle detection. In [146], a synergistic approach to integrated land and vehicle tracking using camera was proposed for driver assistance. In [149], using lane and vehicle information, a maneuvering target was tracked by Radar and image-sensor-based measurement. In [150], an integrated system for vehicle tracking and classification was presented. Table VI highlights key representative works in vehicle detection and tracking.

We performed experiments on KITTI datasets. A total of 278 frames were used and results are obtained with the tracking by detection method [10], the scene flow-based method [158], an L1-based tracking method [159], and a compressive method [160]. These algorithms were applied to tracking single vehicle and comparative results from separate video frames results are given in Fig. 6. A measure of the distance between the true centerline and the estimated centerline is used to evaluate these algorithms. The tracking results are given in Table. VII. It is observed that the L1-based tracker method has the best performance.

C. Behavior analysis

Using the results from the vehicle detection and tracking system, an analysis of the behaviors of other vehicles can be performed. Four characteristics of vehicle behavior are presented, namely context, maneuvers, trajectories and behavior classification [146].

1) Context: The role of context is important for vehicle behavior analysis. In [117], modeling the driving context, the driving environment was classified. In [161], a dynamic visual model was designed to detect critical motions of nearby vehicles. In [154], the behavior of on-coming vehicles was inferred by motion and depth information.

2) Maneuvers: An overtaking monitoring system was presented in [162]. In [163], combining the information provided by Radar and camera, an optical flow method was implemented to detect overtaking vehicles. In [154], an IMM was evaluated for inferring the turning behavior of oncoming vehicles. In [149], a Markov process was constructed to model the behavior of on-road vehicles.

3) Trajectories: In [164], a long-term prediction method of vehicles was proposed. In [165], highway trajectories were clustered using hidden Markov model. In [166], vehicle tracking in combination with a long term motion prediction method was presented.

4) Behavior classification: Efficient models such as Gaussian mixture models, Markov models, and Bayesian networks have been validated for vehicle behavior classification. In [164], the vehicle behavior was classified by a Gaussian mixture model. In [167], the vehicle behavior was modeled by Markov model before their future trajectories was estimated. In [168], the behavior of vehicles was classified by a Bayesian network.

[10] http://www.cvlibs.net/software/trackbydet/

VI. SCENE UNDERSTANDING

Scene understanding is very useful for intelligent vehicles. The procedure of scene understanding can be broadly subdivided to semantic segmentation and scene classification.

A. Semantic segmentation

Semantic segmentation is the first step towards scene understanding. It is mainly based on low-features, such as color, edges, and brightness. The methods for feature selection have been reported in the above subtasks of lane and road detection, traffic sign recognition, and vehicle detection. Furthermore, contextual information is important for semantic segmentation. Using contextual information, the widely applied models of scene understanding can be categorized as graphical models, convolutional networks, cascaded classifiers, and edge detection [169].

1) Graphical models: Markov Random Fields (MRF) and Conditional Random Fields (CRF) are the most popular approaches. In [170], an inference technique was presented for MRF to minimize a unified energy function. In [171], a CRF method was proposed for labeling images. In [172], a hierarchical Dirichlet process was developed to model visual scenes.

2) Convolutional networks: Convolutional networks are widely used [169]. In [173], a convolutional network was trained for scene parsing. In [174], a deep convolutional network in combination with CRF was shown to improve the semantic segmentation performance.

3) Cascaded classifiers: In [175], a different architecture for combining multiple classifiers into a cascaded classifier model was performed for scene understanding. In [176], a feedback enabled cascaded classification model was developed to jointly optimize several subtasks in scene understanding. Since each classifier is considered in series, the training process of a cascaded classifier model is substantially simpler than convolutional networks [169].

4) Edge detection: Various unsupervised methods have been applied for edge detection. In [177], an efficient edge detector was introduced and able to learn different edge patterns. In [169], a contextual hierarchical model was used to distinguish between “patches centered on an edge pixel” and “patches centered on a non-edge pixel”.

TABLE VI
SOME REPRESENTATIVE WORKS IN VEHICLE DETECTION AND TRACKING

Research study	Detection	Tracking	Accuracy reported in the original paper	Processing time reported in the original paper	Sensor type
Liu et al.,2007 [151]	Contour symmetry	Template matching	• Dataset: real-world video sequences • Performance evaluate: tracking continuity • Results: – Urban road 97.9% – Narrow road 93.6% – High way 100% – Mountain road 80.5%	25 frames per second using PC machine (Pentium IV 2.8GHz and 512M of RAM)	Monocular vision
Haselhoff et al.,2009 [152]	Haar-like features	Kalman filter	• Dataset: real-world video sequences • Performance evaluate: overlap factor • Results: around 90%	N/A	Monocular vision
Sivaraman et al.,2009 [153]	Haar-like features	Particle filter	• Dataset: LISA-Q Front FOV data set • Performance evaluate: – True positive rate: 95%(set 1) 91.7%(set 2) 99.8%(set 3) – False detection rate: 6.4%(set 1) 25.5%(set 2) 8.5%(set 3) – Average false positives/frame: 0.29(set 1) 0.39(set 2) 0.28(set 3) – Average true positives/frame: 4.2(set 1) 1.14(set 2) 3.17(set 3) – False positives/vehicle: 0.06(set 1) 0.31(set 2) 0.09(set 3)	N/A	Monocular vision
Barth et al.,2010 [154]	Motion and depth information	Kalman filter, IMM	• Dataset: synthetic stereo images • Performance evaluate: root mean squared corner error • Results: average 0.49m	80 ms per frame on an Intel Quad Core processor,	Stereo vision
Danescu et al.,2011 [120]	Occupancy grids	Particle filter	• Dataset: synthetic stereo images • Performance evaluate: speed and orientation estimation accuracy • Results: quickly converge toward the ground truth and stable	40 ms per frame on an Intel Core 2 Duo processor at 2.1 GHz	Stereo vision
Lim et al.,2013 [155]	HOG	Markov chain Monte Carlo particle filter	• Dataset: real world stereo images • Performance evaluate: the MOTP score and the MOTA score • Results: – MOTA: 94.0% (Scene1) 89.8% (Scene2) – MOTP: 68.2% (Scene1) 69.3% (Scene2)	N/A	Stereo vision
Fortin et al.,2015 [156]	Point sets	Sequential Monte Carlo methods	• Dataset: synthetic data and real data from an IBEO LD Automotive scanning laser telemeter • Performance evaluate: estimation accuracies, cardinality accuracy, and OSPA-T distance	N/A	LiDAR
Chavez-Garcia et al.,2016 [157]	Point sets and HOG	Markov chain Monte Carlo	• Dataset: two datasets from urban areas and two datasets from highways • Performance evaluate: vehicle mis-classifications • Results: 5.4% (Highway 1) 4.5% (Highway 2) 10.2% (Urban 1) 10.3% (Urban 2)	N/A	Radar, camera, and LiDAR

B. *Scene classification*

As most scenes are composed of entities in a highly variable layout, scene classification is an important problem for environment perception. In the literature, scene classification has been focused on binary problems, such as distinguishing indoor from outdoor scenes. Inspired by the way how the human perception system works, numerous efforts have been devoted to classify a large number of scene categories. The

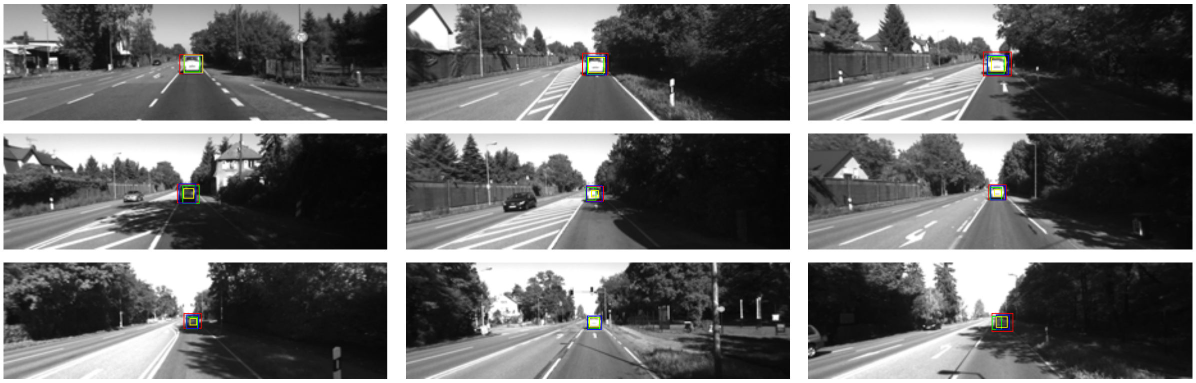


Fig. 6. The samples tracking results by these four algorithms. The red box is the results of the tracking by detection method, the green box is the results of the scene flow-based method, the blue box is the results of the L1-based tracker method, and the yellow box is the results of the compressive tracking method.

most popular method is the "bag-of-features". It represents images as orderless collections. The features can be extracted by the Scale-invariant feature transform (SIFT) [178] or the HOG method [179]. In [180], the SIFT was proposed to extract visual features and these features were encoded to a Fisher kernel framework for scene classification. In [181], a convolutional neural network (CNN) was proposed to perform scene classification.

Recently, the model of "bag-of-semantics" was proposed. In this model, an image is extracted as a semantic feature space. It has the capability to perform a spatially localized semantic mapping. In [182], a set of classifiers for individual semantic attributes was trained for object classification. In [183], a high-level image representation encoding object appearance and spatial location information was proposed. In [184], a semantic Fisher vector, which is an extension of the Fisher vector to bag-of-semantics, was applied to classify image patches. Table VIII highlights some representative works in scene understanding.

C. Datasets

Datasets are publicly available for scene understanding, such as: the KITTI Vision dataset [11] and the CityScapes segmentation benchmark [12]. Furthermore, some researchers have annotated KITTI images with semantic labels, such as Jose Alvarez[13], Philippe Xu[14], Lubor Ladicky[15], Fatma Güney[16], Sunando Sengupta[17], German Ros[18] and Abhijit Kundu[19]. Some other datasets were captured using Kinect or similar devices [20].

VII. Discussion

From Section II to Section VI, the subtasks of environment perception are given. In general, the process of segmentation, detection, classification, and tracking can be treated as a piece or whole framework for each subtasks. In this framework feature extraction represents a key challenge. We summarize some representative feature cues in Table IX.

For the lane and road detection, different methods rely on different assumptions. Many features have been investigated. A better solution is fusion of multiple features to achieve reasonable performance. For model fitting, straight line is a simple and effective model for short range roads or highway. Splines are good models for curved roads.

In the traffic sign recognition systems, most segmentation methods rely on color or shape information. The detection approach is dependent on the choice of features. In most traffic sign recognition system, color, shape, and structural features are typically considered. The Hough transform and its derivatives have been widely applied for the purposes of object detection. The SVM, neural networks, and cascaded classifiers have been used to classify the traffic sign using HOG or Haar wavelet features. The performance of traffic sign recognition can be improved, for instance by creating a combined feature space and by using the map information. A drawback of traffic sign recognition systems is due to the lack of public datasets for training and testing.

For the purpose of vehicle tracking, the tradeoff between accuracy and computational complexity is of primary importance. The motion and depth information add another layer that can help improving the accuracy. However, most current vehicle tracking algorithms do not use information about the fact of vehicle driving on the road. We believe that a joint framework of lane detection, vehicle classification, and vehicle tracking are in the heart of the next generation of intelligent vehicles.

The methods of deep learning [174] and semantic Fisher vector [184] have a big potential in scene understanding.

VIII. Conclusion

The development of environment perception and modeling technology is one of the key aspects for intelligent vehicles. This paper presents an overview of the state of the arts of environment perception and modeling technology. First, the pros and cons of vehicular sensors are presented. Next, popular

[11]http://www.cvlibs.net/datasets/kitti/
[12]https://www.cityscapes-dataset.com/news/
[13]https://rsu.forge.nicta.com.au/people/jalvarez/research_bbdd.php
[14]https://www.hds.utc.fr/x̃uphilip/dokuwiki/en/data
[15]https://www.inf.ethz.ch/personal/ladickyl/
[16]http://www.cvlibs.net/projects/displets/
[17]http://www.robots.ox.ac.uk/t̃vg/projects/SemanticUrbanModelling/index.php
[18]http://adas.cvc.uab.es/s2uad/
[19]http://www.cc.gatech.edu/ãkundu7/projects/JointSegRec/
[20]http://www.cs.ucl.ac.uk/staff/M.Firman/RGBDdatasets/

TABLE VIII
SOME REPRESENTATIVE WORKS IN SCENE UNDERSTANDING

Research study	Indoor or outdoor	Characteristics	Accuracy reported in the original paper	Processing time reported in the original paper	Comment
Lee et al., 2009 [185]	indoor	3D cuboid from a single image	• Dataset: 54 images of indoor scenes • Results: – Classified correctly: 81% of the pixels – Less than 30% misclassified pixels: 76% of the images – 10% misclassified pixels: 44%	N/A	Built on edges and image segments as features; most of them rely on the Manhattan world assumption
Hedau et al., 2009 [186]	indoor	3D occupancy grids	• Dataset: 308 indoor images • Results: – Error for box layout estimation: 21.2%(pixel error), 6.3%(corner error) – Pixel error for surface label estimation: 18.3%(pixel error)	N/A	Modeling the room clutter using 3D occupancy grids
Pero et al., 2012 [187]	indoor	cuboids	• Dataset: UCB dataset and Hedau test • Results: Average error on room layout estimation – UCB dataset: 18.4% – Hedau test: 16.3%	N/A	Modeling the geometry and location of specific objects using cuboids
Delaitre et al., 2012 [188]	indoor	context from observing people	• Dataset: 146 time-lapse videos • Results: – Pose estimation ∗ Average precision: about 50% ∗ Percentage of Correct Parts: about 56% – Semantic labeling of objects ∗ Average precision: 43 ± 4.3(appearance, location and person features combined)	N/A	Functional object description to recognize objects by the way people interact with them
Zeisl et al., 2011 [189]	indoor	structure prior	vertical and non-vertical structures coexist	O(nL) for an m*n image and considering L depth values	A natural assumption of bounded open space for building interiors: (i) by parallel ground and ceiling planes, and (ii) by vertical wall elements
Hoiem et al., 2007 [190]	outdoor	3D layout of a scene from a single image	Results: Average accuracy of varying levels of spatial support and four types of cues are given	N/A	Separating the ground from the vertical structures and the sky
Hoiem et al., 2008 [191]	outdoor	tree-structured modeling (check)	• Dataset: 422 outdoor images from the LabelMe dataset • Performance evaluate: – Object Detection : ROC curves for car and pedestrian detection – Horizon Estimation : median absolute error	N/A	Framework for placing local object detection in the context of the overall 3D scene by modeling the interdependence of objects
Sudderth et al., 2008 [172]	outdoor	hierarchical Dirichlet process	• Dataset: 613 street scenes and 315 pictures of office scenes • Performance evaluate: segmentation results, ROC curves	N/A	Couples topic models originally developed for text analysis with spatial transformations; consistently accounts for geometric constraints
Geiger et al., 2012 [192]	outdoor	Prior, Vehicle Tracklet, Vanishing points, semantic scene label, scene flow, occupancy grid	• Dataset: 113 intersections scenes • Performance evaluate: Topology Accuracy, Location Error, Street Orient, Road Area Overlap, Tracklet Accuracy, Lane Accuracy, Object Orient error, and Object Detection Accuracy	N/A	Understanding traffic situations at intersection

TABLE IX
SOME REPRESENTATIVE FEATURE CUES

Type	Cues	Characteristics or advantages	Limitations
Appearance-based	Color	Simple calculations	Affect from background with same color
	Edge	Low computational load	• Affect from outlier • Difficult to choose the threshold
	Symmetry	• Vertically symmetrical in multi-vehicle in road • Useful to ROI estimation	• High computational load • Difficult to choose the threshold
	Corners	Find corners with edge pixels	Not useful in complex environment
	Multiple features	• More precision • Robustness	High computational load
Motion-based	Occupancy grids	Computing static obstacles and free space	High computational load
	Optical flow	Matching pixels or feature points between two frames	• Not useful to slow moving objects • High computational load

modeling methods and algorithms of lane and road detection, traffic sign recognition, vehicle tracking and behavior analysis, and scene understanding are reviewed. Public datasets and codes of environment perception and modeling technology are also described.

Current challenges for environment perception and modeling technology are due to the complex outdoor environments and the need of efficient methods for their perception in real time. The changeable lighting and weather conditions, and the complex backgrounds, especially the presence of occluding objects still represent significant challenges to intelligent vehicles. Furthermore, it is very important to recognize road in the off-road environment.

As many algorithms have been proposed for environment perception, it is necessary to establish more benchmarks and performance evaluations on environment perception for intelligent vehicles.

Since environment perception and modeling technology stage is the link with the work of localization and map building, path planning and decision-making, and motion control, the next step is to develop the entire system.

REFERENCES

[1] H. Cheng, N. Zheng, X. Zhang, J. Qin, and H. van de Wetering, “Interactive road situation analysis for driver assistance and safety warning systems: Framework and algorithms,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 1, pp. 157–167, 2007.

[2] L. Li, J. Song, F.-Y. Wang, W. Niehsen, and N.-N. Zheng, “IVS 05: new developments and research trends for intelligent vehicles,” *IEEE Intelligent Systems*, vol. 20, no. 4, pp. 10–14, 2005.

[3] H. Cheng, *Autonomous Intelligent Vehicles Theory, Algorithms, and Implementation*. Springer, 2011.

[4] C. Lundquist, “Sensor fusion for automotive applications,” Ph.D. dissertation, Linköping University, Linköping, 2011.

[5] N.-E. E. Faouzi, H. Leung, and A. Kurian, “Data fusion in intelligent transportation systems: Progress and challenges : A survey,” *Information Fusion*, vol. 12, no. 1, pp. 4 – 10, 2011.

[6] S. Tuohy, M. Glavin, C. Hughes, E. Jones, M. M. Trivedi, and L. Kilmartin, “Intra-vehicle networks: A review,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 534–545, 2015.

[7] S. Sivaraman, “Learning, modeling, and understanding vehicle surround using multi-modal sensing,” Ph.D. dissertation, Uinversity of California, San Diego, 2013.

[8] J. Levinson, J. Askeland, J. Becker, J. Dolson, D. Held, S. Kammel, J. Kolter, D. Langer, O. Pink, V. Pratt, M. Sokolsky, G. Stanek, D. Stavens, A. Teichman, M. Werling, and S. Thrun, “Towards fully autonomous driving: Systems and algorithms,” in *Proceedings of IEEE Intelligent Vehicles Symposium*, June 2011, pp. 163–168.

[9] S. Sivaraman and M. M. Trivedi, “Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 4, pp. 1773–1795, 2013.

[10] V. Kastrinaki, M. Zervakis, and K. Kalaitzakis, “A survey of video processing techniques for traffic applications,” *Image and Vision Computing*, vol. 21, no. 4, pp. 359 – 381, 2003.

[11] J. C. McCall and M. M. Trivedi, “Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 20–37, 2006.

[12] A. Bar Hillel, R. Lerner, D. Levi, and G. Raz, “Recent progress in road and lane detection: A survey,” *Machine Vision and Applications*, vol. 25, no. 3, pp. 727–745, 2014.

[13] L. K. Choi, J. You, and A. Bovik, “Referenceless prediction of perceptual fog density and perceptual image defogging,” *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3888–3901, 2015.

[14] Y. Li, P. Hoogeboom, and H. Russchenberg, “Observations and modeling of fog by cloud radar and optical sensors,” in *Proceedings of the 11th European Radar Conference*, Oct 2014, pp. 521–524.

[15] H.-Y. Cheng, B.-S. Jeng, P.-T. Tseng, and K.-C. Fan, “Lane detection with moving vehicles in the traffic scenes,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 4, pp. 571–582, 2006.

[16] I. Katramados, S. Crumpler, and T. Breckon, “Real-time traversable surface detection by colour space fusion and temporal analysis,” in *Computer Vision Systems*, ser. Lecture Notes in Computer Science, M. Fritz, B. Schiele, and J. Piater, Eds. Springer Berlin Heidelberg, 2009, vol. 5815, pp. 265–274.

[17] J. Alvarez, A. Lopez, and R. Baldrich, “Shadow resistant road segmentation from a mobile monocular system,” in *Pattern Recognition and Image Analysis*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2007, vol. 4478, pp. 9–16.

[18] G. Finlayson, S. Hordley, C. Lu, and M. Drew, “On the removal of shadows from images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 59–68, 2006.

[19] M. C. Le, S. L. Phung, and A. Bouzerdoum, “Pedestrian lane detection for assistive navigation of blind people,” in *Proceedings of the 21st International Conference on Pattern Recognition*, 2012, pp. 2594–2597.

[20] M. Bertozzi, L. Bombini, P. Cerri, P. Medici, P. Antonello, and M. Miglietta, “Obstacle detection and classification fusing radar and vision,” in *Proceedings of IEEE Intelligent Vehicles Symposium*, June 2008, pp. 608–613.

[21] C. Keller, M. Enzweiler, M. Rohrbach, D. Fernandez Llorca, C. Schnorr, and D. Gavrila, “The benefits of dense stereo for pedestrian detection,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1096–1106, 2011.

[22] G. Liu, S. Li, and W. Liu, “Lane detection algorithm based on local feature extraction,” in *Proceedings of Chinese Automation Congress*, Nov 2013, pp. 59–64.

[23] Q. Li, N. Zheng, and H. Cheng, “An adaptive approach to lane markings detection,” in *Proceedings of IEEE Intelligent Transportation Systems*, vol. 1, Oct 2003, pp. 510–514.

[24] A. Huang, D. Moore, M. Antone, E. Olson, and S. Teller, “Finding multiple lanes in urban road networks with vision and lidar,” *Autonomous Robots*, vol. 26, no. 2-3, pp. 103–122, 2009.

[25] R. Labayrade, J. Douret, J. Laneurit, and R. Chapuis, “A reliable and robust lane detection system based on the parallel use of three algorithms for driving safety assistance.” *IEICE Transactions on Information and Systems*, vol. 89-d, no. 7, pp. 2092–2100, 2006.

[26] S. J. Wu, H. H. Chiang, J. W. Perng, C. J. Chen, B. F. Wu, and T. T. Lee, “The heterogeneous systems integration design and implementation for lane keeping on a vehicle,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 2, pp. 246–263, 2008.

[27] B. Upcroft, M. Moser, A. Makarenko, D. Johnson, A. Donikian, A. Alempijevic, R. Fitch, W. Uther, E. I. GrAtli, J. Biermeyer, H. Gonzalez, T. Templeton, V. P. Srini, and J. Sprinkle, “DARPA urban challenge technical paper: Sydney-Berkeley driving team,” University of Sydney; University of Technology, Sydney; University of California, Berkeley, Tech. Rep., June 2007. [Online]. Available: http://chess.eecs.berkeley.edu/pubs/379.html

[28] R. Jiang, R. Klette, T. Vaudrey, and S. Wang, “New lane model and distance transform for lane detection and tracking,” *Lecture Notes in Computer Science*, vol. 5702, pp. 1044–1052, 2009.

[29] A. Borkar, M. Hayes, and M. T. Smith, “Robust lane detection and tracking with ransac and kalman filter,” in *Proceedings of the 16th IEEE International Conference on Image Processing*, 2009, pp. 3261–3264.

[30] Y. Alon, A. Ferencz, and A. Shashua, “Off-road path following using region classification and geometric projection constraints,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2006, pp. 689–696.

[31] M. Nieto, L. Salgado, F. Jaureguizar, and J. Arrspide, “Robust multiple lane road modeling based on perspective analysis,” in *Proceedings of the 15th IEEE International Conference on Image Processing*, 2008, pp. 2396–2399.

[32] M. Oliveira, V. Santos, and A. D. Sappa, “Multimodal inverse perspective mapping,” *Information Fusion*, vol. 24, pp. 108 – 121, 2015.

[33] Y. Wang, N. Dahnoun, and A. Achim, “A novel system for robust lane detection and tracking,” *Signal Processing*, vol. 92, no. 2, pp. 319–334, 2012.

[34] Y. Wang, N. Dahnoun, and A. Achim, “A novel lane feature extraction algorithm based on digital interpolation,” in *Proceedings of the 17th European Signal Processing Conference*, 2009, pp. 480–484.

[35] R. Danescu and S. Nedevschi, “Probabilistic lane tracking in difficult road scenarios using stereo vision,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 2, pp. 272–282, 2009.

[36] K. Takagi, K. Morikawa, T. Ogawa, and M. Saburi, “Road environment recognition using on-vehicle lidar,” in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2006, pp. 120–125.

[37] K. Yamaguchi, A. Watanabe, T. Naito, and Y. Ninomiya, “Road region estimation using a sequence of monocular images,” in *Proceedings of the 19th International Conference on Pattern Recognition*, 2009, pp. 1–4.

[38] Y. Yu, J. Li, H. Guan, F. Jia, and C. Wang, “Learning hierarchical features for automated extraction of road markings from 3-D mobile

lidar point clouds," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 2, pp. 709–726, 2015.

[39] M. A. Sotelo, F. J. Rodriguez, and L. Magdalena, "Virtuous: Vision-based road transportation for unmanned operation on urban-like scenarios." *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 2, pp. 69–83, 2004.

[40] Y. He, H. Wang, and B. Zhang, "Color-based road detection in urban traffic scenes." *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 4, pp. 309–318, 2004.

[41] C. Rasmussen and T. Korah, "On-vehicle and aerial texture analysis for vision-based desert road following," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 66–66.

[42] C. Rasmussen, "Grouping dominant orientations for ill-structured road following," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2004, pp. 470 – 477.

[43] C. Tan, T. Hong, T. Chang, and M. Shneier, "Color model-based real-time learning for road following," in *Proceedings of IEEE Intelligent Transportation Systems Conference*, 2006, pp. 939–944.

[44] J. M. Alvarez, A. M. Lopez, T. Gevers, and F. Lumbreras, "Combining priors, appearance, and context for road detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 3, pp. 1168–1178, 2014.

[45] S. A. T. M. Samadzadegan, F., "Automatic lane detection in image sequences for vision based navigation purposes," in *ISPRS image engineering and vision metrology*, 2006.

[46] G. Zhang, N. Zheng, C. Cui, Y. Yan, and Z. Yuan, "An efficient road detection method in noisy urban environment," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2009, pp. 556–561.

[47] Z. W. Kim, "Robust lane detection and tracking in challenging scenarios." *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 1, pp. 16–26, 2008.

[48] Y. Wang, E. K. Teoh, and D. Shen, "Lane detection and tracking using B-snake," *Image and Vision Computing*, vol. 22, no. 4, pp. 269–280, 2004.

[49] W. Yue, D. Shen, and E. K. Teoh, "Lane detection using spline model," *Pattern Recognition Letters*, vol. 21, no. 8, pp. 677–689, 2000.

[50] A. Broggi and S. Cattani, "An agent based evolutionary approach to path detection for off-road vehicle guidance." *Pattern Recognition Letters*, vol. 27, no. 11, pp. 1164 – 1173, 2006.

[51] A. V. Nefian and G. R. Bradski, "Detection of drivable corridors for off-road autonomous navigation." in *Proceedings of International Conference on Image Processing*, 2006, pp. 3025–3028.

[52] S. Lee and W. Kwon, "Robust lane keeping from novel sensor fusion," in *Proceedings of IEEE International Conference on Robotics and Automation*, vol. 4, 2001, pp. 3704–3709.

[53] S. Jung, J. Youn, and S. Sull, "Efficient lane detection based on spatiotemporal images," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 1, pp. 289–295, 2016.

[54] A. Borkar, M. Hayes, and M. T. Smith, "A novel lane detection system with efficient ground truth generation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 1, pp. 365–374, 2012.

[55] H. Yoo, U. Yang, and K. Sohn, "Gradient-enhancing conversion for illumination-robust lane detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1083–1094, 2013.

[56] M. Aly, "Real time detection of lane markers in urban streets," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2008, pp. 7–12.

[57] D. Cui, J. Xue, and N. Zheng, "Real-time global localization of robotic cars in lane level via lane marking detection and shape registration," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 1039–1050, 2016.

[58] X. Shi, B. Kong, and F. Zheng, "A new lane detection method based on feature pattern," in *Proceedings of International Congress on Image and Signal Processing*, 2009, pp. 1–5.

[59] Y. Wang, E. K. Teoh, and D. Shen, "Lane detection using b-snake," in *Proceedings of International Conference on Information Intelligence and Systems*, 1999, pp. 438–443.

[60] M.-Y. Fu and Y.-S. Huang, "A survey of traffic sign recognition," in *Proceedings of International Conference on Wavelet Analysis and Pattern Recognition*, July 2010, pp. 119–124.

[61] H. Fleyeh and M. Dougherty, "Road and traffic sign detection and recognition," in *Proceedings 10th EWGT Meet/ 16th Mini-EURO Conference*, 2005, pp. 644 – 653.

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

[63] A. Mogelmose, M. M. Terivedi, and T. B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance system: Perspectives and survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, pp. 1484–1497, 2012.

[64] A. Mogelmose, D. Liu, and M. M. Trivedi, "Detection of U.S. traffic signs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3116–3125, 2015.

[65] R. Belaroussi and J.-P. Tarel, "Angle Vertex and Bisector Geometric Model for Triangular Road Sign Detection," in *Workshop on Applications of Computer Vision*, 2009, pp. 1–7.

[66] F. Larsson and M. Felsberg, *Using Fourier Descriptors and Spatial Models for Traffic Sign Recognition*. Springer Berlin Heidelberg, 2011.

[67] I. M. Creusen, R. G. J. Wijnhoven, E. Herbschleb, and P. H. N. De With, "Color exploitation in hog-based traffic sign detection," in *Proceedings of IEEE International Conference on Image Processing*, 2010, pp. 2669 – 2672.

[68] X. Bar, S. Escalera, J. Vitri, O. Pujol, and P. Radeva, "Traffic sign recognition using evolutionary Adaboost detection and forest-ECOC classification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 1, pp. 113–126, 2009.

[69] W. J. Kuo and C. C. Lin, "Two-stage road sign detection and recognition," in *Proceedings of IEEE International Conference on Multimedia and Expo*, 2007, pp. 1427 – 1430.

[70] Y.-Y. Nguwi and A. Kouzani, "Detection and classification of road signs in natural environments," *Neural Computing and Applications*, vol. 17, no. 3, pp. 265–289, 2008.

[71] S. Xu, "Robust traffic sign shape recognition using geometric matching," *IET Intelligent Transport Systems*, vol. 3, no. 1, pp. 10–18, 2009.

[72] X. W. Gao, L. Podladchikova, D. Shaposhnikov, K. Hong, and N. Shevtsova, "Recognition of traffic signs based on their colour and shape features extracted using human vision models," *Journal of Visual communication & image representation*, vol. 17, pp. 675–685, 2006.

[73] X. Gao, K. Hong, P. Passmore, L. Podladchikova, and D. Shaposhnikov, "Colour vision model-based approach for segmentation of traffic signs," *Eurasip Journal on Image and Video Processing*, vol. 2008, no. 5, pp. 1–7, 2008.

[74] H. Gmez-Moreno, S. Maldonado-Bascn, P. Gil-Jimnez, and S. Lafuente-Arroyo, "Goal evaluation of segmentation algorithms for traffic sign recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 4, pp. 917 – 930, 2010.

[75] D. Deguchi, M. Shirasuna, K. Doman, I. Ide, and H. Murase, "Intelligent traffic sign detector: Adaptive learning based on online gathering of training samples," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2011, pp. 72 – 77.

[76] A. Ruta, Y. Li, and X. Liu, "Real-time traffic sign recognition from video by class-specific discriminative features," *Pattern Recognition*, vol. 43, no. 4, pp. 416–430, 2010.

[77] R. Kastner, T. Michalke, T. Burbach, J. Fritsch, and C. Goerick, "Attention-based traffic sign recognition with an array of weak classifiers," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2010, pp. 333–339.

[78] S. Houben, "A single target voting scheme for traffic sign detection," in *Proceedings of Intelligent Vehicles Symposium*, 2011, pp. 124 – 129.

[79] G. Loy and N. Barnes, "Fast shape-based road sign detection for a driver assistance system," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2004, pp. 70–75 vol.1.

[80] N. Barnes, A. Zelinsky, and L. S. Fletcher, "Real-time speed sign detection using the radial symmetry detector," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 2, pp. 322–332, 2008.

[81] M. Meuter, C. Nunny, S. Gormer, S. Muller-Schneiders, and A. Kummert, "A decision fusion and reasoning module for a traffic sign recognition system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1126–1134, 2011.

[82] Y. Xie, L. F. Liu, C. R. Li, and Y. Y. Qu, "Unifying visual saliency with hog feature learning for traffic sign detection," *Proceedings of IEEE Intelligent Vehicles Symposium*, pp. 24–29, 2009.

[83] G. Overett and L. Petersson, "Large scale sign detection using hog feature variants," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2011, pp. 326–331.

[84] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler, "A system for traffic sign detection, tracking, and recognition using color, shape, and motion information," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2005, pp. 255–260.

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

[86] P. G. Jimnez, S. M. Bascn, 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, no. 12, pp. 2943–2955, 2008.

[87] F. Moutarde, A. Bargeton, A. Herbin, and L. Chanussot, "Robust on-vehicle real-time visual detection of American and European speed limit signs, with a modular traffic signs recognition system," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2007, pp. 1122–1126.

[88] A. Gonzalez, M. A. Garrido, D. F. Llorca, M. Gavilan, J. P. Fernandez, P. F. Alcantarilla, I. Parra, F. Herranz, L. M. Bergasa, and M. A. Sotelo, "Automatic traffic signs and panels inspection system using computer vision," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 485 – 499, 2011.

[89] N. Barnes and A. Zelinsky, "Real-time radial symmetry for speed sign detection," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2004, pp. 566–571.

[90] G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points of interest," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 959–973, 2003.

[91] N. Pettersson, L. Petersson, and L. Andersson, "The histogram feature - a resource-efficient weak classifier," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2008, pp. 678–683.

[92] H. Liu, D. Liu, and J. Xin, "Real-time recognition of road traffic sign in motion image based on genetic algorithm," in *Proceedings of International Conference on Machine Learning and Cybernetics*, 2002, pp. 83 – 86.

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

[94] F. Zaklouta and B. Stanciulescu, "Real-time traffic-sign recognition using tree classifiers," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1507–1514, 2012.

[95] X. Yuan, X. Hao, H. Chen, and X. Wei, "Robust traffic sign recognition based on color global and local oriented edge magnitude patterns," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1466–1477, 2014.

[96] X. Yuan, J. Guo, X. Hao, and H. Chen, "Traffic sign detection via graph-based ranking and segmentation algorithms," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 12, pp. 1509–1521, 2015.

[97] Y. Yang, H. Luo, H. Xu, and F. Wu, "Towards real-time traffic sign detection and classification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 7, pp. 2022–2031, 2016.

[98] C. Liu, F. Chang, Z. Chen, and D. Liu, "Fast traffic sign recognition via high-contrast region extraction and extended sparse representation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 1, pp. 79–92, 2016.

[99] T. Chen and S. Lu, "Accurate and efficient traffic sign detection using discriminative AdaBoost and support vector regression," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 6, pp. 4006–4015, 2016.

[100] Q. Hu, S. Paisitkriangkrai, C. Shen, A. van den Hengel, and F. Porikli, "Fast detection of multiple objects in traffic scenes with a common detection framework," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 1002–1014, 2016.

[101] S. Zehang, B. George, and M. Ronald, "On-road vehicle detection: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 694 – 711, 2006.

[102] S. D. Buluswar and B. A. Per, "Color machine vision for autonomous vehicles," *Engineering Applications of Artificial Intelligence*, vol. 11, no. 2, pp. 245–256, 1998.

[103] D. Guo, T. Fraichard, M. Xie, and C. Laugier, "Color modeling by spherical influence field in sensing driving environment," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, 2000, pp. 249–254.

[104] A. Kuehnle, "Symmetry-based recognition of vehicle rears," *Pattern Recognition Letters*, vol. 12, no. 91, pp. 249 – 258, 1991.

[105] Y. Du and N. P. Papanikolopoulos, "Real-time vehicle following through a novel symmetry-based approach," in *Proceedings of IEEE International Conference on Robotics and Automation*, 1997, pp. 3160–3165 vol.4.

[106] S. S. Teoh and T. Bräunl, "Symmetry-based monocular vehicle detection system," *Machine Vision and Applications*, vol. 23, no. 5, pp. 831–842, 2012.

[107] M. Cheon, W. Lee, C. Yoon, and M. Park, "Vision-based vehicle detection system with consideration of the detecting location," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1243–1252, 2012.

[108] A. Haselhoff and A. Kummert, "A vehicle detection system based on Haar and triangle features," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2009, pp. 261–266.

[109] P. Negri, X. Clady, S. M. Hanif, and L. Prevost, "A cascade of boosted generative and discriminative classifiers for vehicle detection," *Eurasip Journal on Advances in Signal Processing*, vol. 2008, no. 1, p. 136, 2008.

[110] S. Sivaraman and M. M. Trivedi, "Active learning for on-road vehicle detection: a comparative study," *Machine Vision and Applications*, vol. 25, no. 3, pp. 599–611, 2014.

[111] K. Zu, "Real time obstacle detection and tracking based on constrained delaunay triangulation," in *Proceedings of International IEEE Conference on Intelligent Transportation Systems*, 2006, pp. 548 – 553.

[112] T. Liu, N. Zheng, L. Zhao, and H. Cheng, "Learning based symmetric features selection for vehicle detection," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2005, pp. 124 – 129.

[113] A. Khammari, F. Nashashibi, Y. Abramson, and C. Laurgeau, "Vehicle detection combining gradient analysis and AdaBoost classification," in *Proceedings of IEEE Intelligent Transportation Systems*, 2005, pp. 66–71.

[114] I. Kallenbach, R. Schweiger, G. Palm, and O. Lohlein, "Multi-class object detection in vision systems using a hierarchy of cascaded classifiers," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2006, pp. 383–387.

[115] T. T. Son and S. Mita, "Car detection using multi-feature selection for varying poses," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2009, pp. 507–512.

[116] E. Martinez, M. Diaz, J. Melenchon, J. A. Montero, I. Iriondo, and J. C. Socoro, "Driving assistance system based on the detection of head-on collisions," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2008, pp. 913–918.

[117] A. Geiger and B. Kitt, "Object flow: A descriptor for classifying traffic motion," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2010, pp. 287–293.

[118] S. Bota and S. Nedevschi, "Tracking multiple objects in urban traffic environments using dense stereo and optical flow," in *Proceedings of the 14th International IEEE Conference on Intelligent Transportation Systems*, 2011, pp. 791–796.

[119] M. Perrollaz, J. D. Yoder, A. Negre, A. Spalanzani, and C. Laugier, "A visibility-based approach for occupancy grid computation in disparity space," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1383 – 1393, 2012.

[120] R. Danescu, F. Oniga, and S. Nedevschi, "Modeling and tracking the driving environment with a particle-based occupancy grid," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1331–1342, 2011.

[121] W. R. Wu, "Target tracking with glint noise," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 29, no. 1, pp. 174–185, 1993.

[122] H. Zhu, H. Leung, and Z. He, "A variational Bayesian approach to robust sensor fusion based on Student-t distribution," *Information Sciences*, vol. 221, no. 1, pp. 201–214, 2013.

[123] I. Bilik and J. Tabrikian, "MMSE-based filtering in presence of non-Gaussian system and measurement noise," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 3, pp. 1153–1170, 2010.

[124] D. Huang and H. Leung, "Maximum likelihood state estimation of semi-Markovian switching system in non-Gaussian measurement noise," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 1, pp. 133–146, 2010.

[125] H. Zhu, H. Leung, and Z. He, "State estimation in unknown non-Gaussian measurement noise using variational Bayesian technique," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 49, no. 4, pp. 2601–2614, 2013.

[126] C. Stiller, F. P. León, and M. Kruse, "Information fusion for automotive applications c an overview," *Information Fusion*, vol. 12, no. 4, pp. 244– 252, 2011.

[127] S. Deb, M. Yeddanapudi, K. Pattipati, and Y. Bar-Shalom, "A generalized S-D assignment algorithm for multisensor-multitarget state estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 2, pp. 523 – 538, 1997.

[128] S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 5–18, 2004.

[129] Y. Bar-Shalom and E. Tse, "Tracking in a cluttered environment with probabilistic data association," *Automatica*, vol. 11, no. 5, pp. 451–460, 1975.

[130] T. E. Fortmann, Y. Bar-Shalom, and M. Scheffe, "Sonar tracking of multiple targets using joint probabilistic data association," *IEEE Journal of Oceanic Engineering*, vol. 8, no. 3, pp. 173–184, 1983.

[131] D. Musicki and R. Evans, "Joint integrated probabilistic data association: JIPDA," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 40, no. 3, pp. 1093–1099, 2004.

[132] S. Särkkä, A. Vehtari, and J. Lampinen, "Rao-blackwellized particle filter for multiple target tracking," *Information Fusion*, vol. 8, no. 1, pp. 2–15, 2007.

[133] B. N. Vo and W. K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4091–4104, 2006.

[134] B. Ristic, B. T. Vo, B. N. Vo, and A. Farina, "A tutorial on Bernoulli filters: Theory, implementation and applications," *IEEE Transactions on Signal Processing*, vol. 61, no. 13, pp. 3406 – 3430, 2013.

[135] A. Gning, B. Ristic, and L. Mihaylova, "Bernoulli particle/box-particle filters for detection and tracking in the presence of triple measurement uncertainty," *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2138–2151, 2012.

[136] S. Reuter, B. T. Vo, B. N. Vo, and K. Dietmayer, "The labeled multi-Bernoulli filter," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3246–3260, 2014.

[137] T. De Laet, H. Bruyninckx, and J. De Schutter, "Shape-based online multitarget tracking and detection for targets causing multiple measurements: Variational Bayesian clustering and lossless data association," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2477–2491, 2011.

[138] W. Li and H. Leung, "Simultaneous registration and fusion of multiple dissimilar sensors for cooperative driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 2, pp. 84–98, 2004.

[139] D. Huang and H. Leung, "An expectation-maximization-based interacting multiple model approach for cooperative driving systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 206–228, 2005.

[140] S. Chen, H. Leung, and l. Boss, "A maximum likelihood approach to joint registration, association and fusion for multi-sensor multi-target tracking," in *Proceedings of International Conference on Information Fusion*, 2009, pp. 686–693.

[141] Z. Li, S. Chen, H. Leung, and E. Bosse, "Joint data association, registration, and fusion using EM-KF," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 2, pp. 496–507, 2010.

[142] N. N. Okello and S. Challa, "Joint sensor registration and track-to-track fusion for distributed trackers," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 40, no. 3, pp. 808–823, 2004.

[143] H. Zhu, H. Leung, and K. V. Yuen, "A joint data association, registration, and fusion approach for distributed tracking," *Information Sciences*, vol. 324, pp. 186–196, 2015.

[144] M. Aeberhard, S. Schlichthärle, N. Kaempchen, and T. Bertram, "Track-to-track fusion with asynchronous sensors using information matrix fusion for surround environment perception," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1717–1726, 2012.

[145] R. O. Chavez-Garcia and O. Aycard, "Multiple sensor fusion and classification for moving object detection and tracking," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10, 2015.

[146] S. Sivaraman and M. M. Trivedi, "Integrated lane and vehicle detection, localization, and tracking: A synergistic approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 906–917, 2013.

[147] S. Y. Hung, Y. M. Chan, B. F. Lin, L. C. Fu, P. Y. Hsiao, and S. S. Huang, "Tracking and detection of lane and vehicle integrating lane and vehicle information using PDAF tracking model," in *Proceedings of the 12th International IEEE Conference on Intelligent Transportation Systems*, 2009, pp. 1–6.

[148] S. Sivaraman and M. M. Trivedi, "Improved vision-based lane tracker performance using vehicle localization," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2010, pp. 676–681.

[149] Y. Chen, V. P. Jilkov, and X. R. Li, "Multilane-road target tracking using radar and image sensors," *IEEE Transactions on Aerospace and Electronic Systems*,, vol. 51, no. 1, pp. 65–80, 2015.

[150] S. Battiato, G. M. Farinella, A. Furnari, G. Puglisi, A. Snijders, and J. Spiekstra, "An integrated system for vehicle tracking and classification," *Expert Systems with Applications*, vol. 8, no. 21, pp. 7263–7275, 2015.

[151] W. Liu, X. Wen, B. Duan, and H. Yuan, "Rear vehicle detection and tracking for lane change assist," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2007, pp. 252–257.

[152] A. Haselhoff and A. Kummert, "An evolutionary optimized vehicle tracker in collaboration with a detection system," in *Proceedings of the 12th International IEEE Conference on Intelligent Transportation Systems*, 2009, pp. 1–6.

[153] S. Sivaraman and M. M. Trivedi, "A general active-learning framework for on-road vehicle recognition and tracking," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 267–276, 2010.

[154] A. Barth and U. Franke, "Tracking oncoming and turning vehicles at intersections," in *Proceedings of International IEEE Conference on Intelligent Transportation Systems*, 2010, pp. 861–868.

[155] Y. C. Lim, D. Kim, and C. H. Lee, "MCMC particle filter-based vehicle tracking method using multiple hypotheses and appearance model," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2013, pp. 1131–1136.

[156] B. Fortin, R. Lherbier, and J. C. Noyer, "A model-based joint detection and tracking approach for multi-vehicle tracking with Lidar sensor," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 1883–1895, 2015.

[157] R. O. Chavez-Garcia and O. Aycard, "Multiple sensor fusion and classification for moving object detection and tracking," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 2, pp. 525–534, 2016.

[158] P. Lenz, J. Ziegler, A. Geiger, and M. Roser, "Sparse scene flow segmentation for moving object detection in urban environments," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2011, pp. 926–932.

[159] C. Bao, Y. Wu, H. Ling, and H. Ji, "Real time robust L1 tracker using accelerated proximal gradient approach," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 1830–1837.

[160] L. Z. K. Zhang and M. H. Yang, "Real-time compressive tracking," in *Proceedings of European Conference on Computer Vision*, 2012, pp. 866 – 879.

[161] S. Cherng, C. Y. Fang, C. P. Chen, and S. W. Chen, "Critical motion detection of nearby moving vehicles in a vision-based driver-assistance system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 1, pp. 70–82, 2009.

[162] J. D. Alonso, E. Ros Vidal, A. Rotter, and M. Muhlenberg, "Lane-change decision aid system based on motion-driven vehicle tracking," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 5, pp. 2736 – 2746, 2008.

[163] F. Garcia, P. Cerri, A. Broggi, A. De, la Escalera, and J. M. Armingol, "Data fusion for overtaking vehicle detection based on radar and optical flow," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2012, pp. 494–499.

[164] J. Wiest, U. Hoffken, M.and Kresel, and K. Dietmayer, "Probabilistic trajectory prediction with Gaussian mixture models." in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2012, pp. 141–146.

[165] S. Sivaraman, B. Morris, and M. M. Trivedi, "Learning multi-lane trajectories using vehicle-based vision," in *Proceedings of IEEE International Conference on Computer Vision*, 2011, pp. 2070 – 2076.

[166] C. Hermes, J. Einhaus, M. Hahn, C. WöHler, and F. Kummert, "Vehicle tracking and motion prediction in complex urban scenarios," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2010, pp. 26 – 33.

[167] T. Gindele, S. Brechtel, and R. Dillmann, "A probabilistic model for estimating driver behaviors and vehicle trajectories in traffic environments," in *Proceedings of the 13th IEEE International Conference on Intelligent Transportation Systems*, 2010, pp. 1625–1631.

[168] D. Kasper, G. Weidl, T. Dang, G. Breuel, A. Tamke, A. Wedel, and W. Rosenstiel, "Object-oriented Bayesian networks for detection of lane change maneuvers," in *Proceedings of IEEE Intelligent Vehicles Symposium*, 2011, pp. 673–678.

[169] M. Seyedhosseini and T. Tasdizen, "Semantic image segmentation with contextual hierarchical models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 5, pp. 951 – 964, 2016.

[170] S. Gould, R. Fulton, and D. Koller, "Decomposing a scene into geometric and semantically consistent regions," *Proceedings of the 12th IEEE International Conference on Computer Vision*, vol. 30, no. 2, pp. 1–8, 2009.

[171] X. He, R. Zemel, and M. Carreira-Perpindn, "Multiscale conditional random fields for image labeling," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2004, pp. II–695–II–702 Vol.2.

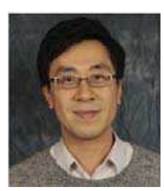
[172] E. B. Sudderth, A. Torralba, W. T. Freeman, and A. S. Willsky, "Describing visual scenes using transformed objects and parts," *International Journal of Computer Vision*, vol. 77, pp. 291–330, 2008.

[173] D. Grangier, L. Bottou, and R. Collobert, "Deep convolutional networks for scene parsing," in *ICML 2009 Deep Learning Workshop*, 2009.
[174] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," in *Proceedings of International Conference on Learning Representations*, 2015.
[175] G. Heitz, S. Gould, A. Saxena, and D. Koller, "Cascaded classification models: Combining models for holistic scene understanding." *Advances in Neural Information Processing Systems*, pp. 641–648, 2008.
[176] C. Li, A. Kowdle, A. Saxena, and T. Chen, "Toward holistic scene understanding: Feedback enabled cascaded classification models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1394–1408, 2012.
[177] P. Dollar and C. Zitnick, "Structured forests for fast edge detection," in *Proceedings of IEEE International Conference on Computer Vision*, 2013, pp. 1841–1848.
[178] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
[179] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, 2005, pp. 886–893.
[180] J. Snchez, F. Perronnin, T. Mensink, and J. Verbeek, "Image classification with the fisher vector: Theory and practice," *International Journal of Computer Vision*, vol. 105, no. 3, pp. 222–245, 2013.
[181] B. Zhou, A. L. Garcia, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using Places database," *Advances in Neural Information Processing Systems*, vol. 1, pp. 487–495, 2015.
[182] Y. Su and F. Jurie, "Improving image classification using semantic attributes," *International Journal of Computer Vision*, vol. 100, no. 1, pp. 1–10, 2010.
[183] L. J. Li, H. Su, Y. Lim, and F. F. Li, "Object bank: An object-level image representation for high-level visual recognition," *International Journal of Computer Vision*, vol. 107, no. 1, pp. 20–39, 2014.
[184] M. Dixit, S. Chen, D. Gao, N. Rasiwasia, and N. Vasconcelos, "Scene classification with semantic fisher vectors," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2974–2983.
[185] D. C. Lee, M. Hebert, and T. Kanade, "Geometric reasoning for single image structure recovery," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 2136–2143.
[186] V. Hedau, D. Hoiem, and D. Forsyth, "Recovering the spatial layout of cluttered rooms," in *Proceedings of IEEE International Conference on Computer Vision*, 2009, pp. 1849–1856.
[187] L. Del Pero, J. Bowdish, D. Fried, B. Kermgard, E. Hartley, and K. Barnard, "Bayesian geometric modelling of indoor scenes," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2719–2726.
[188] V. Delaitre, D. F. Fouhey, I. Laptev, J. Sivic, A. Gupta, and A. A. Efros, "Scene semantics from long-term observation of people," *Lecture Notes in Computer Science*, vol. 7577, no. 1, pp. 284–298, 2012.
[189] B. Zeisl, C. Zach, and M. Pollefeys, "Stereo reconstruction of building interiors with a vertical structure prior," in *Proceedings of International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, 2011, pp. 366–373.
[190] D. Hoiem, A. A. Efros, and M. Hebert, "Recovering surface layout from an image," *International Journal of Computer Vision*, vol. 75, no. 1, pp. 151–172, 2007.
[191] D. Hoiem, A. A. Efros, and M. Hebert, "Putting objects in perspective," *International Journal of Computer Vision*, vol. 80, no. 1, pp. 3–15, 2008.
[192] A. Geiger, M. Lauer, C. Wojek, C. Stiller, and R. Urtasun, "3D traffic scene understanding from movable platforms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 5, pp. 1012–1025, 2014.

Hao Zhu received his Ph.D. degree in computer science from Chongqing University, China, in 2012. He is now an associate professor of Department of Automation, Chongqing University of Posts and Telecommunications, Chongqing, China. His research interests include intelligent vehicles, target tracking, statistical signal processing, and sensor fusion.

Ka-Veng Yuen received his Ph.D. degree from California Institute of Technology, USA, in 2002. He is the Registrar and a professor of the University of Macau. His research interests include Bayesian inference, sensor network, and reliability of engineering system.

Lyudmila Mihaylova (M'98, SM'2008) is Professor of Signal Processing and Control at the Department of Automatic Control and Systems Engineering at the University of Sheffield, United Kingdom. Her research is in the areas of machine learning and autonomous systems with various applications such as navigation, surveillance and sensor network systems. She has given a number of talks and tutorials, including the plenary talk for the IEEE Sensor Data Fusion 2015 (Germany), invited talks University of California, Los Angeles, IPAMI Traffic Workshop 2016 (USA), IET ICWMMN 2013 in Beijing, China. Dr. Mihaylova is an Associate Editor of the IEEE Transactions on Aerospace and Electronic Systems and of the Elsevier Signal Processing Journal. She was elected in March 2016 as a president of the International Society of Information Fusion (ISIF). She is on the board of Directors of ISIF and a Senior IEEE member. She was the general co-chair IET Data Fusion & Target Tracking 2014 and 2012 Conferences, Program co-chair for the 19th International Conference on Information Fusion, Heidelberg, Germany.



Henry Leung is a professor of the Department of Electrical and Computer Engineering of the University of Calgary. Before joining U of C, he was with the Department of National Defence (DND) of Canada as a defence scientist. His main duty there was to conduct research and development of automated surveillance systems, which can perform detection, tracking, identification and data fusion automatically as a decision aid for military operators. His current research interests include big data analytic, chaos and nonlinear dynamics, information fusion, machine learning, signal and image processing, robotics and internet of things. He has published extensively in the open literature on these topics. He has over 200 journal papers and 180 refereed conference papers. Dr. Leung has been the associate editor of various journals such as the IEEE Circuits and Systems Magazine, International Journal on Information Fusion, IEEE Signal Processing Letters, IEEE Trans. Circuits and Systems. He has also served as guest editors for the special issue "Intelligent Transportation Systems" for the International Journal on Information Fusion and "Cognitive Sensor Networks" for the IEEE Sensor Journal. He is the topic editor on "Robotic Sensors" of the International Journal of Advanced Robotic Systems. He is the editor of the Springer book series on "Information Fusion and Data Science". He is a fellow of the IEEE.