Looking at Vehicles on the Road: A Survey of Vision-Based Vehicle Detection, Tracking, and Behavior Analysis

Sayanan Sivaraman, *Member, IEEE*, and Mohan Manubhai Trivedi, *Fellow, IEEE*

***Abstract*—This paper provides a review of the literature in on-road vision-based vehicle detection, tracking, and behavior understanding. Over the past decade, vision-based surround perception has progressed from its infancy into maturity. We provide a survey of recent works in the literature, placing vision-based vehicle detection in the context of sensor-based on-road surround analysis. We detail advances in vehicle detection, discussing monocular, stereo vision, and active sensor–vision fusion for on-road vehicle detection. We discuss vision-based vehicle tracking in the monocular and stereo-vision domains, analyzing filtering, estimation, and dynamical models. We discuss the nascent branch of intelligent vehicles research concerned with utilizing spatiotemporal measurements, trajectories, and various features to characterize on-road behavior. We provide a discussion on the state of the art, detail common performance metrics and benchmarks, and provide perspective on future research directions in the field.**

***Index Terms*—Computer vision, intelligent vehicles, machine learning, object detection, object tracking.**

I. Introduction

IN the United States, tens of thousands of drivers and passengers die on the roads each year, with most fatal crashes involving more than one vehicle [1]. Research and development efforts in advanced sensing, environmental perception, and intelligent driver assistance systems seek to save lives and reduce the number of on-road fatalities. Over the past decade, there has been significant research effort dedicated to the development of intelligent driver assistance systems and autonomous vehicles, which is intended to enhance safety by monitoring the on-road environment.

In particular, the on-road detection of vehicles has been a topic of great interest to researchers over the past decade [2]. A variety of sensing modalities has become available for on-road vehicle detection, including radar, lidar, and computer vision. Imaging technology has immensely progressed in recent years. Cameras are cheaper, smaller, and of higher quality than ever before. Concurrently, computing power has dramatically increased. Furthermore, in recent years, we have seen the emergence of computing platforms geared toward parallelization, such as multicore processing and graphical processing units (GPUs). Such hardware advances allow computer vision approaches for vehicle detection to pursue real-time implementation.

With advances in camera sensing and computational technologies, advances in vehicle detection using monocular vision, stereo vision, and sensor fusion with vision have been an extremely active research area in the intelligent vehicles community. On-road vehicle tracking has been also extensively studied. It is now commonplace for research studies to report the ability to reliably detect and track on-road vehicles in real time, over extended periods [3], [5]. Theoretical, practical, and algorithmic advances have opened up research opportunities that seek higher level of semantic interpretation of on-road vehicle behavior. The aggregate of this spatiotemporal information from vehicle detection and tracking can be used to identify maneuvers and to learn, model, and classify on-road behavior.

Fig. 1 depicts the use of vision for on-road interpretation. At the lowest level, various motion and appearance cues are used for on-road vehicle detection. One level up, detected vehicles are associated across frames, allowing for vehicle tracking. Vehicle tracking measures the dynamics of the motion of detected vehicles. At the highest level, an aggregate of spatiotemporal features allows for characterization of vehicle behavior, recognition of specific maneuvers, behavior classification, and long-term motion prediction. Examples of work in this nascent area include prediction of turning behavior [3], prediction of lane changes [6], and modeling typical on-road behavior [4].

In this paper, we provide a review of vision-based vehicle detection, tracking, and on-road behavior analysis. We concentrate our efforts on works published since 2005, referring the reader to [2] for earlier works. We place vision-based vehicle detection in the context of on-road environmental perception, briefly detailing complimentary modalities that are commonly used for vehicle detection, namely, radar and lidar. We then review vision-based vehicle detection, commenting on monocular vision, stereo vision, monocular–stereo combination, and sensor-fusion approaches to vision-based vehicle detection. We discuss vehicle tracking using vision, detailing image-plane and 3-D techniques for modeling, measuring, and filtering vehicle dynamics on the road. We then discuss the emerging body of literature geared toward analysis of vehicle behavior using spatiotemporal cues, including modeling, learning, classification,

Manuscript received December 6, 2012; revised April 10, 2013; accepted June 1, 2013. Date of publication July 18, 2013; date of current version November 26, 2013. This work was supported in part by Audi Research; by the Volkswagen Electronics Research Laboratory, Belmont, CA; and by the University of California Discovery Program. The Associate Editor for this paper was S. S. Nedevschi.

The authors are with the Laboratory for Intelligent and Safe Automobiles, University of California, San Diego, La Jolla, CA 92093 USA (e-mail: ssivaram@ucsd.edu; mtrivedi@ucsd.edu).

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

Digital Object Identifier 10.1109/TITS.2013.2266661

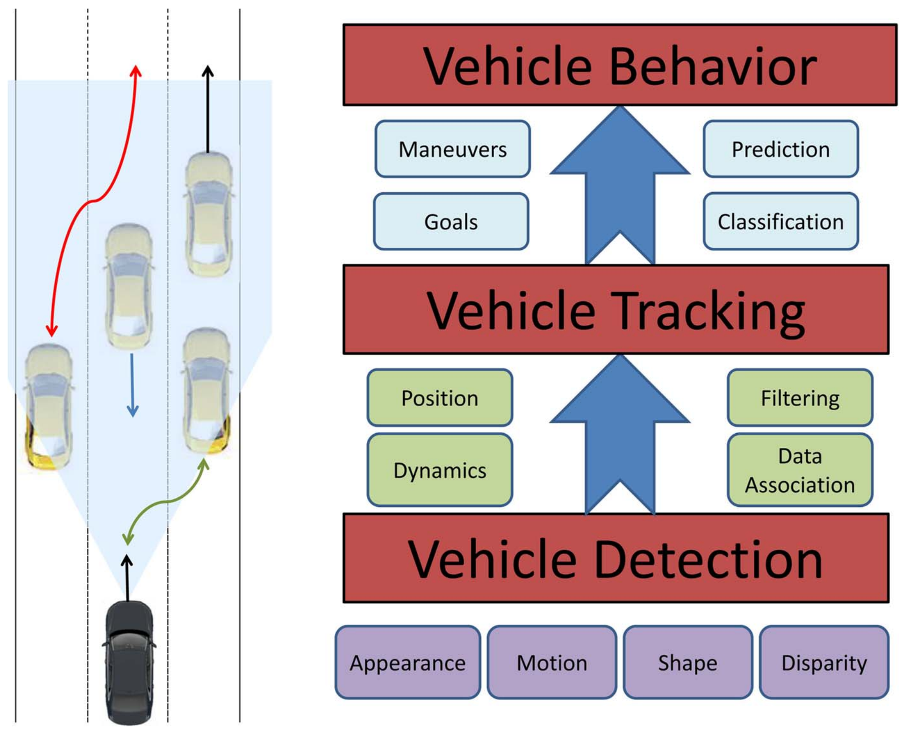

Fig. 1. Illustrating the ascending levels of vision for semantic interpretation of the on-road environment. At the lowest level, features such as appearance, disparity, motion, and size are used to detect vehicles in images and video. One level up, data association, temporal coherence, and filtering are used for tracking, to reidentify and measure the dynamic parameters and to estimate the positions of the vehicles. At the highest level, an aggregate of spatiotemporal features is used to learn, model, classify, and predict the behavior and goals of other vehicles on the road. This area of research includes identification of specific maneuvers [3] and modeling typical on-road behavior [4].

and prediction of vehicle maneuvers and goals. We provide our insights and perspectives on future research directions in vision-based vehicle detection, tracking, and behavior analysis.

II. On-Road Environmental Perception

While the focus of this paper lies in vision-based vehicle detection, it is pertinent to include a brief treatment of complimentary modalities currently used in on-road vehicle detection. We discuss general sensor-based vehicle detection to place vision-based vehicle detection in the overall context of on-road environmental perception. We take this occasion to discuss conceptual similarities and differences that the various sensing modalities bring to vehicle detection and discuss the emerging avenues for data fusion and systems integration. In particular, we briefly discuss the use of millimeter-wave radar and of lidar, alongside computer vision, for on-road vehicle detection. Table I summarizes the comparison between radar, lidar, and vision for vehicle detection.

Millimeter-wave radar is widely used for detecting vehicles on the road. Radar technology has made its way into production-mode vehicles, for applications including adaptive cruise control (ACC) and side warning assist [7], [8]. Typically, a frequency-modulated continuous waveform signal is emitted. Its reflections are received and demodulated, and frequency content is analyzed. The frequency shift in the received signal is used to measure the distance to the detected object. Detected objects are then tracked and filtered based on motion characteristics to identify vehicles and other obstacles [7]. The radar

TABLE I
Comparison of Sensors for Vehicle Detection

Sensing Modality	Perceived Energy	Raw Measurement	Units	Recognizing Vehicles vs. Other Objects
Radar	Millimeter-wave radio signal [emitted]	Distance	Meters	Resolved via tracking
Lidar	600-1000 nanometer-wave laser signal [emitted]	Distance	Meters	Resolved via spatial segmentation, motion
Vision	Visible light [ambient]	Light intensity	Pixels	Resolved via appearance, motion

sensing used for ACC generally features a narrow angular field of view, well suited to detecting objects in the ego vehicle's lane. Fig. 2(a) depicts the operation of radar for on-road vehicle detection.

Radar sensing works quite well for narrow field-of-view applications, detecting and tracking preceding vehicles in the ego lane. Radar vehicle tracking works fairly consistently in different weather and illumination conditions. However, vehicle-mounted radar sensors cannot provide wide field-of-view vehicle detection, struggling with tracking cross traffic at intersections. Furthermore, measurements are quite noisy, requiring extensive filtering and cleaning. Radar-based vehicle tracking does not strictly detect vehicles, rather it detects and

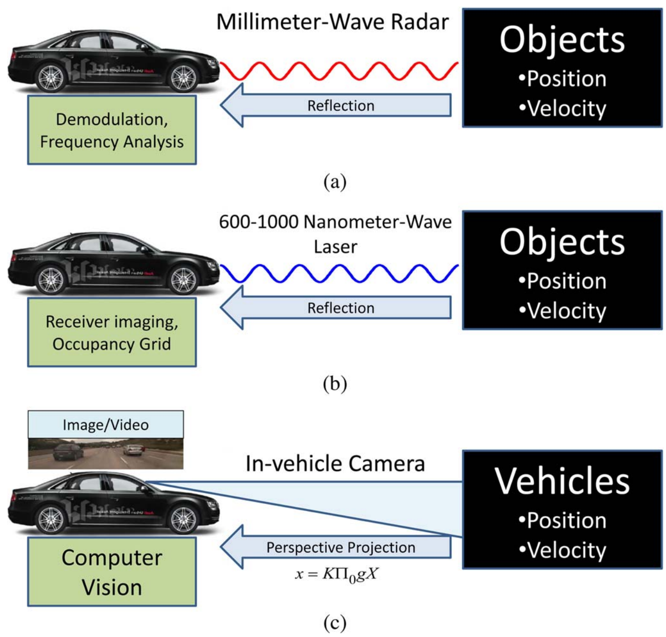

Fig. 2. (a) Radar for on-road vehicle detection uses radar antennas, emitting millimeter-wavelength radio signals. The frequency shift in the reflected signal is used to determine the distance to an object. (b) Lidar for on-road vehicle detection uses laser scanners, emitting illumination at 600- to 1000-nm wavelength, detecting backscattered energy with an imaging receiver, which is used to segment obstacles. (c) Vision for on-road vehicle detection uses cameras, which sense the ambient light. Points in the camera's field of view are mapped to pixels via perspective projection. Computer vision techniques, further detailed in this paper, recognize and localize vehicles from images and video. While radar and lidar detect objects, vision explicitly differentiates vehicles versus nonvehicles.

tracks objects, classifying them as vehicles based on relative motion.

Lidar for on-road vehicle detection has increased in popularity in recent years due to improved costs of lasers, sensor arrays, and computation. Lidar has been extensively used for obstacle detection in autonomous vehicles [9] and are beginning to make their way into driver assistance applications such as ACC [10]. Lidar sensing systems emit laser at wavelengths beyond the visual light spectrum, generally between 600 and 1000 nm, typically scanning the scene at 10–15 Hz [11]. The receiver of the range finder then senses backscattered energy. Using occupancy grid methods [10], objects of interest are segmented from the background. Segmented objects are then tracked and classified as vehicles based on size and motion constraints. Fig. 2(b) depicts the operation of lidar for on-road vehicle detection.

Vehicle-mounted lidar sensing is emerging as a leading technology for research-grade intelligent vehicles, providing cleaner measurements and a much wider field-of-view than radar, allowing for vehicle tracking across multiple lanes. However, lidar sensing is more sensitive to precipitation than radar. While cost remains a factor for lidar systems, the price will continue reduce over the next decade. Lidar-based vehicle tracking does not strictly detect vehicles; rather, it detects, segments, and tracks surfaces and objects, classifying them as vehicles based on size and motion.

Vision-based vehicle detection uses one or more cameras as the primary sensor suite. Unlike lidar and radar, cameras do not emit electromagnetic energy but rather measure the ambient light in the scene. In its simplest form, a digital imaging system consists of a lens and an imaging array, typically charge-coupled device or complementary metal–oxide semiconductor. Within the field of view of an ideal camera, a point X in the 3-D world is mapped to a homogeneous pixel in a digital image via perspective projection, as shown in (1) [12], i.e.,

$$\mathbf{x} = K\Pi_0 g\mathbf{X}$$
$$\mathbf{x} = [x \quad y \quad 1]^T, \quad \mathbf{X} = [X \quad Y \quad Z \quad 1]^T$$
$$g = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}, \quad \Pi_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{1}$$

K contains the camera's intrinsic parameters, and g contains the camera's extrinsic parameters. The mapping converts objects in the real world to representations in the image plane, converting units from meters to pixels. If multiple cameras are used, image rectification according to epipolar constraints is applied [12], followed by stereo matching. The end result of capturing an image from a camera is an array of pixels. In the case of stereo vision, we are left with two arrays of pixels and an array of disparities, which are used to calculate distance, after stereo matching. Fig. 2(c) depicts vehicle detection using vision.

Going from pixels to vehicles is not straightforward. A visual object detection system requires camera-based sensing to measure the scene's light, as well as computational machinery to extract information from raw image data [12]. Unlike lidar or radar, detection cannot rely on a reflected reference signal. Computer vision techniques are necessary to detect the vehicles in images and video. While vehicle detection using cameras often requires more sophisticated computation, it also features several advantages.

Images and video provide a rich data source, from which additional information and context can be surmised. Cameras provide a wide field of view, allowing for detection and tracking across multiple lanes. Cameras feature lower costs than active sensors and are already commonly used for tracking lanes, allowing for system integration [13], shared hardware, and low costs. While active sensors identify objects, vision definitively recognizes objects as *vehicles*. Vision integrates nicely with active sensor suites, allowing sensors like lidar to provide physical measurements, whereas vision classifies objects as vehicle or nonvehicle [14]. The visual domain is also highly intuitive for humans, making vision-based systems attractive for on-road interactivity and driver assistance. The drawbacks to vision-based vehicle detection include sensitivity to light and weather conditions and increased computational cost.

As the intelligent vehicles field advances, computer vision will certainly play a prominent sensing role, either as a primary sensor or as part of multimodal sensor-fusion suites. In the following sections, we detail the recent advances in vision-based vehicle detection, tracking, and behavior analysis. The recent major advances in monocular vision-based vehicle detection have mirrored advances in computer vision, machine learning, and pattern recognition. There has been immense improvement, from template matching to sophisticated feature extraction and classification. In stereo vision, we have seen major advances in stereo matching, scene segmentation, and detection of moving and static objects. The next section of this paper details the recent advances in monocular, stereo vision, and fusion of vision with other sensors for on-road vehicle detection.

III. Vision-Based Vehicle Detection

The lowest level depicted in Fig. 1 involves detecting vehicles using one or more cameras. From a computer vision standpoint, on-road vehicle detection presents myriad challenges. The on-road environment is semistructured, allowing for only weak assumptions to be made about the scene structure. Object detection from a moving platform requires the system to detect, recognize, and localize the object in video, often without reliance on background modeling.

Vehicles on the road are typically in motion, introducing effects of ego and relative motion. There is variability in the size, shape, and color of vehicles encountered on the road [2]. The on-road environment also features variations in illumination, background, and scene complexity. Complex shadowing, man-made structures, and ubiquitous visual clutter can introduce erroneous detections. Vehicles are also encountered in a variety of orientations, including preceding, oncoming, and cross traffic. The on-road environment features frequent and extensive scene clutter, limiting the full visibility of vehicles, resulting in partially occluded vehicles. Furthermore, a vehicle detection system needs to operate at real-time speeds in order to provide the human or autonomous driver with advanced notice of critical situations.

Here, we review on-road vehicle detection. We detail the various cues, assumptions, and classification approaches taken by researchers in the field. We split this section into studies using monocular vision and those using stereo vision for on-road vehicle detection. Table II highlights representative works in vision-based vehicle detection. Fig. 3 shows qualitative results from monocular and stereo-vision-based vehicle detection studies.

A. Monocular Vehicle Detection

We divide vehicle detection approaches into two broad categories: appearance-based and motion-based methods. Generally speaking, appearance-based methods are more common in the monocular vehicle detection literature. Appearance-based methods recognize vehicles directly from images, that is to say that they go directly from pixels to vehicles. Motion-based approaches, by contrast, require a sequence of images in order to recognize vehicles. Monocular images lack direct depth measurements. Even though ego-motion compensation and structure from motion methods have been used for vehicle detection in [15], generally speaking, appearance-based methods are more direct for monocular vehicle detection.

Here, we discuss camera placement and the various applications of monocular vehicle detection. We then detail common features and common classification methods. We detail motion-based approaches. We discuss nighttime vehicle detection and monocular pose estimation.

1) Camera Placement: Vehicle detection using a single camera aims to detect vehicles in a variety of locations with respect to the ego vehicle. The vast majority of monocular vehicle detection studies position the camera looking forward, to detect preceding and oncoming vehicles, as detailed in later subsections. However, various novel camera placements have yielded valuable insight and safety-critical applications.

Mounting the camera on the side-view mirror, facing toward the rear of the vehicle, has allowed for monitoring of the vehicle's blind spot. Detecting vehicles with this camera placement presents difficulty because of the field of view and high variability in the appearance of vehicles, depending on their relative positions. In [16], this camera placement was used to detect overtaking vehicles using optical flow. In the absence of a vehicle in the blind spot, the optical flow of static objects moves backward with respect to the ego vehicle. Oncoming vehicles exhibit forward flow. Vehicles were tracked using Kalman filtering [16]. Optical flow was also used for blind spot detection in [17]. Blind spot vehicle detection is also presented in [18], using edge features and support vector machine (SVM) classification. In [19], a camera is mounted on the vehicle to monitor the blind spot area. The study uses a combination of speeded-up robust features (SURF) and edge segments. Classification is performed via probabilistic modeling, using a Gaussian-weighted voting procedure to find the best configuration.

Mounting an omnidirectional camera on top of the vehicle has been used to acquire full panoramic view of the on-road scene. In [20], omnidirectional vision was used to estimate the vehicle's ego motion, detecting static objects using optical flow. Moving objects were also detected and tracked over long periods of time using Kalman filtering. In [21], a pair of omnidirectional cameras was mounted on the ego vehicle, performing binocular stereo matching on the rectified images for a dynamic panoramic surround map of the region around the vehicle.

Detection of vehicles traveling parallel to the ego vehicle, on either side, has been also pursued. In [22], using a camera looking out the side passenger's window, vehicles in adjacent lanes are detected by first detecting the front wheel and then the rear wheel. The combined parts are tracked using Kalman filtering. In [23], the camera was similarly mounted on the side of the TerraMax autonomous experimental vehicle test-bed. An adaptive background model of the scene was built, and motion cues were used to detect vehicles in the side view.

In [24], the camera was positioned looking backward, out of the rear windshield. The application was detection of the front faces of following vehicles, to advise the driver on the safety of ego lane change. Symmetry and edge operators were used to generate regions of interest; vehicles were detected using Haar wavelet feature extraction and SVM classification.

2) Appearance—Features: A variety of appearance features have been used in the field to detect vehicles. Many earlier works used local symmetry operators, measuring the symmetry of an image patch about a vertical axis in the image plane. Often, the symmetry was computed on image patches after

TABLE II
REPRESENTATIVE WORKS IN VISION-BASED VEHICLE DETECTION

Research Study	Motion/ Appearance	Description	Comments
Monocular Vision			
Sun et al., 2006 [41]	Appearance	HOG and Gabor features, SVM and neural network classification	Feature and classifier evaluation. Evaluation on static images.
Zhu et al., 2006 [77]	Motion	Dynamic background modeling of overtake area	Validation on real-world video, with ego-motion compensation.
Wang and Lien, 2008 [57]	Appearance	Statistical modeling of local features	Detection of sedans in statics image. Evaluation is performed on static images.
Diaz-Alonso et al., 2008 [16]	Motion	Optical flow for blind spot detection	Detection results were validated with lidar for ground truth, and TTC validation.
Chang and Cho, 2010 [67]	Appearance	Haar-like features, boosted classification, online learning	Online learning allows for adaptation to new environments.
Sivaraman and Trivedi, 2010 [49]	Appearance	Haar-like features, Adaboost classification, active learning	Active learning shown to improve detection and false alarm rates, evaluated on highway video.
Yuan et al., 2011 [61]	Appearance	HOG features, SVM classification. Orientation determined using multiplicative kernel learning	Vehicles are oriented using matched detectors. The same framework was shown to work for hand gestures and head rotation.
Jazayeri et al., 2011 [69]	Motion	Optical flow, hidden Markov model classification	Modeling the position and motion of preceding vehicles in the image plane.
Niknejad et al., 2012 [75]	Appearance	HOG features, deformable parts-based model	Adaptive threshold for detection in urban environments.
Lin et al. 2012 [19]	Appearance	SURF and edge features, probabilistic classification, blind spot detection	Front and side car models were evaluated to accommodate different views of blind-spot vehicles.
Stereo Vision			
Research Study	**Motion/ Appearance**	**Description**	**Comments**
Chang et al. 2005 [78]	Appearance	Size, width, height, image intensity features, Bayesian classification	A combination of object geometry, template-matching, image features, and depth map features were used for vehicle detection from single stereo pair. Evaluation in parking lot.
Cabani et al. 2005 [79]	Appearance	Color, 3D vertical edges	Sparse stereo matching using L*a*b* color image pairs, and vertical edges, in order to detect vehicles and obstacles.
Franke et al., 2005 [80]	Motion	Optical flow	Optical flow interest points are tracked in the image plane, and their corresponding 3D positions and velocities are tracked using Kalman filtering.
Badino et al., 2007 [81]	Motion	Occupancy grid, free space computation	6D vision points are tracked. Stochastic occupancy grids are solved using dynamic programming, and free space is computed.
Barrois et al., 2009	Appearance	Clustering of 3D points, vehicle orientation estimation	Clustering of points in 3D using polar iterative closest point algorithm. Points are fit to a cuboid model, and pose is inferred.
Barth and Franke, 2009 [82]	Motion	Optical flow, clustering 6D points	6D vision points are tracked over time, with objects formed by clustering using the Mahalanobis distance.
Broggi et al., 2010 [83]	Appearance	V-disparity, clustering in the disparity space	Detection in the disparity space image.
Danescu et al., 2011 [84]	Motion	Optical flow, particle-based occupancy grid	Occupancy grid cells are represented by particles that serve a dual purpose. In a conventional particle filtering framework, each cell as a position and velocity. Particles also carry a probability of the cell's occupancy.
Erbs et al., 2011 [85]	Motion	Tracking stixels, fitting probabilistic cuboid model	Stixels, vertical intermediate representations of 3D points, are tracked using Kalman filtering. Stixels with similar motion are fit to a cuboid model for vehicle detection and tracking.
Perrollaz et al., 2012 [86]	Motion	Optical flow, spatio-temporally smoothed occupancy grid	The occupancy grid is also smoothed in the time and spatial domains to account for noise and outliers.

evaluating edge operators over the image, to recognize the vertical sides of the rear face of a vehicle [25]–[27]. Edge information helps highlight the sides of the vehicle, as well as its cast shadow [18], [28]–[30]. Symmetry was used along with detected circular headlights and edge energy to detect vehicles at nighttime in [31]. Symmetry and edges were also used in [32] and [33], with longitudinal distance and time to collision (TTC) estimated using assumptions on the 3-D width of vehicles and the pinhole camera model 1.

In recent years, there has been a transition from simpler image features like edges and symmetry to general and robust feature sets for vehicle detection. These feature sets, now common in the computer vision literature, allow for direct classification and detection of objects in images. Histogram

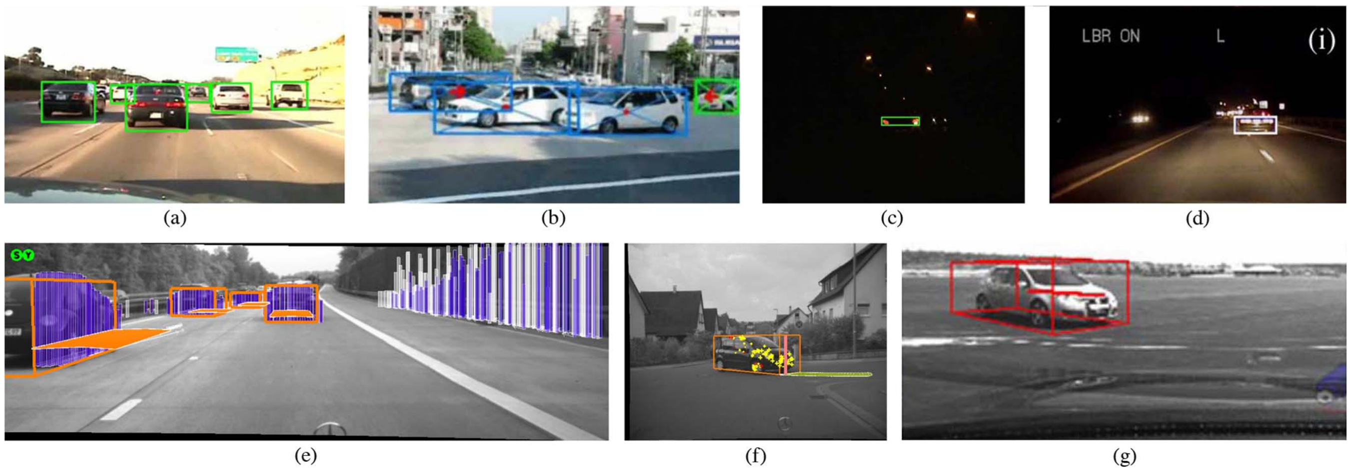

Fig. 3. Images from representative vehicle detection studies, highlighting real-world system performance. (Top row) Monocular: (a) Sivaraman and Trivedi, 2010 [49]; (b) Niknejad *et al.*, 2012 [75]; (c) O'Malley *et al.*, 2010 [103]; (d) Jazayeri *et al.*, 2011 [69]. (Bottom row) Stereo vision; (e) Erbs *et al.*, 2011 [85]; (f) Barth and Franke, 2010 [3]; (g) Danescu *et al.*, 2011 [84].

of oriented gradient (HOG) features and Haar-like features are extremely well represented in the vehicle detection literature, as they are in the object detection literature [34], [35].

HOG features [34] are extracted by first evaluating edge operators over the image and then discretizing and binning the orientations of the edge intensities into a histogram. The histogram is then used as a feature vector. HOG features are descriptive image features, exhibiting good detection performance in a variety of computer vision tasks, including vehicle detection, but they are generally slow to compute. HOG features have been used in a number of studies [36], [37]. In [38], the symmetry of the HOG features extracted in a given image patch, along with the HOG features themselves, was used for vehicle detection. Beyond vehicle detection, HOG features have been used for determining vehicle pose [39]. The main drawback of HOG features is that they are quite slow to compute. Recent work has tackled the speed bottleneck by implementing HOG feature extraction on a GPU [40].

Haar-like features [35] are composed of sums and differences of rectangles over an image patch. Highly efficient to compute, Haar-like features are sensitive to vertical, horizontal, and symmetric structures, making them well suited for real-time detection of vehicles or vehicle parts. In [24], Haar features were extracted to detect the front faces of following vehicles, which were captured with a rear-facing camera. Haar-like features have been extensively used to detect the rear faces of preceding vehicles, using a forward-facing camera [37], [41]–[49]. Side profiles of vehicles have been also detected using Haar-like features [22], by detecting the front and rear wheels. Haar-like features have been also used to track vehicles in the image plane [50]. In [51], Haar features were used to detect parts of vehicles.

While studies that use either HOG or Haar-like features comprise a large portion of recent vehicle detection works, other general image feature representations have been used. In [52], a combination of HOG and Haar-like features [35] was used to detect vehicles. Scale invariant feature transform (SIFT) features [53] were used in [54] to detect the rear faces of vehicles, including during partial occlusions. In [19], a combination of SURF [55] and edges was used to detect vehicles in the blind spot. In [56], Gabor and Haar features were used for vehicle detection. Gabor features were used in [41], in concert with HOG features. Dimensionality reduction of the feature space, using a combination of principal component analysis and independent component analysis, was used in [57] for detecting parked sedans in static images.

3) Appearance—Classification: Classification methods for appearance-based vehicle detection have followed the general trends in the computer vision and machine learning literature. Classification can be broadly split into two categories: discriminative and generative. Discriminative classifiers, which learn a decision boundary between two classes, have been more widely used in vehicle detection. Generative classifiers, which learn the underlying distribution of a given class, have been less common in the vehicle detection literature.

While in [41] and [58] artificial neural network classifiers were used for vehicle detection, they have recently fallen somewhat out of favor. Neural networks can feature many parameters to tune, and the training converges to a local optimum. The research community has moved toward classifiers whose training converges to a global optimum over the training set, such as SVMs [59] and AdaBoost [60].

SVMs [59] have been widely used for vehicle detection. In [24], SVM classification was used to classify feature vectors consisting of Haar wavelet coefficients. The combination of HOG features and SVM classification has been also used [36], [37], [41]. The HOG-SVM formulation was extended to detect and calculate vehicle orientation using multiplicative kernels in [61]. Edge features were classified for vehicle detection using SVM in [18] and [29]. In [56], vehicles were detected using Haar and Gabor features, using SVM classification.

AdaBoost [60] has been also widely used for classification, largely owing to its integration in cascade classification in [35]. In [62], AdaBoost classification was used for detecting vehicles based on symmetry feature scores. In [63], edge features were classified using AdaBoost. The combination of Haar-like feature extraction and AdaBoost classification has been used to detect rear faces of vehicles in [44], [46], and [64]–[66]. In [22], Haar features and AdaBoost classification were used to detect the front and rear wheels of vehicles from the side view. The combination of Haar features and AdaBoost classification was used to detect parts of vehicles in [51]. AdaBoost classification was used in an active learning framework for vehicle detection in [37] and [49]. In [67], online boosting was used to train a vehicle detector. In [68], WaldBoost was used to train the vehicle detector.

Generative classifiers have been less common in the vehicle detection literature. This is because it often makes sense to model the classification boundary between vehicles and nonvehicles rather than the distributions of each class. In [19], a probabilistically weighted vote was used for detecting vehicles in the blind spot. In [69], motion-based features were tracked over time and classified using hidden Markov models. In [57], Gaussian mixture modeling was used to detect vehicles in static images. In [54], hidden random field classification was used to detect the rear faces of vehicles.

Recently, there has been interest in detecting vehicles as a combination of parts. The motivation consists of two main goals: encoding the spatial configuration of vehicles for improved localization and using the parts to eliminate false alarms. In [19], a combination of SURF and edge features was used to detect vehicles, with vehicle parts identified by keypoint detection. In [54], vehicles were detected as a combination of parts, using SIFT features and hidden conditional random field classification. In [70], spatially constrained detectors for vehicle parts were trained; the detectors required manual initialization of a reference point. The deformable parts-based model [71], [72], using HOG features and the latent SVM, has been used for on-road vehicle detection in [73]–[75]. In [51] and [76], the front and rear parts of vehicles were independently detected and matched using structural constraints, which were encoded by an SVM.

4) Motion-Based Approaches: Motion-based monocular vehicle detection has been less common than appearance-based methods. It is often more direct to use appearance cues in monocular vision because monocular images do not directly provide 3-D depth measurements. Adaptive background models have been used in some studies, in an effort to adapt surveillance methods to the dynamic on-road environment. In [23], an adaptive background model was constructed, with vehicles detected based on motion that differentiated them from the background. Adaptive background modeling was also used in [87], specifically to model the area where overtaking vehicles tend to appear in the camera's field of view. Dynamic modeling of the scene background in the area of the image where vehicles typically overtake was implemented in [77]. A similar concept, i.e., dynamic visual maps, was developed in [88] for detecting vehicles and identifying unusual maneuvering in the scene. In [89], a homography matrix was computed between adjacent video frames; image regions that did not cleanly map between frames were assumed to be vehicles. This method seems likely to return many false alarms, but quantitative performance analysis was not included.

Optical flow [90], a fundamental machine vision tool, has been used for monocular vehicle detection [91]. In [27], a combination of optical flow and symmetry tracking was used for vehicle detection. In [69], interest points that persisted over long periods of time were detected as vehicles traveling parallel to the ego vehicle. Optical flow was used in conjunction with appearance-based techniques in [44]. Ego-motion estimation using optical flow and integrated detection of vehicles was implemented in [92]–[94]. Ego-motion estimation using an omnidirectional camera and detection of vehicles was implemented in [20]. In [16], optical flow was used to detect overtaking vehicles in the blind spot. A similar approach for detecting vehicles in the blind spot was reported in [17]. Cross traffic was detected in [95]. In [96], optical flow was used to form a spatiotemporal descriptor, which was able to classify the scene as either intersection or nonintersection. Optical flow was used in [97] for segmentation of the on-road scene using video. In [98], ego-motion compensation and motion cues were used for tomographical reconstruction of objects in the scene.

5) Nighttime Vision: The vast majority of vision-based vehicle detection papers are dedicated to daytime conditions. Nighttime conditions may be dealt with in a few ways. Using high dynamic range cameras [41] allows for the same appearance and classification model to be used during daytime or nighttime conditions. Well-illuminated nighttime scenes can also accommodate vehicle detection models that have been designed for daytime conditions [28]. Absent specialized hardware or illumination infrastructure, various studies have trained specific models for detecting vehicles at nighttime, often by detecting the headlights and taillights of vehicles encountered on the road.

Color space thresholding can often serve as the initial segmentation step in detecting vehicle lights in low-light conditions. In [99], vehicles are detected at nighttime using stereo vision to extract vertical edges and 3-D and color in the $L*a*b*$ color space. In [100], taillights are localized by thresholding the grayscale image. Vehicles are detected based on fitting a bounding box around pairs of detected taillights, and 3-D range information is inferred by making assumptions on the typical width of a vehicle and solving for the longitudinal distance using the common pinhole model. In [31], symmetry, edge energy, and detected circles are used to track vehicles using particle filters. In [101], vehicle taillights are paired using cross correlation and validated by tracking. Using the pinhole model and the assumptions on the 3-D dimensions of vehicles, the TTC is computed for forward collision warning applications. The use of the pinhole model to compute longitudinal distance for nighttime vehicles is also featured in [102]. In [103], vehicles are detected by localizing pairs of red taillights in the hue–saturation–value color space. The camera has been configured to reliably output colors by controlling the exposure, optimizing the appearance of taillights for segmentation. The segmented taillights are detected as pairs using cross correlation and symmetry. Vehicles are then tracked in the image plane using Kalman filtering. In [104], multiple vehicles are detected by tracking headlight and taillight blobs, a detection-by-tracking approach. The tracking problem is formulated as a maximum *a posteriori* inference problem over a random Markov field.

6) Vehicle Pose: Determining vehicle pose can be useful for understanding how a detected vehicle is oriented with respect to the ego vehicle. The orientation can serve to predict the vehicle's motion. In the absence of 3-D measurements, tracking information coupled with a set of geometric constraints was used in [30] to determine the vehicles' pose information. In [105], color, location, and texture features were used, with detection and orientation using conditional random field classification.

Simultaneously detecting vehicles and determining their orientations has been pursued with the use of HOG features [34] in various works. HOG features, while expensive to compute, are descriptive and well suited to distinguishing orientations within an object class. In [106], a set of HOG-SVM classifiers was trained for several orientations. Vehicles were detected in static frames using the all-versus-one trained detectors. However, it was found that a general HOG-SVM detector performed better at detecting vehicles than the oriented detectors. In [61], multiplicative kernels were used to train a family of HOG-SVM classifiers for simultaneous vehicle detection and orientation estimation. In [39], HOG features were used to discover vehicle orientations in a partition-based unsupervised manner, using simple linear classifiers.

B. Stereo Vision for Vehicle Detection

Motion-based approaches are more common than appearance-based approaches to vehicle detection using stereo vision. Multiview geometry allows for direct measurement of 3-D information, which provides for understanding of scene, motion characteristics, and physical measurements. The ability to track points in 3-D and distinguish moving from static objects affects the direction of many stereo-vision studies. While monocular vehicle detection often relies on appearance features and machine learning, stereo vehicle detection often relies on motion features, tracking, and filtering. Stereo-vision approaches have access to the same image pixels as monocular approaches, but two views allow spatial reasoning, and many research studies concentrate their efforts on this problem domain.

While [107] places stereo cameras looking sideways for cross traffic, most studies place the stereo rig looking forward out the front windshield to detect vehicles ahead of the ego vehicle. Here, we discuss stereo matching, appearance-based approaches, and motion-based approaches to vehicle detection using stereo vision.

1) Stereo Matching: Epipolar rectification between the two cameras in a stereo rig transforms the epipolar lines into horizontal scan lines in the respective image planes. This transformation confines the search for point correspondences between two images to the horizontal direction. The set of solved point correspondences yields a disparity map and is achieved by performing stereo matching [12]. Various techniques are available in the vision community for dense matching. Advances in dense stereo matching, filtering, and interpolation have been of great interest in the intelligent vehicles community [80], as better stereo matching allows for better interpretation of the on-road scene. While classic correlation-based stereo matching has been implemented and highly optimized [108], new advances in stereo matching are actively pursued in the computer vision and intelligent vehicles communities. In particular, there has been a transition from local correlation-based approaches [108] toward semiglobal matching [109]–[111], which features denser disparity maps and lower errors.

2) Compact Representations: Stereo-vision studies have made extensive use of compact representations of measured data, including occupancy grids [86], elevation maps [112], free space understanding [81], ground surface modeling [113], and dynamic stixels [85]. Compact representations serve to facilitate segmentation of the scene [113], identify obstacles [114], and reduce computational load. We discuss compact representations in the following subsections, dividing them between appearance-based and motion-based methods.

3) Appearance-Based Approaches: Exclusive reliance on appearance cues for vehicle detection is not as common in stereo vision as in monocular vision. While motion-based approaches are more common, even studies that rely on motion for vehicle detection often utilize some appearance-based stereo-vision techniques for initial scene segmentation.

The v-disparity [113] has been widely used to model the ground surface, in order to identify objects that lie above the ground. The v-disparity forms a histogram of disparity values for pixel locations with the same v, i.e., vertical image coordinate. Starting with an $n \times m$ disparity map, the result is an image consisting of n stacked histograms of disparity for the image. Using curve-fitting techniques, such as the Hough transform [115] or the RANdom SAmple Consensus (RANSAC) [116], disparity can be modeled as a function of the v coordinate of the disparity map, and pixel locations can be classified as belonging to the ground surface if they fit this model [113]. The v-disparity has been widely used in stereo vision for intelligent vehicles [83], [113], [117]–[122].

Free space understanding from disparity maps has been implemented using the u-disparity [123], which forms a similar histogram of stacked disparities, for pixel locations sharing the same u coordinate. Instead of fitting a road model, the u-disparity is used to infer free space directly. Free space computation heavily features in the stereo-vision literature, for scene segmentation and highlighting of potential obstacles. In [81] and [124], free space was directly computed from the disparity and depth maps using dynamic programming. In [125], convolutional noise and image degradation are added to stereo image pairs to model the corresponding errors introduced to stereo matching and 3-D localization of tracked interest points. Corresponding methods are introduced to compensate for the errors in localization.

Monocular appearance features are sometimes used for vehicle detection using a stereo rig, including color [79], and image intensity [78]. Disparity and depth appearance features are generally more common. In [78], features such as size, width, height, and image intensity were combined in a Bayesian model to detect vehicles using a stereo rig. In [126], a histogram of depths, which was computed from stereo matching, was used to segment out potential vehicles. Operations directly on the monocular frame also include Delaunay triangulation [127].

Various studies have utilized clustering in the depth map for object detection, often using Euclidean distance to cluster point clouds into objects [128], [129]. Clustering was also used for object detection in [123]. In [130], clustering was implemented using a modified version of iterative closest point, using polar coordinates to segment objects. The implementation was able to detect vehicles and infer the vehicle's pose with respect to the ego vehicle. Clustering was used in tandem with image-based mean shift algorithm for vehicle detection in [131]. The mean-shift algorithm was also used in [132] for object detection.

4) Motion-Based Approaches: The use of motion heavily features in stereo-based vehicle detection. The foundation for a large portion of stereo-vision analysis of the on-road scene starts with optical flow [90]. In many studies, interest points are tracked in the monocular image plan of one of the stereo rig's cameras and then localized in 3-D using the disparity and depth maps [133]. In [133], the concept of 6D vision, i.e., the tracking of interest points in 3-D using Kalman filtering, along with ego-motion compensation, is used to identify moving and static objects in the scene. Optical flow is also used as a fundamental component of stereo-vision analysis of the on-road scene in [84], [117], [126], [127], [131], and [133]–[137]. A 3-D version of optical flow, in which a least squares solution to 3-D points' motion is solved, is used in [138].

There are various modifications and uses of optical flow point tracking in the literature. In [136], block-based coarse-to-fine optical flow is compared with the classical Lucas–Kanade optical flow and is found to be more robust to drifting. The object flow descriptor [96] is used to understand whether the ego vehicle is at an intersection or arterial road by modeling the aggregate flow of the scene over time. Scene flow is used to model the motion of the background, and regions whose motion differs from the scene flow are categorized as candidate vehicles in [134], where integration with geometric constraints improves vehicle detection performance.

Via tracking, ground plane estimation is implemented by tracking feature points from optical flow [127], [134]. In [117], the ground plane model is fit using total least squares. In [126] and [134], the ground plane is estimated using RANSAC to fit the plane parameters [116]. The ground is estimated as a quadratic surface in [112], which serves as a scene segmentation for obstacle detection using rectangular digital elevation maps [114]. This work is enhanced in [139] by radial scanning of the digital elevation map, for detecting static and dynamic objects, which are tracked with Kalman filtering. Interest points are also tracked in order to utilize structure from motion techniques for scene reconstruction and understanding. The Longuet-Higgins equations are used for scene understanding in [3], [127], and [140]. In [127], tracked interest points are used to estimate ego motion. In [128], ego-motion estimation is performed by tracking SURF interest points.

In [85], tracked 3-D points, using 6D vision, are grouped into an intermediate representation consisting of vertical columns of constant disparity, which are termed stixels. Stixels are initially formed by computing the free space in the scene and using the fact that structures of near-constant disparity stand upon the ground plane. The use of the stixel representation considerably reduces the computation expense over tracking all the 6D vision points individually. The tracked stixels are classified as vehicles using probabilistic reasoning and fitting to a cuboid geometric model.

Occupancy grids are widely used in the stereo-vision literature for scene segmentation and understanding. Static and moving points are tracked in 3-D, which are used to populate an occupancy grid and compute the free space in the scene using dynamic programming in [81]. Dynamic programming is also used in [141] for computing the free space and populating the occupancy grid. The comparison of spatial representation of the scene is presented, detailing Cartesian coordinates, column disparity, and polar coordinates, in a stochastic occupancy framework. We note that the column disparity representation of [81] is equivalent to the u-disparity representation of [123]. In [123] and [142], scene tracking and recursive Bayesian filtering are used to populate the occupancy grid in each frame, whereas objects are detected via clustering. In [117], the occupancy grid's state is inferred using a recursive estimation technique termed the sequential probability ratio test. In [86], the occupancy grid is filtered both temporally and spatially. In [140], the occupancy grid is set up in polar coordinates, and the cells are assigned depth-adaptive dimensions to model the field of view and depth resolution of the stereo rig. In [84], the occupancy grid is populated using motion cues, with particles representing the cells, their probabilities the occupancy, and their velocities estimated for object segmentation and detection.

IV. On-Road Vehicle Tracking

Beyond recognizing and identifying vehicles from a given captured frame, vehicle tracking aims to reidentify and measure dynamics and motion characteristics and predict and estimate the upcoming position of vehicles on the road. Implicit in vehicle tracking are issues like measurement and sensor uncertainty, data association, and track management. Here, we detail the common vehicle tracking methods employed in the research literature. We split our discussion into portions detailing monocular and stereo-vision approaches. While there are estimation and filtering methods common to both camera configurations, often, the estimated parameters differ based on the available measurements. Many monocular tracking methods measure and estimate dynamics in terms of pixels, whereas stereo-vision methods estimate dynamics in meters.

We continue this section by discussing works that combine or fuse monocular and stereo vision for on-road vehicle detection and tracking. We then discuss papers that focus on optimized system architecture and real-time implementation of on-road vehicle detection and tracking. We conclude this section with the discussion of studies that fuse vision with other sensing modalities for on-road vehicle detection and tracking. Table III features representative papers and their tracking approaches for monocular and stereo vision.

A. Monocular Vehicle Tracking

Using monocular vision, vehicles are typically detected and tracked in the image plane. Tracking using monocular vision serves two major purposes. Tracking facilitates estimation of motion and prediction of vehicle position in the image plane. Second, tracking enforces temporal coherence, which helps to maintain awareness of previously detected vehicles that were

TABLE III
REPRESENTATIVE WORKS IN VISION-BASED VEHICLE TRACKING

Monocular Vision			Stereo Vision		
Research Study	**Tracking Method**	**Comments**	**Research Study**	**Tracking Method**	**Comments**
Zhu et al., 2005 [143]	Optical flow, geometric constraints, Kalman filtering	Vehicles are tracked in the image plane using optical flow and Kalman filtering for motion estimation	Rabe et al., 2007 [133]	Kalman filtering	Ego-motion estimation is achieved using Kalman filtering on information from image features and inertial sensor. Stationary and moving feature points are detected.
Liu et al., 2007 [24]	Template matching	Templates of detected vehicles are created, and updated each frame to track vehicles in the image plane.	Barth and Franke, 2009 [82]	Extended Kalman filtering	Extended Kalman filtering is used to estimate the state of tracked vehicles, including the yaw-rate, to determine turning behavior.
Haselhoff and Kummert, 2009 [50]	Feature-based tracking, Kalman filtering	Vehicles are detected, and the feature values are used as hypotheses for tracking in the image plane in subsequent frames. Kalman filtering smooths the estimation.	Barth and Franke, 2010 [3]	Kalman filtering, interacting multiple models	Interacting multiple models are used to determine the best estimate of the vehicle's motion, using separate motions modes and corresponding motion models.
Sivaraman and Trivedi, 2010 [49]	Particle filtering	Vehicles are tracked in the image plane using particle filtering.	Bota and Nedevschi, 2011 [137]	Kalman filtering	The position and velocity of each corner of the vehicle is tracked using Kalman filtering.
Quan et al., 2011 [61]	Particle filtering	Particle filtering is used to track, and determine the orientation of vehicles in the image plane.	Lim et al., 2011 [118]	Extended Kalman filtering	Vehicles are tracked using extended Kalman filtering as part of a unified track management framework.
Xue and Ling, 2011 [144]	Particle filtering	Vehicles are tracked by best match to sparsely-represented templates in a particle filter framework.	Danescu et al., 2011 [84]	Particle filtering	Occupancy grid cells are represented by particles that serve a dual purpose. In a conventional particle filtering framework, each cell as a position and velocity. Particles also carry a probability of the cell's occupancy.
Niknejad et al., 2012 [75]	Particle filtering	Particle filtering estimates the vehicle's position in the image plane, as well as adaptively changing the decision threshold for the vehicle detector, to accommodate change in appearance.	Lefebvre and Ambellouis, 2012 [132]	Mean-shift on 3D points	Mean-shift tracking of 3D point clouds, forming objects using sparse stereo-matching is used to track vehicles.

not detected in a given frame [50], while filtering out spurious false positives [49].

The goal in monocular vision is to measure the motion and predict the position of vehicles in pixel position and pixel velocity. The observation space, based on pixels, gives way to uniquely vision-based tracking methods, based on the appearance of the vehicle in the image plane. An example of uniquely vision-based tracking is template matching. In [24], vehicles were detected in the image plane using Haar wavelet coefficients and SVM classification. Vehicles were tracked from frame to frame by taking a measurement of the similarity in appearance.

Often, the appearance-based tracking is based on cross-correlation scores. Vision-based tracking is taken one step further using feature-based tracking [145]. In [50], vehicles were detected using Haar-like features and AdaBoost cascade classification. Candidate vehicles' locations were predicted using Kalman filtering in the image plane. The measurements in the image plane were determined by a local search over the image patch for similar feature scores, allowing for a measurement even if the detector failed in a particular frame. Optical flow has been also used to track vehicles by directly measuring the new position and the displacement of interest points [143].

Conventional tracking and Bayesian filtering techniques have been widely used in the monocular vehicle tracking literature. The state vector typically consists of the pixel coordinates that parametrize a rectangle in the image plane and their interframe pixel velocities [28]. In [22], [27], and [33], Kalman filtering was used to estimate the motion of detected vehicles in the image plane. Particle filtering has been also widely used for monocular tracking in the image plane [28], [31], [49], [73], [75], [144].

Estimating longitudinal distance and 3-D information from monocular vision has been attempted in various vehicle tracking studies. Typically, the ground plane is assumed flat [44], [146] or its parameters estimated using interest point detection and a robust mode-fitting step, such as RANSAC [116]. In [30], a set of constraints and assumptions was used to estimate 3-D coordinates from monocular vision, and Kalman filtering was used to track vehicles in 3-D. In [25], 3-D information was inferred using ground plane estimation, and interacting multiple models were used to track vehicles, each model consisting of a Kalman filter. In [93] and [94], ego motion was estimated using monocular vision, and moving objects were tracked in 3-D using Kalman filtering. While various tracking studies have estimated 3-D vehicle position and velocity information from monocular measurements, few such studies have compared their measurements to a ground truth reference 3-D measurement, from radar, lidar, or stereo vision, for example.

B. Stereo-Vision Vehicle Tracking

Vehicle tracking using stereo vision concerns itself with measuring and estimating the position and velocity, in meters, of detected vehicles on the road. The state vector often consists of the vehicle's lateral and longitudinal position, width and height, as well as velocity. Estimation is most often implemented using Kalman filtering, which is considered optimal, assuming linear motion and Gaussian noise [80]. In reality, vehicle motion is nonlinear, with the vehicle's yaw rate describing the vehicle's turning behavior. Using the extended Kalman filter (EKF) is often used for estimating nonlinear parameters by linearizing the motion equations for estimation [82]. Particle filtering has been used as an alternative to measure both linear and nonlinear motion parameters [138], using sample importance re-sampling in place of the linearization of the EKF.

Kalman filtering for stereo-vision vehicle tracking has been widely used [135] for vehicle tracking, as well as disparity filtering. Noise in stereo matching is generally modeled as white Gaussian noise [125], [133], and filtering over time can produce cleaner disparity maps [80]. Kalman filtering is used to track individual 3-D points in [81] and [133]. Kalman filtering is used to track stixels, i.e., intermediate vertical elements of near-constant depth, which are fit to cuboid vehicle models [85]. In [5] and [126], vehicles are detected in the monocular plane using an AdaBoost-based classification and tracked in 3-D using Kalman filtering in the stereo domain. In [137], vehicles' positions and velocities are estimated using Kalman filtering. In [147], Kalman filtering is used to track objects detected by clustering, stereo matching linear cameras. In [3], Kalman filtering is used to estimate the vehicles' yaw rate, as well as position and velocity.

The EKF has been also widely used in stereo-vision vehicle tracking, specifically to account for nonlinearity in motion and observational model quantities. The EKF was used to estimate the yaw rate and the corresponding turning behavior of vehicles in [82] and [148]. Extended Kalman filtering for vehicle tracking was particularly apt due to camera positioning in [107], with the side-mounted stereo rig observing particularly nonlinear motion of tracked vehicles with respect to the camera's frame of reference. Extended Kalman filtering was used to model the nonlinearity of mapping a vehicle's 3-D position into stereo image position and disparity in [117]. In [127], extended Kalman filtering was used to estimate the ego motion, with independently moving objects' position and motion estimated using Kalman filtering. Vehicles were also tracked using extended Kalman filtering in [118].

Particle filtering for vehicle tracking has been also fairly widely used. In particular, particle filtering offers an alternative to the EKF's estimation of nonlinear parameters, as the particle filter's multiple hypotheses are weighted by a likelihood function. In [138], vehicles were tracked using a particle filter, estimating their 3-D position and yaw rate. In [131], the motion of tracked vehicles was estimated using a particle filter that mapped the motion to full trajectories, which were learned from prior observational data. In [84], the on-road environment is modeled using particles that serve a dual purpose, i.e., as occupancy cells and as tracking states for detected objects and obstacles.

Interacting multiple models have been used in tracking to estimate the motion of a vehicle given different motion modes. In [3], four different predominating modes were used to model the motion of oncoming vehicles at intersections, in terms of their velocity and yaw rate characteristics. The goal was to identify whether the velocity was constant or accelerated and whether the yaw rate was constant or accelerated. The model fit was determined using the error covariance of each estimator. A state transition probability was used to switch between competing modes after the model fit was determined [3]. Interacting multiple models were also used in [107]. The use of interacting multiple models will likely increase in popularity, as measurements become more precise and as it becomes apparent that all the motion parameters cannot be well estimated by a single linear or linearized filter.

C. Fusing Monocular and Stereo-Vision Cues

Various studies have fused monocular and stereo vision for on-road vehicle tracking. We draw a distinction between papers that use optical flow and stereo vision for vehicle detection and those papers that use monocular computer vision for full vehicle detection, typically relying on machine learning and stereo vision for tracking in 3-D.

The use of both monocular and stereo-vision cues typically manifests itself in the use of monocular vision for detection and stereo vision for 3-D localization and tracking. In [149], it was noted that monocular vision can detect objects that stereo-vision approaches typically miss, such as disambiguation of two objects that lie close together in 3-D space. This problem was addressed by detecting in the monocular plane but localizing in 3-D using stereo vision. In [150], monocular symmetry was used to generate vehicle candidate regions, and stereo vision was used to verify those regions as vehicles, by searching for vertical objects in the 3-D domain. In [5], vehicles were tracked in the image plane using a monocular vehicle detector [49] and in 3-D using stereo vision and Kalman filtering. Clustering on aggregates vehicle tracking data from the system presented in [5] was used for learning typical vehicle behavior on highways in [4].

In [118], vehicles were detected using an AdaBoost classifier on the monocular plane. The v-disparity was used to estimate the ground surface, and vehicles were tracked using extended

Kalman filtering in the stereo-vision domain. Track management for reduction of false alarms and improved precision was presented. In [126], vehicles' candidate regions were selected in the image plane using a set of AdaBoost detectors, which are trained for multiple vehicle views. The candidate regions were verified by looking for peaks in the disparity range. Stereo-vision was used for 3-D ranging and for estimating the ground plane.

D. Real-Time Implementation and System Architecture

Eventually, for a vehicle detection and tracking system to be of utility on the road, real-time implementation is necessary, typically processing above ten frames per second. While some detection algorithms run in real time on standard central processing units, many do not, and further efforts are necessary to optimize the implementation in hardware and software.

Efforts to implement vehicle detection algorithms using embedded systems implementations have garnered attention in recent years. In [151], vehicle detection using shadow features was implemented on an embedded system. In [152], a boosting-based vehicle detector was implemented on an embedded system. In [153], nighttime vehicle detection was implemented on an embedded system. Embedded implementation of stereo-vision-based lane and vehicle detection and tracking was reported in [154]. Commercialization of embedded vision-based vehicle detection has also hit the market [155]. Embedded systems for correlation-based stereo matching have been also commercialized [108].

In recent years, the availability of GPUs has enabled real-time implementation and parallelization of computationally expensive vision algorithms for vehicle detection and tracking. In [156], an early GPU was used for monocular vehicle detection. In [157], real-time stereo matching using semiglobal matching [109] was implemented on the GPU. In [40], the GPU was used to implement real-time vehicle detection using HOG [34] features. GPU implementation was used in [86] for real-time occupancy grid computation. In [158], vehicle detection was implemented using a fusion of stereo vision and lidar on the GPU.

E. Fusing Vision With Other Modalities

Over the past decade, the availability and cost of a variety of sensors has become favorable for integration in intelligent vehicles, for driver assistance, and for autonomous driving. It is widely accepted that fully autonomous vehicles will need an advanced sensor suite, covering a variety of sensing modalities, in order to sense, perceive, and respond to the on-road environment in a safe and efficient manner. Leading research in autonomous vehicles features sensor suites that include cameras, lidars, and radar sensing arrays [9]. Often, the vision algorithms used in sensor-fusion studies closely resemble those in vision-only studies, with information fusion performed across modalities, to reduce uncertainty, cover blind spots, or perform ranging with monocular cameras.

Radar-vision fusion for on-road vehicle detection and perception has received quite a bit of attention in recent years [159]. Radars have good longitudinal ranging coupled with crude lateral resolution; monocular vision can localize well in the camera's field of view but lacks ranging. The combination of the two can ameliorate the weakness of each sensor [160], [161]. In [162] and [163], information fusion between radar and vision sensors was used to probabilistically estimate the positions of vehicles and to propagate estimation uncertainty into decision making, for lane change recommendations on the highway. In [164], vision and radar were combined to detect overtaking vehicles on the highway, using optical flow to detect vehicles entering the camera's field of view. Radar and vision were combined in [165], with radar detecting side guardrails and vision detecting vehicle using symmetry cues.

Several studies perform extrinsic calibration between radar and camera sensors. In [166], obstacles are detected using vision operations on the inverse perspective mapped image and ranged using radar. In [167], vehicles are detected with a boosted classifier using Haar and Gabor features and ranged using radar. In [160], camera and radar detections were projected into a common global occupancy grid; vehicles were tracked using Kalman filtering in a global frame of reference. In [168], potential vehicles were detected using saliency operations on the inverse perspective mapped image and combined with radar. In [169], vehicles were detected using a combination of optical flow, edge information, and symmetry; ranged with radar; and tracked using interacting multiple models with Kalman filtering. In [170], symmetry was used to detect vehicles, with radar ranging. In [171], vehicles were detected using HOG features and SVM classification and ranged using radar. In [15], monocular vision was used to solve structure from motion, with radar providing probabilities for objects and the ground surface. In [172], a radar-vision online learning framework was utilized for vehicle detection. Stereo vision has been also used in conjunction with radar sensing [173], [174].

Fusion of lidar with monocular vision has been explored in recent years. Several studies perform extrinsic calibration between lidar and camera sensors, using monocular vision for vehicle detection and lidar for longitudinal ranging. In [175], monocular vision was used to detect vehicles using Haar-like features, and ranging was performed using lidar. A similar system was presented in [14] and [176]. Saliency was used as the vision cue in [177], which was fused with lidar in a Bayesian framework. Fusion of stereo vision with lidar was performed in [178]–[181].

V. ON-ROAD BEHAVIOR ANALYSIS

Analysis of the behavior of tracked vehicles has emerged as an active and challenging research area in recent years. While considerable research effort has been dedicated to on-road detection and tracking of vehicles in images and video, going from pixels to vehicles with positions and velocity, the highest level of semantic interpretation lies in characterizing the behavior of vehicles on the road. In order to analyze the on-road behavior of other vehicles, robust vehicle detection and tracking are prerequisite. While various studies have modeled the vehicle dynamics [182] and driver gestures [183] associated with the ego vehicle's maneuvering, research into the on-road

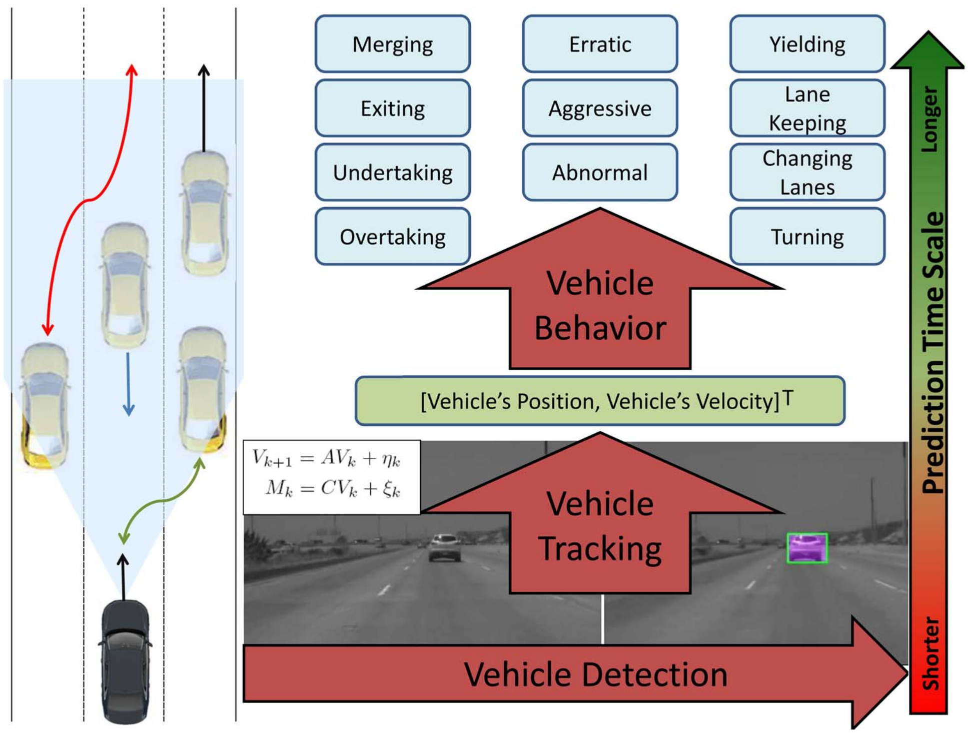

Fig. 4. Ascending levels vehicle behavior interpretation. At the lowest level, vehicles are detected using vision. Vehicle tracking estimates the motion of previously detected vehicles. At the highest level of interpretation, vehicle behavior is characterized. Behavior characterization includes maneuver identification, on-road activity description, and long-term motion prediction.

behavior of other vehicles is a relatively recent development. Fig. 4 depicts on-road behavior analysis in the context of vision-based understanding of the driving environment. At the lowest level, detection takes place, recognizing and localizing vehicles on the road. One level up, tracking reidentifies vehicles, measuring their motion characteristics using a motion model. Often, linear or linearized models are used. At the highest level, using spatial and temporal information from vehicle detection and vehicle tracking, vehicle behavior analysis is performed.

Research studies in this area take a variety of approaches to characterize on-road behavior. Certain studies try to categorize observed vehicle behavior as normal or abnormal [88], identifying and highlighting critical situations. Other studies try to identify specific maneuvers, such as overtaking [164], turning [3], or lane changes [6]. Most recently, studies in the literature have tried to make long-term classification and prediction of vehicle motion. While vehicle tracking, often based on Kalman filtering, can make optimal estimation of the vehicle state one frame (1/25 s) ahead of time, trajectory modeling approaches try to predict vehicle motion up to 2 s ahead, based on models of typical vehicle trajectories [184]. Fig. 5 depicts trajectory prediction.

Broadly speaking, we categorize studies that address the characterization of on-road vehicle behavior based on four main criteria. First, we consider the role of context in the analysis of on-road behavior, loosely defined to encompass considerations such as urban driving versus highway driving or intersection versus nonintersection driving. Second, we consider the identification of prespecified maneuvers, such as turning, lane change, or overtaking maneuvers, of tracked vehicles on the road. Third, we consider the use of *trajectories*, i.e., long-term

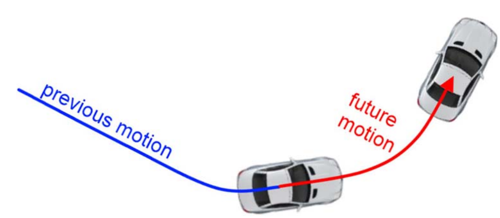

Fig. 5. Depiction of trajectory prediction, aiming to map the most likely future vehicle motion, based on observed motion [185].

sequences of positions and velocities, in characterizing on-road behavior. Finally, we consider the classification and modeling found in the literature.

A. Context

The use of context is a vital component of many studies that characterize on-road vehicle behavior. The motion model used in [69] models the distribution of vehicles in the image plane, using it as a prior probability on vehicle detections. The vehicle detection in [69] can be viewed as a detection-by-tracking approach, which is enabled by spatiotemporal modeling of the driving context. In [96], histograms of scene flow vectors are used to classify the driving environment as intersection or nonintersection driving, modeling the driving context using spatiotemporal information. In [4], a context-specific spatiotemporal model of highway driving is developed by performing clustering on observed vehicle trajectories on highways. In [131], the trajectories of vehicles are recorded as they navigate a roundabout and are used to model the long-term behavior of vehicles in roundabouts. In [3], the context of interest is intersections, i.e., the turning behavior of oncoming vehicles inferred. In [88], a dynamic visual model is developed of the driving environment, with saliency alerting the system to unusual and critical on-road situations.

B. Maneuvers

A body of work has been dedicated to the detection of specific maneuvers of vehicles on the road. In [16], overtaking behavior is detected by detecting vehicles in the blind spot of the ego vehicle. Overtaking behavior is specifically detected in [77], this time with the camera pointing forward and vehicles detected as they overtake in front of the ego vehicle. In [164], overtaking behavior is also detected in front of the ego vehicle, using a fusion of vision and radar. Overtaking behavior is also detected in [87], also for vehicles in front of the ego vehicle. In these studies, the overtaking maneuver is detected by virtue of detecting the vehicle, as the search space includes only vehicles that are in the process of overtaking.

By contrast, specific maneuvers are identified in other works via inference on tracking information. In [82], the turning behavior of tracked vehicles is identified by measuring the yaw rate using extended Kalman filtering. Using the yaw rate in the vehicle motion model, the system is able to detect turning behavior. In [3], turning behavior is further addressed, using interacting multiple models to characterize the motion of the oncoming vehicle. The model with the highest likelihood, based on observations, characterizes the turning behavior of the oncoming vehicle, with a transition probability handling change of states. Turning behavior is addressed in [130] by solving the vehicle's pose with respect to the ego vehicle using clustering of 3-D points.

On-road vehicle behavior is modeled in [186] as a Markov process and inferred using a dynamic Bayesian network, based on tracking observations. However, the experimental evaluation is performed using simulation data. In [6], the lane change behavior of tracked vehicles is modeled using dynamic Bayesian networks, and the experiments are performed on real-world vision data.

C. Trajectories

The use of vehicle trajectories to characterize and learn on-road vehicle behaviors has emerged in the past few years. A trajectory is typically defined as a data sequence, consisting of several concatenated state vectors from tracking, meaning an indexed sequence of positions and velocities over a given time window. Using a time window of 1 s, for example, can mean trajectories consisting of 25–30 samples, depending on the frame rate of the camera.

In [185], variational Gaussian mixture modeling is used to classify and predict the long-term trajectories of vehicles, using simulated data. In [4], highway trajectories are recorded using stereo vision, and clustering is performed to model the typical trajectories encountered in highway driving, with classification performed using hidden Markov modeling. In [184], trajectories are classified using a rotation-invariant version of the longest common subsequence as the similarity metric between trajectories. Vehicle trajectories are used to characterize behavior at roundabouts in [131], using the quaternion-based rotationally invariant longest common subsequence (QRLCS) metric to match observed trajectories to a database of prerecorded trajectories. Similar work is carried out in [187] for vehicles at intersections.

D. Behavior Classification

Classification of vehicle behavior is performed using a variety of techniques, which is dependent on the objective of the study. In [16], [77], [87], and [164], the vehicle detection task encompasses the classification of vehicle behavior. This is to say that these studies aim to detect vehicles that overtake the ego vehicle; thus, in these cases, vehicle detection is synonymous with vehicle behavior characterization. By contrast, yaw rate information is used in [82] to characterize the turning behavior of tracked vehicles. In this case, measured motion characteristics describe the specific maneuver.

In studies that explicitly classify vehicle behavior, we see a preponderance of generative modeling. In [185], Gaussian mixture modeling is used, which provides distribution over the prediction, complete with a point estimate (the conditional mean) and a covariance to convey uncertainty. In [3], the likelihood of the interacting multiple model tracking is used to classify the tracked vehicle's turning behavior, complete with a transition probability. In [6], Bayesian networks are used for classifying the vehicle's behavior and predicting the vehicle's lane change. In [4], hidden Markov modeling is used to model each of the prototypical trajectories, which are learned using clustering. In [186], the vehicle behavior is also modeled as a Markov process, with observations coming from the vehicle's instantaneous state vector. Table IV highlights some representative works in vision-based on-road behavior analysis.

VI. Discussion and Future Directions

Here, we provide discussion, critiques, and perspective on vision-based vehicle detection, tracking, and on-road behavior analysis.

A. Vehicle Detection

In recent years, the feature representations used in monocular vehicle detection have transitioned from simpler image features like edges and symmetry to general and robust feature sets for vehicle detection. These feature sets, now common in the computer vision literature, allow for direct classification and detection of objects in images. HOG and Haar-like features are extremely well represented in the vehicle detection literature, as they are in the object detection literature [34], [35]. While early works heavily relied upon symmetry, symmetry is typically not robust enough to detect vehicles by itself. The on-road environment features many objects that feature high symmetry, such as man-made structures and traffic signs. Many daytime vehicle detection studies that have used symmetry as the main feature do not provide an extensive experimental evaluation. Symmetry is more likely to serve to generate regions of interest

TABLE IV
REPRESENTATIVE WORKS IN ON-ROAD BEHAVIOR ANALYSIS

Research Study	Context-Specific?	Specific Maneuver	Trajectory-based?	Classification/ Inference	Description	Comments
Diaz-Alonso et al., 2008 [16]	No	Overtaking vehicles	No	Template matching score	Detection and tracking of overtaking vehicles	Maneuver-specific classification. Overtake maneuvers are identified by virtue of detecting the oncoming vehicle in the blind spot.
Cherng et al., 2009 [88]	Yes	None	No	Neural network	Dynamic visual model of typical on-road behavior; saliency used to detect unusual and critical situations.	The saliency based model is fairly unique, in that critical situations are identified by their visual manifestation in the image, and not by 3D motion.
Barth and Franke, 2010 [3]	Yes	Turning behavior	No	Interacting multiple model likelihood	Velocity and yaw-rate estimation used to infer turning behavior of oncoming vehicles.	Turning behavior was characterized by estimating velocity and yaw rate, using multiple motion models, choosing the best fit with IMM. The motion models incorporated constant velocity and constant acceleration modes, for vehicle velocity and vehicle yaw rate.
Geiger and Kitt, 2010 [96]	Yes	No	No	Support vector machine	Histograms of scene flow used to classify intersection vs. non-intersection driving environment.	The approach allowed the system to characterize the current driving context using coarse motion cues.
Hermes et al., 2010 [131]	Yes	No	Yes	Augmented particle filter	Vehicle motion is matched to 44 prototypes using QRLCS distance.	Vehicles were tracked at a roundabout, and their motion over time was matched them to previously-observed trajectories.
Sivaraman et al., 2011 [4]	Yes	No	Yes	hidden Markov model	Unsupervised clustering of observed on-road trajectories.	Vehicles were tracked on highways, and clustering was performed on the archived vehicle trajectories, from which typical highway driving patterns emerged.
Kasper et al., 2011 [6]	No	Lane change	No	Dynamic Bayesian network	Dynamic Bayesian network is used to predict lane changes of other vehicles.	Modeling each vehicle's driving environment with its own occupancy grid, and measuring the motion, a probabilistic framework based on Bayesian Networks for predicting the likelihood of changing lanes.
Garcia et al., 2012 [164]	No	Overtaking vehicles	No	Optical flow direction, intensity	Optical flow is used to detect overtaking vehicles.	Overtake maneuvers are identified by virtue of detecting the oncoming vehicle in the peripheral of the forward-looking camera.

for further feature extraction and classification [18], [24], [36], [62]. In [38], a novel analysis of the symmetery of HOG features is used to detect vehicles.

Learning and classification approaches have also transitioned in recent years. While neural networks can deliver acceptable performance, their popularity in research communities has waned. This is mainly due to the fact that neural network training features many parameters to tune and converges to a local optimum over the training set. Competing discriminative methods converge to a global optimum over the training set, which provides nice properties for further analysis, such as fitting posterior probabilities [188]. Studies that use SVM and AdaBoost classifications are far more prevalent in the modern vehicle literature, a trend which mirrors similar movement in the computer vision and machine learning research communities.

While monocular vehicle detection has been an active research area for quite some time, open challenges still remain. It is challenging to develop a single detector that works equally well in all the varied conditions encountered on the road.

Scene-specific classifiers, categorizing the on-road scene as urban versus highway and/or cloudy versus sunny, could augment the performance of vehicle detectors, utilizing image classification as a preprocessing step [189].

Monocular vehicle detection largely relies on a feature extraction–classification paradigm, which is based on machine learning. This approach works very well when the vehicle is fully visible. In particular, robustly detecting partially occluded vehicles using monocular vision remains an open challenge. Early work in this area is ongoing based on detecting vehicles as a combination of independent parts [51], but detecting partially occluded vehicles remains a challenging research area. Using parts to detect vehicles has been implemented in [75], but the recognition still has difficulty with occlusions. Future works will need to include motion cues into monocular vehicle detection to identify vehicles as they appear while seamlessly integrating them into machine learning frameworks.

Object detection using stereo vision has also made great progress over the past decade. Advances in stereo matching yield much cleaner, less noisy, and denser disparity maps [109]. Improved stereo matching enables more robust scene segmentation, based on motion and structural cues [85]. While stereo matching and reconstruction has improved, stereo-vision methods typically recognize vehicles in a bottom-up manner. This is to say that the typical paradigm consists of ego-motion compensation, tracking feature points in 3-D, distinguishing static from moving points, and associating moving points into moving objects [84]. Finally, moving objects are labeled as vehicles by fitting a cuboid model [137] or clustering [130]. While these methods have made great progress, complex scenes still present difficulty [85]. Integration of machine learning methodology could increase the robustness of existent stereo-vision approaches and has the potential to simplify the vehicle detection task. Research along these lines has been performed by using machine-learning-based detection on the monocular plane, integrating stereo vision for validation and tracking [5], [118], [126]. Future work could involve a more principled machine learning approach, learning on motion cues, image cues, and disparity or depth cues.

As the cost of active sensors, such as radar and lidar, continues to decrease, integration of these sensing modalities with vision will continue to increase in prevalence. Automotive radar and lidar systems are fairly mature in their ability to detect objects and obstacles, but their ability to distinguish vehicles from other objects is limited. Currently, radar and lidar systems distill their detections into multiple object lists. As lane-tracking cameras become standard options on serial production vehicles, the opportunity to integrate vision with active sensing technology will present itself. Vision allows an intuitive level of semantic abstraction that is not otherwise available with lidar or radar. Many studies detect vehicles with one modality and validate with the other [164], [178]. Others detect vehicles with vision and detect range with radar or lidar [14]. Future works will need a principled object-level fusion of vision and active sensors for vehicle detection [177]. Such an information fusion could reduce estimation covariance and enhance robustness, although the asynchronous nature of the multiple modalities will need to be handled [190].

B. Vehicle Tracking

While early works in monocular on-road vehicle tracking used template matching [24], recursive Bayesian filtering approaches, such as Kalman filtering [50] and particle filtering [75], have become the norm. Estimation of a tracked vehicle's position in the image plane can be augmented by using optical flow, as vision-based vehicle detectors can fluctuate in their pixel locations from frame to frame, even for a static object. Future work in monocular vehicle tracking will pursue information fusion of the optical flow motion and the motion from vehicle detections in consecutive frames. To this end, low-level motion cues can improve monocular vehicle tracking and provide a basis for enhanced track management, when detections are dropped in certain frames.

Vehicle tracking in the stereo-vision domain, by contrast, is extremely mature. Indeed, most vehicle detection approaches using stereo vision are based on motion and tracking, from interest points to 3-D points, to clustered objects, to cuboid vehicle models [85]. Estimation of the vehicle's yaw rate has emerged as a very important cue, for identifying turning behavior and for improved prediction of the vehicle's motion [82]. Extended Kalman filtering for vehicle tracking has increased in popularity to accommodate the nonlinear observation and motion models [82], [118]. Particle filtering has also increased in popularity for vehicle tracking while dispensing with some assumptions required for Kalman filtering [84]. Interacting multiple models, however, seem to best account for the different modes exhibited by vehicle motion on the road. In [3], motion modes for constant velocity and yaw rate, constant velocity and yaw acceleration, constant acceleration and yaw rate, and constant acceleration and yaw acceleration were all available to best model the motion of a tracked vehicle, with a likelihood measurement to choose the best fit. Such modeling allows for sudden changes in vehicle motion, without simply compensating for the changes with the noise terms. Including a transition probability between modes increases the estimation stability and draws powerful parallels with established techniques in the analysis of Markov chains.

C. On-Road Behavior Analysis

On-road behavior analysis speaks to a higher level of semantic interpretation of the driving environment and is the least mature area of research. While this level of analysis is dependent on robust vehicle detection and vehicle tracking, the aim is to answer questions beyond those answered by detection and tracking. These issues include identification of maneuvers, characterization of vehicle behavior, and long-term motion prediction. Only in the past few years have vehicle detection and tracking methodologies become sufficiently mature to enable the exploration of these deeper questions.

Recognition of specific maneuvers has so far been coincident with detection of vehicles in a particular location relative to the ego vehicle [16] or directly inferred by dynamical information from tracking vehicles [82]. While dynamical models of vehicle motion are well established, identification of vehicle maneuvering has so far been implemented in a maneuver-specific manner. Future works will need to formulate general models of vehicle maneuvering, which allow for picking the most likely current maneuver from a pool of available classes [6]. This could be implemented using generative models [186], or all-versus-one discriminative classification for maneuver detection. Ongoing research will also have to account for various traffic and road conditions, with different models for urban versus highway driving, arterial versus intersections, and free-flow versus congested driving conditions. Predicting a vehicle's maneuver requires making a decision with a partially observed sequence of tracking data. Integration of latent variable models [71], [191] will play a role in the identification of maneuvers.

Recognition of overtaking maneuvers has been an active research area [16], [164]. The main difference between an overtake and a lane change is the presence of a vehicle in the target lane and acceleration to keep a safe distance. A similar distinction exists between so-called undertaking and lane changing. In real-world settings, vehicle motion is constrained by traffic, infrastructure, and other vehicles. Modeling the interactions between vehicles is an open area of research, in the context of vehicle behavior characterization. In [192], the distances and time gaps between vehicles were used as a feature for predicting the driver's intent to change lanes. Further research will need to be conducted to characterize the interactions between vehicles and their role in on-road behavior. The vehicle dynamics associated with a specific maneuver can be learned, in a data-driven manner, from the controller area network (CAN) bus of the ego vehicle and the presence and absence of vehicles in the target lane. It would then be a research question to determine an appropriate and sufficient distillation of the data to classify and predict those maneuvers in unlabeled data. The inertial sensors available on the ego vehicle provide more data signals, each of higher precision, than can be reasonably measured using vision. The research question will be concerned with detecting maneuvers based on the parameters that are observable and robustly estimated.

In the analysis and characterization of vehicle behavior, a major challenge will be in identifying erratic, abnormal, and aggressive driving by other vehicles. While identifying specific maneuvers can be formulated as a well-defined problem, for example, turning [3], characterizing another vehicle's behavior remains an open question. Given the tracking of vehicles with respect to their own lanes, weak cues such as vehicle's veering within its lane or crossing over lane boundaries could be used. More likely, research studies will try to characterize normal driving behavior for a given context in a data-driven manner and identify abnormal trajectories by measuring the model fit of an observed vehicle trajectory [4].

Filtering approaches like the Kalman filter can provide a good estimate of a vehicle's position, velocity, and other parameters one frame ahead of time, at typical camera frame rates between 10 and 25 frames per second. Long-term motion prediction requires an estimate of the vehicle's motion 1–2 s, or 25–50 frames, ahead of time, outside the capabilities of conventional filtering techniques. Long-term motion classification and prediction will involve further research into learning and modeling of vehicle trajectories. An enhanced understanding of vehicle trajectories will allow onboard systems to infer the intent of other vehicle's drivers, based on sequential tracking measurements from vision-based systems.

A transition will need to take place from measuring a tracked vehicle's motion in the coordinate frame of the ego vehicle to position-independent coordinate frames, such as the sequence of angles approach used in [193]. While vision-based vehicle tracking research tends to measure the position and motion of vehicles in the coordinate frame of the ego vehicle, there needs to be a movement toward understanding the motion and behavior of other vehicles as independent traffic agents. Trajectories, i.e., sequences of vehicle tracking data, may be well modeled by Markov chains, but there will need to be a transition probability between sequences, to account for drivers changing their minds at the last minute. To this end, we foresee learned trajectory models working in concert with established tracking approaches such as interacting multiple models. A full vehicle motion understanding engine would include multiple trackers with distinct motion models to estimate vehicle state in the short term, interacting multiple models to identify vehicle maneuvering in the medium term, and trajectory learning to predict vehicle motion in the long term. Associated issues, such as data windowing and online model updates, will also need to be addressed.

D. Benchmarks

We briefly discuss the benchmarks that are publicly available and commonly used performance metrics in vehicle detection, tracking, and behavior analysis. Table V provides a summary of some of the publicly available and widely used data sets. While these data sets are available, we note that it is still common practice for research groups to capture their own video data, for use in training and testing. Like many computer vision and machine learning research areas, vehicle detection is not so easily summarized by one standard data set. The driving environment features high variability in illumination, weather, and road conditions. Furthermore, vehicle models and road structure differ across the globe, meaning European, Asian, and North American data sets will certainly differ.

Until recently, few vehicle detection studies, particularly in the monocular domain, have evaluated their performance in realistic conditions, using real-world on-road video. Prior works would report results on static images, often subsets of databases that were created for object recognition, such as Caltech 1999, 2001 [194], [195]. In the past few years, this has changed, and published research works now regularly report results on real-world video. A lot of research are funded by automotive manufacturers, and their CAN signals are proprietary information. As such, on-road vehicle detection does not have a strong history of standard benchmark data sets, unlike other computer vision disciplines [196]. It is now becoming a standard practice for researchers to release data sets, source code, and even camera calibration parameters, which will help the research community make further progress in on-road vehicle detection. However, only very recently have published works started to make their data sets publicly available, for evaluation and benchmarking by others in the research community.

TABLE V
VEHICLE DETECTION BENCHMARKS

Dataset	Description	Metrics
Caltech 1999 [194], 2001 [195]	Static images of vehicles in a variety of poses.	Recall and Precision
PETS 2001 [197]	Testing set of some 2867 frames, from two cameras. Includes videos of preceding vehicles viewed through the front windshield, and a video of oncoming vehicles viewed through the rear windshield.	Recall and Precision
LISA, 2010. [49]	Three short videos, 1500, 300, and 300 frames, comprising highway and urban driving. Monocular detection of preceding vehicles only.	Recall and Precision
Caraffi, 2012. [68]	Several videos, comprising some 20 minutes of driving on italian highways.	Recall and Precision
KITTI, 2012 [198]	Extended videos from stereo pairs, complete with CAN-bus and lidar data. Monocular, stereo-vision, and sensor fusion evaluation is possible.	Recall, precision, orientation, 3D position, 3D speed, object re-identification

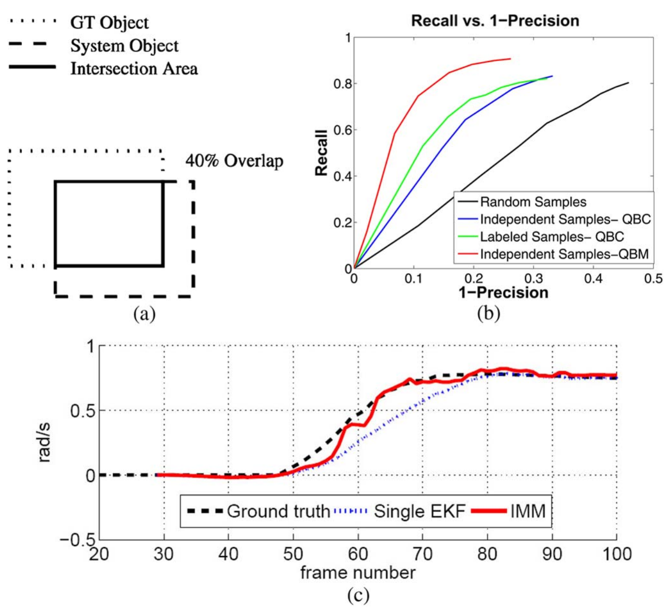

Fig. 6. Performance metrics. (a) Overlap criterion used for labeling detections as true or false positives in the image plane [196]. (b) Plotting the recall versus 1 − precision for monocular vehicle detection [37]. (c) Plotting the estimated yaw rate versus time, along with ground truth [3].

In monocular vehicle detection, commonly used benchmarks quantify recall of true positive vehicles and false alarms. Given a ground truth annotation G_i in a frame and a detection D_i, the detection is deemed a true positive if the overlap of the two exceeds a threshold τ, as shown in (2). Fig. 6(a) depicts the overlap criterion for detections and ground truth bounding boxes in the image plane. Thus

$$D_i = \begin{cases} \text{True positive}, & \text{if } \frac{G_i \cap D_i}{G_i \cup D_i} > \tau \\ \text{False positive}, & \text{otherwise.} \end{cases} \quad (2)$$

TABLE VI
MONOCULAR VEHICLE DETECTION METRICS

Performance Metric	Definition
Detection Rate/True Positive Rate/ Recall	$\frac{\#\text{ True Positives}}{\#\text{ Vehicles}}$
False Positive Rate	$\frac{\#\text{ False Positives}}{\#\text{ Possible Bounding Boxes}}$
False Detection Rate/ 1 − Precision	$\frac{\#\text{False Positives}}{\#\text{ True Positives}+\#\text{ False Positives}}$
False Positives per Frame/ False Positives per Image	$\frac{\#\text{ False Positives}}{\#\text{ Frames}}$

TABLE VII
STEREO-VISION TRACKING METRICS

Performance Metric	Definition
Mean Absolute Error	$\frac{1}{N}\sum \lvert x_{\text{estimated}} - x_{\text{ground truth}}\rvert$
Standard Deviation of Error	$\sqrt{\frac{1}{N}\sum (x_{\text{estimated}} - x_{\text{ground truth}})^2}$

A detection D_i that does not have sufficient overlap with the ground truth annotation, including zero overlap, is deemed a false positive. Detection and false positives are counted over a video sequence. Dividing the number of true and false positives by a variety of denominators yields a set of metrics that have been used in the field. For detections, the true positive rate, or recall, is almost uniformly used. For false alarms, common metrics include 1 − precision or false detection rate, false positive rate, false positives per frame, and false positives per object. Table VI defines these terms. Monocular tracking studies typically use the same metrics for tracking as for detection. While [196] defined useful metrics to quantify the consistency of track identification numbers, their use is virtually unseen in on-road vehicle tracking works.

Stereo-vision vehicle detection studies typically do not report true positive and false positive rates, although this has recently begun to change [118]. Instead, stereo-vision studies tend to focus on estimation accuracy for the motion parameters of a tracked vehicle, including position, velocity, and yaw rate [3], [84]. The performance is typically quantified with the mean absolute error and the standard deviation in estimation. Table VII defines these terms, using x for the various parameters estimated using a given tracker, and N is the number of frames.

Vehicle behavior studies, lacking a uniformly established system definition, lack a standard set of performance metrics. Context-based detection [16], [125] studies will often use similar metrics to those used in monocular vision. Studies concerned with trajectories [4], [131] will often report classification accuracy by casting the problem as a multiclass classification task. As this area of study matures, a standard set of performance metrics will emerge. Performance evaluation will need to include a combination of metrics for motion prediction, e.g., mean absolute error, and metrics for classification accuracy for multiclass inference, e.g., confusion matrices.

Publicly available vehicle detection benchmarks are becoming more common, but they are still quite rare for vehicle tracking. Part of the difficulty has lied with generating ground truth for the 3-D positions and velocities of vehicles in video sequences. The newly released KITTI database [198] contains extensive video data captured with a calibrated stereo rig, as well as synchronized lidar data, which can be used as a ground truth for vehicle localization. The recently released data set in [68] has also contained lidar data so that vision-based detection accuracy can be evaluated as a function of longitudinal distance. However, ground truth for dynamical parameters such as velocity and vehicle yaw rate must necessarily come from the tracked vehicle's own CAN, which is not feasible outside of controlled and orchestrated trials.

Benchmark data sets for on-road behavior analysis do not currently exist. This lack of benchmarks largely has to do with the infancy of this research area. Semantically meaningful labels for ground truth are still not standard. The various studies in the field pursue different objectives: identifying critical and abnormal situations [88], detecting specific maneuvers [3], [164], and predicting long-term motion [131]. As such, there is currently not a unified objective or a set of performance metrics for this area of research. Furthermore, capturing and labeling a relevant data set is a challenging task, as such a benchmark requires all steps, from vehicle detection to tracking to behavior analysis, all labeled and ground truth, with globally applicable and accepted performance metrics and interpretations. As this area of research matures, meaningful and widely accepted research objectives, goals, and performance metrics will emerge; and standard benchmarks will become more common. More comprehensive benchmark data sets will need to be published in order to streamline the efforts of the research community.

VII. Concluding Remarks

In this study, we have provided a review of the literature addressing on-road vehicle detection, vehicle tracking, and behavior analysis using vision. We have placed vision-based vehicle detection in the context of sensor-based on-road perception and provided comparisons with complimentary technologies, namely, radar and lidar. We have provided a survey of the past decade's progress in vision-based vehicle detection, for monocular and stereo-vision sensor configurations. Included in our treatment of vehicle detection is the treatment of camera placement, nighttime algorithms, sensor-fusion strategies, and real-time architecture. We have reviewed vehicle tracking in the context of vision-based sensing, addressing monocular applications in the image plane, and stereo-vision applications in the 3-D domain, including various filtering techniques and motion models. We have reviewed the state of the art in on-road behavior analysis, addressing specific maneuver detection, context analysis, and long-term motion classification and prediction. Finally, we have provided critiques, discussion, and outlooks on the direction of the field. While vision-based vehicle detection has matured significantly over the past decade, a deeper and more holistic understanding of the on-road environment will remain an active area of research in the coming years.

Acknowledgment

The authors would like to thank the reviewers for their comments and suggestions and Dr. B. Morris and A. Mogelmose, their colleagues, for their valuable assistance and interactions.

References

[1] Department of Transportation National Highway Traffic Safety Administration, Traffic safety facts 2011. [Online]. Available: http://www-nrd.nhtsa.dot.gov

[2] Z. Sun, G. Bebis, and R. Miller, "On-road vehicle detection: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 5, pp. 694–711, May 2006.

[3] A. Barth and U. Franke, "Tracking oncoming and turning vehicles at intersections," in *Proc. 13th Int. IEEE ITSC*, Sep. 2010, pp. 861–868.

[4] S. Sivaraman, B. T. Morris, and M. M. Trivedi, "Learning multi-lane trajectories using vehicle-based vision," in *Proc. IEEE Int. Conf. Comput. Vision Workshop*, 2011, pp. 2070–2076.

[5] S. Sivaraman and M. M. Trivedi, "Combining monocular and stereo-vision for real-time vehicle ranging and tracking on multilane highways," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2011, pp. 1249–1254.

[6] D. Kasper, G. Weidl, T. Dang, G. Breuel, A. Tamke, and W. Rosenstiel, "Object-oriented Bayesian networks for detection of lane change maneuvers," in *Proc. IEEE IV*, Jun. 2011, pp. 673–678.

[7] X. Mao, D. Inoue, S. Kato, and M. Kagami, "Amplitude-modulated laser radar for range and speed measurement in car applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 1, pp. 408–413, Mar. 2012.

[8] S. Tokoro, K. Kuroda, A. Kawakubo, K. Fujita, and H. Fujinami, "Electronically scanned millimeter-wave radar for pre-crash safety and adaptive cruise control system," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2003, pp. 304–309.

[9] 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 *Proc. IEEE IV*, Jun. 2011, pp. 163–168.

[10] S. Sato, M. Hashimoto, M. Takita, K. Takagi, and T. Ogawa, "Multilayer lidar-based pedestrian tracking in urban environments," in *Proc. IEEE IV*, Jun. 2010, pp. 849–854.

[11] F. Garcia, P. Cerri, A. Broggi, J. M. Armingo, and A. de la Escalera, *Vehicle Detection Based on Laser Radar*. New York, NY, USA: Springer-Verlag, 2009.

[12] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An Invitation to 3-D Vision: From Images to Geometric Models*. New York, NY, USA: Springer-Verlag, 2003.

[13] S. Sivaraman and M. Trivedi, "Integrated lane and vehicle detection, localization, and tracking: A synergistic approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 907–917, Jun. 2013.

[14] M. Mahlisch, R. Schweiger, W. Ritter, and K. Dietmayer, "Sensorfusion using spatio-temporal aligned video and lidar for improved vehicle detection," in *Proc. IEEE Intell. Veh. Symp.*, 2006, pp. 424–429.

[15] A. Wedel and U. Franke, "Monocular video serves radar-based emergency braking," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2007, pp. 93–98.

[16] J. Diaz Alonso, E. Ros Vidal, A. Rotter, and M. Muhlenberg, "Lane-change decision aid system based on motion-driven vehicle tracking," *IEEE Trans. Veh. Technol.*, vol. 57, no. 5, pp. 2736–2746, Sep. 2008.

[17] K.-T. Song and H.-Y. Chen, "Lateral driving assistance using optical flow and scene analysis," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2007, pp. 624–629.

[18] N. Blanc, B. Steux, and T. Hinz, "LaRASideCam: A fast and robust vision-based blindspot detection system," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2007, pp. 480–485.

[19] B.-F. Lin, Y.-M. Chan, L.-C. Fu, P.-Y. Hsiao, L.-A. Chuang, S.-S. Huang, and M.-F. Lo, "Integrating appearance and edge features for sedan vehicle detection in the blind-spot area," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 737–747, Jun. 2012.

[20] T. Gandhi and M. M. Trivedi, "Parametric ego-motion estimation for vehicle surround analysis using an omnidirectional camera," *Mach. Vis. Appl.*, vol. 16, no. 2, pp. 85–95, Feb. 2005.

[21] T. Gandhi and M. Trivedi, "Vehicle surround capture: Survey of techniques and a novel omni-video-based approach for dynamic panoramic surround maps," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 3, pp. 293–308, Sep. 2006.

[22] W.-C. Chang and C.-W. Cho, "Real-time side vehicle tracking using parts-based boosting," in *Proc. IEEE Int. Conf. SMC*, 2008, pp. 3370–3375.

[23] A. Broggi, A. Cappalunga, S. Cattani, and P. Zani, "Lateral vehicles detection using monocular high resolution cameras on TerraMax," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2008, pp. 1143–1148.

[24] W. Liu, X. Wen, B. Duan, H. Yuan, and N. Wang, "Rear vehicle detection and tracking for lane change assist," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2007, pp. 252–257.

[25] C. Hoffmann, "Fusing multiple 2D visual features for vehicle detection," in *Proc. IEEE Intell. Veh. Symp.*, 2006, pp. 406–411.

[26] C. Hilario, J. Collado, J. Armingol, and A. de la Escalera, "Pyramidal image analysis for vehicle detection," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2005, pp. 88–93.

[27] J. Arrospide, L. Salgado, M. Nieto, and F. Jaureguizar, "On-board robust vehicle detection and tracking using adaptive quality evaluation," in *Proc. 15th IEEE ICIP*, Oct. 2008, pp. 2008–2011.

[28] Y.-M. Chan, S.-S. Huang, L.-C. Fu, and P.-Y. Hsiao, "Vehicle detection under various lighting conditions by incorporating particle filter," in *Proc. IEEE ITSC*, 2007, pp. 534–539.

[29] Z. Kim, "Realtime obstacle detection and tracking based on constrained Delaunay triangulation," in *Proc. IEEE ITSC*, Sep. 2006, pp. 548–553.

[30] J. Nuevo, I. Parra, J. Sjoberg, and L. Bergasa, "Estimating surrounding vehicles' pose using computer vision," in *Proc. 13th ITSC*, Sep. 2010, pp. 1863–1868.

[31] C. Idler, R. Schweiger, D. Paulus, M. Mahlisch, and W. Ritter, "Real-time vision based multi-target-tracking with particle filters in automotive applications," in *Proc. IEEE Intell. Veh. Symp.*, 2006, pp. 188–193.

[32] H.-Y. Chang, C.-M. Fu, and C.-L. Huang, "Real-time vision-based preceding vehicle tracking and recognition," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2005, pp. 514–519.

[33] B. Aytekin and E. Altug, "Increasing driving safety with a multiple vehicle detection and tracking system using ongoing vehicle shadow information," in *Proc. IEEE Int. Conf. SMC*, Oct. 2010, pp. 3650–3656.

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

[35] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Vis. Pattern Recog. Conf.*, 2001, pp. I-511–I-518.

[36] S. Teoh and T. Brunl (2012, Sep.). Symmetry-based monocular vehicle detection system. *Mach. Vis. Appl.* [Online]. *23(5)*, pp. 831–842. Available: http://dx.doi.org/10.1007/s00138-011-0355-7

[37] S. Sivaraman and M. M. Trivedi, "Active learning for on-road vehicle detection: A comparative study," *Mach. Vis. Appl.—Special Issue Car Navigation and Vehicle Systems*, pp. 1–13, Dec. 2011.

[38] M. Cheon, W. Lee, C. Yoon, and M. Park, "Vision-based vehicle detection system with consideration of the detecting location," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1243–1252, Sep. 2012.

[39] R. Wijnhoven and P. de With, "Unsupervised sub-categorization for object detection: Finding cars from a driving vehicle," in *Proc. IEEE ICCV Workshops*, Nov. 2011, pp. 2077–2083.

[40] T. Machida and T. Naito, "GPU and CPU cooperative accelerated pedestrian and vehicle detection," in *Proc. IEEE ICCV Workshops*, Nov. 2011, pp. 506–513.

[41] R. M. Z. Sun and G. Bebis, "Monocular precrash vehicle detection: Features and classifiers," *IEEE Trans. Image Process.*, vol. 15, no. 7, pp. 2019–2034, Jul. 2006.

[42] D. Ponsa, A. Lopez, F. Lumbreras, J. Serrat, and T. Graf, "3D vehicle sensor based on monocular vision," in *Proc. IEEE Intell. Transp. Syst.*, Sep. 2005, pp. 1096–1101.

[43] D. Ponsa, A. Lopez, J. Serrat, F. Lumbreras, and T. Graf, "Multiple vehicle 3D tracking using an unscented Kalman," in *Proc. IEEE Intell. Transp. Syst.*, Sep. 2005, pp. 1108–1113.

[44] J. Cui, F. Liu, Z. Li, and Z. Jia, "Vehicle localization using a single camera," in *Proc. IEEE IV*, Jun. 2010, pp. 871–876.

[45] G. Y. Song, K. Y. Lee, and J. W. Lee, "Vehicle detection by edge-based candidate generation and appearance-based classification," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2008, pp. 428–433.

[46] D. Withopf and B. Jahne, "Learning algorithm for real-time vehicle tracking," in *Proc. IEEE ITSC*, Sep. 2006, pp. 516–521.

[47] A. Haselhoff, S. Schauland, and A. Kummert, "A signal theoretic approach to measure the influence of image resolution for appearance-based vehicle detection," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2008, pp. 822–827.

[48] A. Haselhoff and A. Kummert, "A vehicle detection system based on Haar and triangle features," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2009, pp. 261–266.

[49] S. Sivaraman and M. Trivedi, "A general active-learning framework for on-road vehicle recognition and tracking," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 267–276, Jun. 2010.

[50] A. Haselhoff and A. Kummert, "An evolutionary optimized vehicle tracker in collaboration with a detection system," in *Proc. 12th Int. IEEE Conf. ITSC*, Oct. 2009, pp. 1–6.

[51] S. Sivaraman and M. M. Trivedi, "Real-time vehicle detection using parts at intersections," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2012, pp. 1519–1524.

[52] P. Negri, X. Clady, S. M. Hanif, and L. Prevost, "A cascade of boosted generative and discriminative classifiers for vehicle detection," *EURASIP J. Adv. Signal Process.*, vol. 2008, p. 136, Jan. 2008.

[53] D. Lowe, "Object recognition from local scale-invariant features," in *Proc. Int. Conf. Comput. Vis.*, 1999, pp. 1150–1157.

[54] X. Zhang, N. Zheng, Y. He, and F. Wang, "Vehicle detection using an extended hidden random field model," in *Proc. 14th Int. IEEE Conf. ITSC*, Oct. 2011, pp. 1555–1559.

[55] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.

[56] Y. Zhang, S. Kiselewich, and W. Bauson, "Legendre and Gabor moments for vehicle recognition in forward collision warning," in *Proc. IEEE ITSC*, Sep. 2006, pp. 1185–1190.

[57] C.-C. R. Wang and J.-J. Lien, "Automatic vehicle detection using local features: A statistical approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 1, pp. 83–96, Mar. 2008.

[58] O. Ludwig and U. Nunes, "Improving the generalization properties of neural networks: An application to vehicle detection," in *Proc. 11th Int. IEEE Conf. ITSC*, Oct. 2008, pp. 310–315.

[59] C. Cortes and V. Vapnik, "Support vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[60] Y. Freund and R. Schapire, "A short introduction to boosting," *J. Jpn. Soc. Artif. Intell.*, vol. 14, no. 5, pp. 771–780, Sep. 1999.

[61] Q. Yuan, A. Thangali, V. Ablavsky, and S. Sclaroff, "Learning a family of detectors via multiplicative kernels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 514–530, Mar. 2011.

[62] T. Liu, N. Zheng, L. Zhao, and H. Cheng, "Learning based symmetric features selection for vehicle detection," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2005, pp. 124–129.

[63] A. Khammari, F. Nashashibi, Y. Abramson, and C. Laurgeau, "Vehicle detection combining gradient analysis and AdaBoost classification," in *Proc. IEEE Intell. Transp. Syst.*, Sep. 2005, pp. 66–71.

[64] I. Kallenbach, R. Schweiger, G. Palm, and O. Lohlein, "Multi-class object detection in vision systems using a hierarchy of cascaded classifiers," in *Proc. IEEE Intell. Veh. Symp.*, 2006, pp. 383–387.

[65] T. Son and S. Mita, "Car detection using multi-feature selection for varying poses," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2009, pp. 507–512.

[66] D. Acunzo, Y. Zhu, B. Xie, and G. Baratoff, "Context-adaptive approach for vehicle detection under varying lighting conditions," in *Proc. IEEE ITSC*, 2007, pp. 654–660.

[67] W.-C. Chang and C.-W. Cho, "Online boosting for vehicle detection," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 3, pp. 892–902, Jun. 2010.

[68] C. Caraffi, T. Vojii, J. Trefny, J. Sochman, and J. Matas, "A system for real-time detection and tracking of vehicles from a single car-mounted camera," in *Proc. 15th Int. IEEE Conf. ITSC*, 2012, pp. 975–982.

[69] A. Jazayeri, H. Cai, J. Y. Zheng, and M. Tuceryan, "Vehicle detection and tracking in car video based on motion model," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 583–595, Jun. 2011.

[70] A. Chavez-Aragon, R. Laganiere, and P. Payeur, "Vision-based detection and labeling of multiple vehicle parts," in *Proc. 14th Int. IEEE Conf. ITSC*, Oct. 2011, pp. 1273–1278.

[71] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.

[72] P. Felzenszwalb, R. Girshick, and D. McAllester, "Cascade object detection with deformable part models," in *Proc. IEEE Comput. Vis. Pattern Recog.*, 2010, pp. 2241–2248.

[73] A. Takeuchi, S. Mita, and D. McAllester, "On-road vehicle tracking using deformable object model and particle filter with integrated likelihoods," in *Proc. IEEE IV*, Jun. 2010, pp. 1014–1021.

[74] H. Niknejad, S. Mita, D. McAllester, and T. Naito, "Vision-based vehicle detection for nighttime with discriminately trained mixture of weighted deformable part models," in *Proc. 14th Int. IEEE Conf. ITSC*, 2011, pp. 1560–1565.

[75] H. Tehrani Niknejad, A. Takeuchi, S. Mita, and D. McAllester, "On-road multivehicle tracking using deformable object model and particle filter with improved likelihood estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 748–758, Jun. 2012.

[76] S. Sivaraman and M. Trivedi, "Vehicle detection by independent parts for urban driver assistance," *IEEE Trans. Intell. Transp. Syst.*, 2013, DOI 10.1109/TITS.2013.2264314.

[77] Y. Zhu, D. Comaniciu, M. Pellkofer, and T. Koehler, "Reliable detection of overtaking vehicles using robust information fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 4, pp. 401–414, Dec. 2006.

[78] P. Chang, D. Hirvonen, T. Camus, and B. Southall, “Stereo-based object detection, classification, and quantitative evaluation with automotive applications,” in *Proc. IEEE Comput. Soc. Conf. CVPR Workshops*, Jun. 2005, p. 62.
[79] I. Cabani, G. Toulminet, and A. Bensrhair, “Contrast-invariant obstacle detection system using color stereo vision,” in *Proc. 11th Int. IEEE Conf. ITSC*, Oct. 2008, pp. 1032–1037.
[80] U. Franke, C. Rabe, H. Badino, and S. Gehrig, “6D-vision: Fusion of stereo and motion for robust environment perception,” in *Proc. DAGM*, 2005, pp. 216–223.
[81] H. Badino, R. Mester, J. Wolfgang, and U. Franke, “Free space computation using stochastic occupancy grids and dynamic programming,” in *Proc. ICCV Workshop Dyn. Vis.*, 2007, pp. 144–156.
[82] A. Barth and U. Franke, “Estimating the driving state of oncoming vehicles from a moving platform using stereo vision,” *IEEE Trans. Intell Transp. Syst.*, vol. 10, no. 4, pp. 560–571, Dec. 2009.
[83] A. Broggi, A. Cappalunga, C. Caraffi, S. Cattani, S. Ghidoni, P. Grisleri, P. Porta, M. Posterli, and P. Zani, “TerraMax vision at the urban challenge 2007,” *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 1, pp. 194–205, Mar. 2010.
[84] R. Danescu, F. Oniga, and S. Nedevschi, “Modeling and tracking the driving environment with a particle-based occupancy grid,” *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1331–1342, Dec. 2011.
[85] F. Erbs, A. Barth, and U. Franke, “Moving vehicle detection by optimal segmentation of the dynamic stixel world,” in *Proc. IEEE IV*, Jun. 2011, pp. 951–956.
[86] M. Perrollaz, J.-D. Yoder, A. Nègre, A. Spalanzani, and C. Laugier, “A visibility-based approach for occupancy grid computation in disparity space,” *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1383–1393, Sep. 2012.
[87] J. Wang, G. Bebis, and R. Miller, “Overtaking vehicle detection using dynamic and quasi-static background modeling,” in *Proc. IEEE CVPR*, Jun. 2005, p. 64.
[88] 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 Trans. Intell. Transp. Syst.*, vol. 10, no. 1, pp. 70–82, Mar. 2009.
[89] J. Arrospide, L. Salgado, and M. Nieto, “Vehicle detection and tracking using homography-based plane rectification and particle filtering,” in *Proc. IEEE IV*, Jun. 2010, pp. 150–155.
[90] B. D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” in *Proc. Imag. Understand. Workshop*, 1981, pp. 674–679.
[91] E. Martinez, M. Diaz, J. Melenchon, J. Montero, I. Iriondo, and J. Socoro, “Driving assistance system based on the detection of head-on collisions,” in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2008, pp. 913–918.
[92] D. Baehring, S. Simon, W. Niehsen, and C. Stiller, “Detection of close cut-in and overtaking vehicles for driver assistance based on planar parallax,” in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2005, pp. 290–295.
[93] K. Yamaguchi, T. Kato, and Y. Ninomiya, “Vehicle ego-motion estimation and moving object detection using a monocular camera,” in *Proc. 18th ICPR*, vol. 4, pp. 610–613.
[94] K. Yamaguchi, T. Kato, and Y. Ninomiya, “Moving obstacle detection using monocular vision,” in *Proc. IEEE Intell. Veh. Symp.*, 2006, pp. 288–293.
[95] I. Sato, C. Yamano, and H. Yanagawa, “Crossing obstacle detection with a vehicle-mounted camera,” in *Proc. IEEE IV*, Jun. 2011, pp. 60–65.
[96] A. Geiger and B. Kitt, “Object flow: A descriptor for classifying traffic motion,” in *Proc. IEEE Intell. Veh. Symp.*, San Diego, USA, Jun. 2010, pp. 287–293.
[97] A. Geiger, “Monocular road mosaicing for urban environments,” in *Proc. IEEE IV*, Xi’an, China, Jun. 2009, pp. 140–145.
[98] J. Velten, S. Schauland, A. Gavriilidis, T. Schwerdtfeger, F. Boschen, and A. Kummert, “Tomographical scene reconstruction in the active safety car project,” in *Proc. 15th Int. IEEE Conf. ITSC*, 2012, pp. 307–312.
[99] I. Cabani, G. Toulminet, and A. Bensrhair, “Color-based detection of vehicle lights,” in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2005, pp. 278–283.
[100] Y.-L. Chen, C.-T. Lin, C.-J. Fan, C.-M. Hsieh, and B.-F. Wu, “Vision-based nighttime vehicle detection and range estimation for driver assistance,” in *Proc. IEEE Int. Conf. SMC*, Oct. 2008, pp. 2988–2993.
[101] S. Gormer, D. Muller, S. Hold, M. Meuter, and A. Kummert, “Vehicle recognition and TTC estimation at night based on spotlight pairing,” in *Proc. 12th Int. IEEE Conf. ITSC*, Oct. 2009, pp. 1–6.
[102] A. Fossati, P. Schönmann, and P. Fua, “Real-time vehicle tracking for driving assistance,” *Mach. Vis. Appl.*, vol. 22, no. 2, pp. 439–448, Mar. 2010.
[103] R. O’Malley, E. Jones, and M. Glavin, “Rear-lamp vehicle detection and tracking in low-exposure color video for night conditions,” *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 2, pp. 453–462, Jun. 2010.
[104] J. C. Rubio, J. Serrat, A. M. Lopez, and D. Ponsa, “Multiple-target tracking for intelligent headlights control,” *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 594–605, Jun. 2012.
[105] X. Zhang and N. Zheng, “Vehicle detection under varying poses using conditional random fields,” in *Proc. 13th Int. IEEE Conf. ITSC*, Sep. 2010, pp. 875–880.
[106] P. Rybski, D. Huber, D. Morris, and R. Hoffman, “Visual classification of coarse vehicle orientation using histogram of oriented gradients features,” in *Proc. IEEE IV Symp.*, Jun. 2010, pp. 921–928.
[107] M. Grinberg, F. Ohr, and J. Beyerer, “Feature-based probabilistic data association (FBPDA) for visual multi-target detection and tracking under occlusions and split and merge effects,” in *Proc. 12th IEEE Int. Conf. ITSC*, Oct. 2009, pp. 1–8.
[108] J. I. Woodlill, R. Buck, D. Jurasek, G. Gordon, and T. Brown, “3D vision: Developing an embedded stereo-vision system,” *Computer*, vol. 40, no. 5, pp. 106–108, May 2007.
[109] H. Hirschmuller, “Accurate and efficient stereo processing by semi-global matching and mutual information,” in *Proc. IEEE Comput. Vis. Pattern Recog.*, 2005, pp. 807–814.
[110] A. Geiger, M. Roser, and R. Urtasun, “Efficient large-scale stereo matching,” in *Proc. ACCV*, Queenstown, New Zealand, Nov. 2010, pp. 25–38.
[111] I. Haller, C. Pantilie, F. Oniga, and S. Nedevschi, “Real-time semi-global dense stereo solution with improved sub-pixel accuracy,” in *Proc. IEEE IV Symp.*, Jun. 2010, pp. 369–376.
[112] F. Oniga and S. Nedevschi, “Processing dense stereo data using elevation maps: Road surface, traffic isle, and obstacle detection,” *IEEE Trans. Veh. Technol.*, vol. 59, no. 3, pp. 1172–1182, Mar. 2010.
[113] R. Labayrade, D. Aubert, and J.-P. Tarel, “Real time obstacle detection on non flat road geometry through v-disparity representation,” in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2002, pp. 646–651.
[114] A. Vatavu, R. Danescu, and S. Nedevschi, “Real-time dynamic environment perception in driving scenarios using difference fronts,” in *Proc. IEEE IV Symp.*, Jun. 2012, pp. 717–722.
[115] R. O. Duda and P. E. Hart, “Use of the Hough transformation to detect lines and curves in pictures,” *Commun. ACM*, vol. 15, no. 1, pp. 11–15, Jan. 1972.
[116] M. A. Fischler and R. C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.
[117] H. Lategahn, T. Graf, C. Hasberg, B. Kitt, and J. Effertz, “Mapping in dynamic environments using stereo vision,” in *Proc. IEEE IV Symp.*, Jun. 2011, pp. 150–156.
[118] Y.-C. Lim, C.-H. Lee, S. Kwon, and J. Kim, “Event-driven track management method for robust multi-vehicle tracking,” in *Proc. IEEE IV Symp.*, Jun. 2011, pp. 189–194.
[119] A. Broggi, C. Caraffi, R. Fedriga, and P. Grisleri, “Obstacle detection with stereo vision for off-road vehicle navigation,” in *Proc. IEEE Comput. Soc. Conf. CVPR*, Jun. 2005, p. 65.
[120] A. Broggi, A. Cappalunga, C. Caraffi, S. Cattani, S. Ghidoni, P. Grisleri, P.-P. Porta, M. Posterli, P. Zani, and J. Beck, “The passive sensing suite of the TerraMax autonomous vehicle,” in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2008, pp. 769–774.
[121] A. Broggi, C. Caraffi, P. Porta, and P. Zani, “The single frame stereo vision system for reliable obstacle detection used during the 2005 DARPA Grand Challenge on TerraMax,” in *Proc. IEEE ITSC*, Sep. 2006, pp. 745–752.
[122] N. Ben Romdhane, M. Hammami, and H. Ben-Abdallah, “A generic obstacle detection method for collision avoidance,” in *Proc. IEEE IV Symp.*, Jun. 2011, pp. 491–496.
[123] M. Perrollaz, A. Spalanzani, and D. Aubert, “Probabilistic representation of the uncertainty of stereo-vision and application to obstacle detection,” in *Proc. IEEE IV Symp.*, Jun. 2010, pp. 313–318.
[124] S. Kubota, T. Nakano, and Y. Okamoto, “A global optimization algorithm for real-time on-board stereo obstacle detection systems,” in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2007, pp. 7–12.
[125] A. Broggi, S. Cattani, E. Cardarelli, B. Kriel, M. McDaniel, and H. Chang, “Disparity space image’s features analysis for error prediction of a stereo obstacle detector for heavy duty vehicles,” in *Proc. 14th Int. IEEE Conf. ITSC*, Oct. 2011, pp. 80–86.
[126] T. Kowsari, S. Beauchemin, and J. Cho, “Real-time vehicle detection and tracking using stereo vision and multi-view AdaBoost,” in *Proc. 14th Int. IEEE Conf. ITSC*, Oct. 2011, pp. 1255–1260.

[127] B. Kitt, B. Ranft, and H. Lategahn, "Detection and tracking of independently moving objects in urban environments," in *Proc. 13th Int. IEEE Conf. ITSC*, Sep. 2010, pp. 1396–1401.

[128] A. Bak, S. Bouchafa, and D. Aubert, "Detection of independently moving objects through stereo vision and ego-motion extraction," in *Proc. IEEE IV Symp.*, Jun. 2010, pp. 863–870.

[129] W. van der Mark, J. van den Heuvel, and F. Groen, "Stereo based obstacle detection with uncertainty in rough terrain," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2007, pp. 1005–1012.

[130] B. Barrois, S. Hristova, C. Wohler, F. Kummert, and C. Hermes, "3D pose estimation of vehicles using a stereo camera," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2009, pp. 267–272.

[131] C. Hermes, J. Einhaus, M. Hahn, C. Wöhler, and F. Kummert, "Vehicle tracking and motion prediction in complex urban scenarios," in *Proc. IEEE IV Symp.*, Jun. 2010, pp. 26–33.

[132] S. Lefebvre and S. Ambellouis, "Vehicle detection and tracking using mean shift segmentation on semi-dense disparity maps," in *Proc. IEEE IV Symp.*, Jun. 2012, pp. 855–860.

[133] C. Rabe, U. Franke, and S. Gehrig, "Fast detection of moving objects in complex scenarios," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2007, pp. 398–403.

[134] P. Lenz, J. Ziegler, A. Geiger, and M. Roser, "Sparse scene flow segmentation for moving object detection in urban environments," in *Proc. IEEE Intell. Veh. Symp.*, Baden-Baden, Germany, Jun. 2011, pp. 926–932.

[135] Y.-C. Lim, C.-H. Lee, S. Kwon, and J.-H. Lee, "A fusion method of data association and virtual detection for minimizing track loss and false track," in *Proc. IEEE IV Symp.*, Jun. 2010, pp. 301–306.

[136] J. Morat, F. Devernay, and S. Cornou, "Tracking with stereo-vision system for low speed following applications," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2007, pp. 955–961.

[137] S. Bota and S. Nedevschi, "Tracking multiple objects in urban traffic environments using dense stereo and optical flow," in *Proc. 14th Int. IEEE Conf. ITSC*, Oct. 2011, pp. 791–796.

[138] G. Catalin and S. Nedevschi, "Object tracking from stereo sequences using particle filter," in *Proc. 4th Int. Conf. ICCP*, Aug. 2008, pp. 279–282.

[139] A. Vatavu and S. Nedevschi, "Real-time modeling of dynamic environments in traffic scenarios using a stereo-vision system," in *Proc. 15th Int. IEEE Conf. ITSC*, 2012, pp. 722–727.

[140] C. Pantilie and S. Nedevschi, "Real-time obstacle detection in complex scenarios using dense stereo vision and optical flow," in *Proc. 13th Int. Conf. ITSC*, Sep. 2010, pp. 439–444.

[141] K. Y. Lee, J. W. Lee, and M. R. Cho, "Detection of road obstacles using dynamic programming for remapped stereo images to a top-view," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2005, pp. 765–770.

[142] M. Perrollaz, J. Yoder, and C. Laugier, "Using obstacles and road pixels in the disparity-space computation of stereo-vision based occupancy grids," in *Proc. 13th Int. IEEE Conf. ITSC*, Sep. 2010, pp. 1147–1152.

[143] Y. Zhu, D. Comaniciu, V. Ramesh, M. Pellkofer, and T. Koehler, "An integrated framework of vision-based vehicle detection with knowledge fusion," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2005, pp. 199–204.

[144] X. Mei and H. Ling, "Robust visual tracking and vehicle classification via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2259–2272, Nov. 2011.

[145] A. Yilmaz, O. Javed, and M. Shah (2006, Dec.). Object tracking: A survey. *ACM Comput. Surv.* [Online]. *38(4)*, p. 13. Available: http://doi.acm.org/10.1145/1177352.1177355

[146] C. Hilario, J. Collado, J. Armingol, and A. de la Escalera, "Visual perception and tracking of vehicles for driver assistance systems," in *Proc. IEEE Intell. Veh. Symp.*, 2006, pp. 94–99.

[147] S. Moqqaddem, Y. Ruichek, R. Touahni, and A. Sbihi, "A spectral clustering and Kalman filtering based objects detection and tracking using stereo vision with linear cameras," in *Proc. IEEE IV Symp.*, Jun. 2011, pp. 902–907.

[148] A. Barth and U. Franke, "Where will the oncoming vehicle be the next second?" in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2008, pp. 1068–1073.

[149] G. Stein, Y. Gdalyahu, and A. Shashua, "Stereo-assist: Top-down stereo for driver assistance systems," in *Proc. IEEE IV Symp.*, Jun. 2010, pp. 723–730.

[150] G. Toulminet, M. Bertozzi, S. Mousset, A. Bensrhair, and A. Broggi, "Vehicle detection by means of stereo vision-based obstacles features extraction and monocular pattern analysis," *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2364–2375, Aug. 2006.

[151] F. Rattei, P. Kindt, A. Probstl, and S. Chakraborty, "Shadow-based vehicle model refinement and tracking in advanced automotive driver assistance systems," in *Proc. 9th IEEE Symp. ESTIMedia*, Oct. 2011, pp. 46–55.

[152] B. Alefs, "Embedded vehicle detection by boosting," in *Proc. IEEE ITSC*, Sep. 2006, pp. 536–541.

[153] G. Gritsch, N. Donath, B. Kohn, and M. Litzenberger, "Night-time vehicle classification with an embedded, vision system," in *Proc. 12th IEEE Int. Conf. ITSC*, Oct. 2009, pp. 1–6.

[154] J. Kaszubiak, M. Tornow, R. Kuhn, B. Michaelis, and C. Knoeppel, "Real-time vehicle and lane detection with embedded hardware," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2005, pp. 619–624.

[155] G. Stein, E. Rushinek, G. Hayun, and A. Shashua, "A computer vision system on a chip: A case study from the automotive domain," in *Proc. IEEE Comput. Soc. Conf. CVPR*, Jun. 2005, p. 130.

[156] S. Toral, F. Barrero, and M. Vargas, "Development of an embedded vision based vehicle detection system using an ARM video processor," in *Proc. 11th Int. IEEE Conf. ITSC*, Oct. 2008, pp. 292–297.

[157] C. Banz, H. Blume, and P. Pirsch, "Real-time semi-global matching disparity estimation on the GPU," in *Proc. IEEE ICCV Workshops*, Nov. 2011, pp. 514–521.

[158] F. Homm, N. Kaempchen, J. Ota, and D. Burschka, "Efficient occupancy grid computation on the GPU with lidar and radar for road boundary detection," in *Proc. IEEE IV Symp.*, Jun. 2010, pp. 1006–1013.

[159] X. Liu, Z. Sun, and H. He, "On-road vehicle detection fusing radar and vision," in *Proc. IEEE ICVES*, Jul. 2011, pp. 150–154.

[160] R. Chavez-Garcia, J. Burlet, T.-D. Vu, and O. Aycard, "Frontal object perception using radar and mono-vision," in *Proc. IEEE IV Symp.*, Jun. 2012, pp. 159–164.

[161] M. Nishigaki, S. Rebhan, and N. Einecke, "Vision-based lateral position improvement of radar detections," in *Proc. IEEE ITSC*, 2012, pp. 90–97.

[162] R. Schubert, G. Wanielik, and K. Schulze, "An analysis of synergy effects in an omnidirectional modular perception system," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2009, pp. 54–59.

[163] E. Richter, R. Schubert, and G. Wanielik, "Radar and vision based data fusion—Advanced filtering techniques for a multi object vehicle tracking system," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2008, pp. 120–125.

[164] 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 *Proc. IEEE IV Symp.*, Jun. 2012, pp. 494–499.

[165] G. Alessandretti, A. Broggi, and P. Cerri, "Vehicle and guard rail detection using radar and vision data fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 1, pp. 95–105, Mar. 2007.

[166] M. Bertozzi, L. Bombini, P. Cerri, P. Medici, P. Antonello, and M. Miglietta, "Obstacle detection and classification fusing radar and vision," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2008, pp. 608–613.

[167] U. Kadow, G. Schneider, and A. Vukotich, "Radar-vision based vehicle recognition with evolutionary optimized and boosted features," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2007, pp. 749–754.

[168] J. Fritsch, T. Michalke, A. Gepperth, S. Bone, F. Waibel, M. Kleinehagenbrock, J. Gayko, and C. Goerick, "Towards a human-like vision system for driver assistance," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2008, pp. 275–282.

[169] F. Liu, J. Sparbert, and C. Stiller, "IMMPDA vehicle tracking system using asynchronous sensor fusion of radar and vision," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2008, pp. 168–173.

[170] B. Alefs, D. Schreiber, and M. Clabian, "Hypothesis based vehicle detection for increased simplicity in multi-sensor ACC," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2005, pp. 261–266.

[171] Y. Tan, F. Han, and F. Ibrahim, "A radar guided vision system for vehicle validation and vehicle motion characterization," in *Proc. IEEE ITSC*, Oct. 3–30, 2007, pp. 1059–1066.

[172] Z. Ji, M. Luciw, J. Weng, and S. Zeng, "Incremental online object learning in a vehicular radar-vision fusion framework," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 402–411, Jun. 2011.

[173] T. Shen, G. Schamp, T. Cooprider, and F. Ibrahim, "Stereo vision based full-range object detection and tracking," in *Proc. 14th Int. IEEE Conf. ITSC*, Oct. 2011, pp. 925–930.

[174] S. Wu, S. Decker, P. Chang, T. Camus, and J. Eledath, "Collision sensing by stereo vision and radar sensor fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 4, pp. 606–614, Dec. 2009.

[175] L. Huang and M. Barth, "Tightly-coupled lidar and computer vision integration for vehicle detection," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2009, pp. 604–609.

[176] C. Premebida, G. Monteiro, U. Nunes, and P. Peixoto, "A lidar and vision-based approach for pedestrian and vehicle detection and tracking," in *Proc. IEEE ITSC*, Oct. 3–30, 2007, pp. 1044–1049.

[177] S. Matzka, A. Wallace, and Y. Petillot, "Efficient resource allocation for attentive automotive vision systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 859–872, Jun. 2012.

[178] S. Rodriguez, F. V. Freandmont, P. Bonnifait, and V. Cherfaoui, "Visual confirmation of mobile objects tracked by a multi-layer lidar," in *Proc. 13th Int. IEEE Conf. ITSC*, Sep. 2010, pp. 849–854.
[179] Q. Baig, O. Aycard, T. D. Vu, and T. Fraichard, "Fusion between laser and stereo vision data for moving objects tracking in intersection like scenario," in *Proc. IEEE IV Symp.*, Jun. 2011, pp. 362–367.
[180] O. Aycard, Q. Baig, S. Bota, F. Nashashibi, S. Nedevschi, C. Pantilie, M. Parent, P. Resende, and T.-D. Vu, "Intersection safety using lidar and stereo vision sensors," in *Proc. IEEE IV Symp.*, Jun. 2011, pp. 863–869.
[181] M. Haberjahn and M. Junghans, "Vehicle environment detection by a combined low and mid level fusion of a laser scanner and stereo vision," in *Proc. 14th Int. IEEE Conf. ITSC*, Oct. 2011, pp. 1634–1639.
[182] W. Yao, H. Zhao, F. Davoine, and H. Zha, "Learning lane change trajectories from on-road driving data," in *Proc. IEEE IV Symp.*, Jun. 2012, pp. 885–890.
[183] A. Doshi and M. Trivedi, "On the roles of eye gaze and head dynamics in predicting driver's intent to change lanes," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 3, pp. 453–462, Sep. 2009.
[184] C. Hermes, C. Wohler, K. Schenk, and F. Kummert, "Long-term vehicle motion prediction," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2009, pp. 652–657.
[185] J. Wiest, M. Hoffken, U. Kresel, and K. Dietmayer, "Probabilistic trajectory prediction with Gaussian mixture models," in *Proc. IEEE IV Symp.*, Jun. 2012, pp. 141–146.
[186] T. Gindele, S. Brechtel, and R. Dillmann, "A probabilistic model for estimating driver behaviors and vehicle trajectories in traffic environments," in *Proc. 13th Int. IEEE Conf. ITSC*, Sep. 2010, pp. 1625–1631.
[187] E. Kafer, C. Hermes, C. Woandhler, F. Kummert, and H. Ritter, "Recognition and prediction of situations in urban traffic scenarios," in *Proc. 20th ICPR*, Aug. 2010, pp. 4234–4237.
[188] H.-T. Lin, C.-J. Lin, and R. C. Weng, "A note on Platt's probabilistic outputs for support vector machines," *Mach. Learn.*, vol. 68, no. 3, pp. 267–276, Oct. 2007.
[189] A. Joshi, F. Porikli, and N. Papanikolopoulos, "Scalable active learning for multiclass image classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2259–2273, Nov. 2012.
[190] A. Westenberger, B. Duraisamy, M. Munz, M. Muntzinger, M. Fritzsche, and K. Dietmayer, "Impact of out-of-sequence measurements on the joint integrated probabilistic data association filter for vehicle safety systems," in *Proc. IEEE IV Symp.*, Jun. 2012, pp. 438–443.
[191] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc.*, vol. 39, no. 1, pp. 1–38, 1977.
[192] A. Doshi, B. Morris, and M. Trivedi, "On-road prediction of driver's intent with multimodal sensory cues," *IEEE Pervasive Comput.*, vol. 10, no. 3, pp. 22–34, Jul.–Sep. 2011.
[193] S. Calderara, A. Prati, and R. Cucchiara, "Mixtures of von Mises distributions for people trajectory shape analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 4, pp. 457–471, Apr. 2011.
[194] Caltech computational vision Caltech cars 1999. [Online]. Available: http://www.vision.caltech.edu/html-files/archive.html
[195] Caltech computational vision Caltech cars 2001. [Online]. Available: http://www.vision.caltech.edu/html-files/archive.html
[196] R. Kasturi, D. Goldgof, P. Soundararajan, V. Manohar, J. Garofolo, R. Bowers, M. Boonstra, V. Korzhova, and J. Zhang, "Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 319–336, Feb. 2009.
[197] Performance evaluation of tracking and surveillance, PETS 2001. [Online]. Available: http://www.cvg.cs.rdg.ac.uk/PETS2001/pets2001-dataset.html
[198] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving?" in *Proc. CVPR*, Providence, USA, Jun. 2012, pp. 3354–3361.

Sayanan Sivaraman (M'07) received the B.S. degree in electrical engineering from the University of Maryland, College Park, MD, USA, in 2007 and the M.S. degree in electrical engineering from the University of California, San Diego, La Jolla, CA, USA, in 2009. He is currently working toward the Ph.D. degree in intelligent systems, robotics, and controls in the Computer Vision and Robotics Research Laboratory, University of California, San Diego.

His research interests include computer vision, machine learning, intelligent vehicles, and intelligent transportation systems.

Mr. Sivaraman serves as a Reviewer for the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEM, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and *Machine Vision and Applications*. He also served on the Organizing Committee of the IEEE Intelligent Vehicles Symposium 2010.

Mohan Manubhai Trivedi (F'09) received the B.E. (Hons.) degree from Birla Institute of Technology and Science, Pilani, India, and the Ph.D. degree from Utah State University, Logan, UT, USA.

He is a Professor of electrical and computer engineering and the Founding Director of the Computer Vision and Robotics Research Laboratory at the University of California, San Diego, La Jolla, CA, USA. He and his team are currently working toward research in machine and human perception, machine learning, distributed video systems, multimodal effect and gesture analysis, human-centered interfaces, and intelligent driver assistance systems. He regularly serves as a Consultant to industry and government agencies in the U.S. and abroad. He has given over 65 keynote/plenary talks.

Prof. Trivedi is a Fellow of the IEEE, The International Association for Pattern Recognition, and The International Society for Optical Engineers. He has served as the Editor-in-Chief of the *Machine Vision and Applications* journal and is currently an Editor of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS and *Image and Vision Computing*. He served as the General Chair of the IEEE Intelligent Vehicles Symposium 2010. His team designed and deployed the Eagle Eyes system on the U.S.–Mexico border in 2006. He served as a Charter Member and the Vice Chair of the Executive Committee of the University of California (UC) system-wide UC Discovery Program and as an Expert Panelist for the Strategic Highway Research Program of the Transportation Research Board of the National Academies. He has been a recipient of the Distinguished Alumnus Award from Utah State University, the Pioneer Award and the Meritorious Service Award from the IEEE Computer Society, and several Best Paper awards.