State Estimation and Motion Prediction of Vehicles and Vulnerable Road Users for Cooperative Autonomous Driving: A Survey

Prasenjit Ghorai, Azim Eskandarian, *Senior Member, IEEE*, Young-Keun Kim, and Goodarz Mehr

***Abstract*—The recent progress in autonomous vehicle research and development has led to increasingly widespread testing of fully autonomous vehicles on public roads, where complex traffic scenarios arise. Along with these vehicles, partially autonomous vehicles, manually-driven vehicles, pedestrians, cyclists, and some animals can be present on the road, to which autonomous vehicles must react. This study focuses on a comprehensive survey of the literature on motion prediction and state estimation of vehicles and VRUs, which are essential for path planning and navigation functionalities of an autonomous vehicle. Motion prediction and state estimation methods utilize the vehicle's own sensory perception capabilities and information obtained through cooperative perception from V2V and V2X connections. This survey summarizes the significant progress that has been made in both categories, discusses the most promising results to date and outlines critical research challenges that need to be overcome to achieve full autonomy, from an ego vehicle's perspective in mixed traffic environments.**

***Index Terms*—Cooperative autonomous driving, motion prediction, perception, state estimation, vulnerable road users.**

List of Acronyms

ACC	Adaptive Cruise Control
AD	Autonomous Driving
ADAS	Advanced Driver Assistant Systems
ADS	Automated Driving System
AP	Average Precision
AV	Autonomous Vehicle
BEV	Bird's Eye View
CACC	Cooperative Adaptive Cruise Control
CAS	Collision Avoidance System
CAVs	Connected Autonomous Vehicles
CCAD	Connected and Cooperative Autonomous Driving
CNN	Convolutional Neural Network
CP	Cooperative Perception
CPN	Cooperative Perception and Navigation
DL	Deep Learning
DSRC	Dedicated Short-Range Communication
FoV	Field of View
HMM	Hidden Markov Model
HOG	Histogram of Oriented Gradients
LoS	Line of Sight
ML	Machine Learning
MLP	Multi-layer Perceptron
NHTSA	National Highway Traffic Safety Administration
R-CNN	Regions with Convolutional Neural Network
RoI	Regions of Interest
SLAM	Simultaneous Localization and Mapping
SSD	Single Shot Detector
SVM	Support Vector Machine
SVM-BF	Support Vector Machines-Bayesian Filtering
V2I	Vehicle to Infrastructure
V2V	Vehicle to Vehicle
V2X	Vehicle to Everything
VEC	Vehicular Edge Computing
VRUs	Vulnerable Road Users
YOLO	You Only Look Once

Manuscript received February 13, 2021; revised September 24, 2021; accepted March 4, 2022. This work was supported in part by the Department of Mechanical Engineering, Virginia Tech; and in part by the Safe-D Project (Safety through Disruption and Safe-D National UTC) under Grant 69A3551747115. The Associate Editor for this article was Y. Chen. *(Corresponding author: Prasenjit Ghorai.)*

Prasenjit Ghorai, Azim Eskandarian, and Goodarz Mehr are with the Autonomous Systems and Intelligent Machines (ASIM) Laboratory, Department of Mechanical Engineering, Virginia Tech, Blacksburg, VA 24061 USA (e-mail: prasenjitg@vt.edu; eskandarian@vt.edu; goodarzm@vt.edu).

Young-Keun Kim is with the School of Mechanical and Control Engineering, Handong Global University, Pohang 37554, Republic of Korea (e-mail: ykkim@handong.edu).

Digital Object Identifier 10.1109/TITS.2022.3160932

I. Introduction

RECENT breakthroughs in deep learning-based computer vision have advanced autonomous driving technology to the next level. The worldwide research and development carried out by academia and vehicle manufacturers has significantly expanded the knowledge base for autonomous driving, reducing the time horizon to deploy fully autonomous vehicles on public roads. This progress, however, has brought forth new problems and challenges arising from the operation of autonomous vehicles in dynamic and heterogeneous traffic scenarios where manually-driven vehicles and partially-automated vehicles will be on the road along with pedestrians, bicyclists, and other VRUs.

A critical challenge facing fully autonomous vehicles is an improper or inaccurate response to the surrounding environment in a driving scenario that may endanger other vehicles or VRUs. This can be because the vehicle has not encountered that specific scenario before, because of detection or classification failure, because of sensor FoV blockage or failure, or because of extreme weather conditions. Take, for example,

TABLE I
A SUMMARY OF ADAS AND AV SURVEY PAPERS

Author(s)	**Year**	**Survey Topic(s)**
Sivaraman and Trivedi [10]	2013	Vision-based on-road vehicle detection, tracking, and behavior analysis
Lefèvre et al. [11]	2014	Motion prediction and risk assessment for intelligent vehicles
Mukhtar et al. [12]	2015	On-road vision-based vehicle detection and tracking systems for CAS
Paden et al. [13]	2016	Planning and control algorithms for self-driving vehicles in urban traffic/scenario
Gonzalez et al. [14]		Motion planning techniques implemented in intelligent transportation literature
Abboud et al. [15]		DSRC and cellular solutions for V2X communications already adopted and deployed in vehicles by car manufacturers
Bresson et al. [16]	2017	Localization, mapping, and SLAM
Pendleton et al. [22]		Perception, planning, control, and coordination for AVs
Zhu et al. [18]		Environment perception: lane and road detection, traffic sign recognition, vehicle tracking, behavior analysis, and scene understanding
Kuutti et al. [23]	2018	Ego vehicle localization techniques using on-board sensors and information obtained from V2X communication channels and their applicability to AVs
Van Brummelen et al. [24]		Perception, localization, and mapping methods currently implemented in AV research
Schwarting et al. [19]		Integrated perception for behavior-aware planning
Bighashdel and Dubbelman [25]	2019	Path prediction techniques/approaches for VRUs
Montanaro et al. [26]		Connected autonomous driving
Eskandarian et al. [27]		Methods and algorithms for sensing, perception, planning, and control of CAVs
Badue et al. [28]	2020	Architectural autonomy of AVs for perception and decision making
Yurtsever et al. [17]		Localization, mapping, perception, planning, and human-machine interfaces
Feng et al. [29]		Deep multi-modal object detection with semantic segmentation
Yu and Marinov [30]		Obstacle detection in extreme weather and in urban areas
Rasouli and Tsotsos [20]		Pedestrian behavior studies and interaction problems with AVs
Rudenko et al. [21]		Human motion trajectory prediction
Wu et al. [31]		Intrusion detection for in-vehicle networks
Hu et al. [32]		Multi-sensor fusion-based obstacle detection for intelligent ground vehicles in off-road environments
Hu et al. [33]		Research on traffic conflicts based on intelligent vehicles
Pilz et al. [34]	2021	Components of CP
Yeong et al. [35]		Sensor and sensor fusion technology in AVs

the Tesla on Autopilot's crash in California [1], where the car's sensors could not recognize a parked fire truck on the side of the raod. In another crash involving Tesla and Autopilot in Florida [2], the vehicle could not discern a white crossing truck against the bright sky background. There are still many other instances of circumstances leading up to bad decisions, such as the Uber incident [3]. In that case, the vehicle did detect an unknown object, a pedestrian walking with a bicycle, from a distance. As the vehicle approached the unknown object, it first classified that object as a vehicle and later as a bicycle, but it was too late by then. While these incidents highlight the challenges facing autonomous vehicles and the importance of perception failure mitigation, we should not gloss over the incredible progress that has been made in autonomous vehicle research, nor the benefits of having fully autonomous vehicles on the road, given that according to NHTSA, 94% of road accidents are caused by human error [4].

An enormous amount of research work has been carried out to introduce and implement ADAS [5] such as CAS, lane-keeping [6], ACC [7], and CACC [8], [9] to counteract a signal loss, reduce human error and improve vehicle safety. The same is true for the methods and algorithms enabling vehicle autonomy in areas ranging from perception to motion planning and control. Overall, the progress made toward intelligent transportation systems over the past several years has be reviewed by researchers in different areas, with important surveys listed in Table I each highlighting a core area of research and the advances made in that area; namely, on-road vehicle detection [10], motion prediction and risk assessment [11], vehicle detection techniques for collision avoidance [12], motion planning [13], [14] and control techniques [13], DSRC and cellular solutions for V2X communication for intelligent vehicles [15], localization and mapping [16], [17], environment perception and traffic sign detection [18], perception for behavior-aware planning [19], pedestrian behavior [20] and it's motion trajectory prediction [21], etc.

Most of these works only cover a few aspects of connected autonomous driving, which is reflective of the current approach to autonomy that has focused on building small and disparate intelligences that are closed off to the rest of the world. In the current approach, even if several autonomous vehicles are traveling in the same environment at the same time, they each have to carry expensive sensing, navigation, and processing hardware and still, lacking coordination with other road users, they may get into accidents. A future with a mixed traffic of CAVs and other vehicles on the road requires a paradigm shift in communications and coordination, cooperative sensing, and real-time dynamic planning and controls to be effective

TABLE II
A SUMMARY OF THE DISCUSSED EXTEROCEPTIVE SENSORS USED IN AVS

SM	IA	WA	C	A	Benefits	Drawbacks
Standard Camera	Yes	Yes	Lowest	High	Good lane and obstacle detection, object classification, 3D mapping (using a stereo camera), long-range detection, depth information can be extracted	Image processing may become computationally expensive, distance and velocity measurements are not easy, performance degrades in extreme weather conditions, sensitive to scene lighting (except thermal Camera)
Stereo camera			Low			
Thermal Camera	No		Low			
Lidar	Yes		High	High	Direct distance measurement and obstacle detection, large FoV, robust 3D mapping, intensity measurement can lead to lane detection	Poor object classification and indirect velocity estimation, difficulty detecting an object with high reflectivity or in bad weather conditions, difficulty in short-distance measurements
Radar	No	No	Low	Medium	Direct distance and velocity measurement, can operate in extreme weather conditions	Poor object classification, poor performance in short distance measurement and pedestrian and static object detection, susceptible to interference

SM: Sensor Modality; IA: Illumination Affects; WA: Weather Affects; A: Accuracy; C: Cost;

at improving traffic congestion, road user safety, and overall efficiency. This future can be imagined as a multi-lane highway or a city block with a mix of autonomous and manually-driven cars which are communication-enabled, each having a navigation plan, and a generated trajectory and a maneuver of some sort to meet that plan. The autonomous ones have situational awareness by virtue of their sensors, and this awareness can be shared with the surrounding road users within a region or area. This will ultimately improve traffic congestion, minimize driver load, increase the effective usage of on-road vehicles, and improve fuel efficiency. Observing this untapped potential, researchers are moving towards connected and cooperative intelligent transportation systems by merging established and developing technologies from diverse areas. Therefore, the prime objective of this comprehensive study is to connect all the relevant research areas, summarize the existing developments, and highlight the challenges in each area so that a bird's-eye view is available to the new researchers in this field.

This survey begins with a discussion of exteroceptive sensor types used in AVs and a comparison of their range, accuracy, cost, weather performance, and a discussion of each sensor type's drawbacks. We then review DL-based 2D and 3D dynamic object detection methods used in AV research, with a focus on the applications and limitations of these methods. Next, we discuss and categorize different approaches for detection, motion prediction, intent estimation, and behavior analysis of other vehicles and pedestrians from a practical point of view, along with a summary of existing data sets for training, validation, and testing of these methods. We will also highlight open challenges in mixed driving traffic scenarios for future research. Considering the critical nature of perception failure mitigation, in this survey we focus on detection and tracking, state and intent estimation, and motion prediction of dynamic agents and objects an autonomous ego vehicle encounters. In this survey, dynamic agents include pedestrians and other vehicles - primarily passenger cars. The recent emergence of cooperative perception and navigation plays an important role in the development of CAVs, which should ultimately help them take appropriate actions in heterogeneous traffic scenarios. Therefore, we provide a summary of major developments in cooperative perception and navigation and present an overall analysis of current implementations and their limitations. As CCAD [36] seems a promising approach for the widespread adoption of vehicle autonomy, we think this survey will be beneficial to researchers who are working in or entering this area.

The remainder of this paper is organized as follows. Section II highlights major developments of exteroceptive perception sensors used in AVs, sensor fusion, egocentric dynamic object detection methods using DL and machine intelligence, their limitations, and open challenges. Section III summarizes the state-of-the-art classical methods of state estimation and motion prediction of pedestrians and vehicles. Section IV discusses the progress of cooperative perception for autonomous driving and a detailed analysis of existing implementation issues in AVs. Future research directions are discussed in Section V, and finally, Section VI concludes our review of the literature.

II. EGO VEHICLE PERCEPTION OF ON-ROAD OBJECTS

An AV's level of intelligence depends on its sensors and the sophistication of the algorithms that interpret information from those sensors. This section first reviews perception sensors commonly used in CAVs, then discusses various object detection methods, and finally highlights the existing challenges of ego-centric object detection.

A. Perception Sensors

Based on their application, AV sensors can be divided into onboard exteroceptive and proprioceptive or interoceptive sensors. The primary task of exteroceptive sensors is the perception of static and dynamic objects in the surrounding environment and prediction of their motion and behavior. This subsection focuses on exteroceptive perception sensors, particularly camera, lidar, and radar, and discusses their purpose, major advantages and disadvantages, cost-effectiveness, level of uncertainty, and suitability for different weather conditions. A comparative summary of our discussion on these sensors is available in Table II.

TABLE III
ONBOARD SENSOR COMBINATIONS FOR SOME AV PLATFORMS [17]

Platform	360° rotating lidar (No.)	Stationary lidar (No.)	Radar (No.)	Camera (No.)
Yurtsever et al. 2020 [17]	1	-	-	4
Boss [49]	1	9	5	2
Junior [48]	1	2	6	4
BRAiVE [56]	-	5	1	10
RobotCar [50]	-	3	-	4
Google car (Pirus) [57]	1	-	4	1
Uber car (XC 90) [52]	1	-	10	7
Uber car (Fusion) [52]	1	7	7	20
Bertha [54]	-	-	6	3
Apollo Auto [53]	1	3	2	2

1) Camera: cameras are passive sensors in the sense that they do not interfere with other systems or sensors by affecting the environment. They can distinguish color, which is critical to AVs for recognizing traffic lights and signs, lane markings, other vehicles, and pedestrians on the road. A recent survey [17] has highlighted the state-of-the-art computer vision algorithms utilizing monocular, omnidirectional, and event cameras, comparing their advantages and limitations. Though event and thermal cameras have drawn some interest for ADS, they still suffer from problems arising from scene illumination and weather conditions. Another survey paper [27] has detailed computer vision-based algorithms for object and traffic sign detection. Additional details regarding the performance of standard, stereo, and thermal cameras are highlighted below.

a) Standard camera: standard cameras are cost and computationally efficient but subject to performance degradation due to scene illumination and weather conditions. They are mainly utilized for vehicle [10], [37], pedestrian [38]–[40], lane marking [41]–[43], and traffic sign [18] detection in AVs. 360° or omnidirectional cameras can be used to obtain a panoramic view for navigation, localization, and mapping [44]. It is generally difficult to obtain accurate depth information from a single camera, but promising studies to improve monocular camera-based depth estimation are ongoing.

b) Stereo camera: depth information from a scene can be measured by a stereo camera system, similar to the human eye. Stereo cameras are commonly used for 3D mapping, better target classification, and long-range detection with better detection capacity than standard vision. Image processing of stereo cameras is more computationally demanding, and camera performance suffers in poor weather or lighting conditions.

c) Thermal camera: thermal cameras are used as standalone or with standard color cameras in object detection to overcome poor lighting [45], [46]. They are effective at pedestrian detection in low light conditions [47] and are useful for vehicle detection and tracking at night. Information from a thermal camera can be fused with data from other sources such as standard color cameras and lidar to get depth information in normal weather conditions.

2) Lidar: lidar is a relatively expensive sensor and utilizes IR light to measure its distance to targets, outputting a 3D point cloud. Lidars calculate target distance through either pulse measurement or phase shift measurement. Phase shift measurement is used for small distances and has a higher accuracy compared to pulse measurement, which is commonly used for long-range distance measurement and hence suitable for AVs. Lidar is suitable for the identification and recognition of road markings, pedestrians, bicyclists, and cars. A perception process utilizing lidar is generally divided into three steps: segmentation, fragmentation clustering, and tracking. The range of lidars is generally below 300 m, but is subject to performance degradation especially in extreme weather conditions such as fog and snow. Overall, lidars are most effective for mid-near range and multi-target object detection, though they cost more compared to other exteroceptive sensors.

3) Radar: compared to lidar, radar has a lower cost, is lightweight, and is small in size, but also has a lower accuracy. In AV applications, it is primarily used to measure the position and velocity of an object and is more reliable in extreme weather conditions than lidar or camera. As its performance is not affected by scene illumination, radar can also cover some of the shortcomings of camera. Radar is good at detecting vehicle-sized objects, but the detection task becomes challenging if the object is smaller. Moreover, due to its lower resolution precise shape estimation is challenging, though fusing with camera images can increase the precision and accuracy of such an operation.

Research groups and vehicle manufacturers worldwide have developed different AV platforms utilizing various sensor combinations, indicating each platform's approach to achieving full autonomy. Among them are not only platforms from academia such as Stanford's Junior [48], CMU's Boss [49], and RobotCar [50], but also commercial ones like the Tesla Autopilot [51], Uber Car (Ford Fusion) [52], Apollo Auto [53], Bertha [54], and Google's self-driving car [55]. A summary of various full-size AVs and their sensor combination is provided in Table III. Some of these platforms prioritize vision data while others favor that of lidar, with a few pursuing a balance between these two types of perception sensors. Further study is needed to understand the optimal number, type, and combination of sensors that achieve the best overall perception quality and redundancy while maintaining some level of cost-effectiveness.

B. Deep Learning-Based 2D Object Detection

Detection, state estimation, and motion prediction of dynamic objects on the road is the most challenging task facing an AV, as the ego vehicle needs to frequently update its path based on the predicted behavior of surrounding objects to prevent any hazardous situations. Computer vision research over the past few decades has enabled the detection and classification of thousands of static and dynamic objects in a scene (image frame) [58], first using traditional detection methods and from 2012 using DL [59]. This can be seen in the road-map of object detection milestones shown in Fig. 1. Detection of static objects has allowed AVs to understand traffic signs and traffic lights and obey basic driving rules.

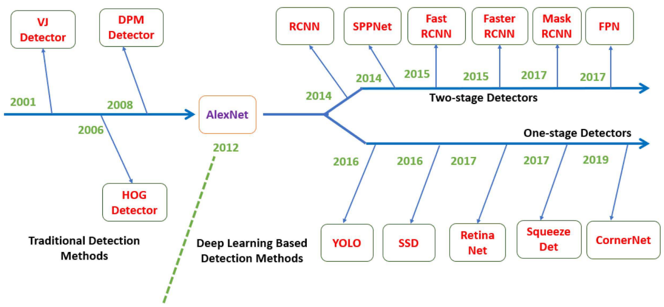

Fig. 1. Milestones of object detection over the last 20 years [58].

TABLE IV

COMPARISON OF THE ACCURACY OF DL OBJECT DETECTION ARCHITECTURES ON THE IMAGENET 1K TEST SET [17]

Architecture	No. of Parameters ($\times 10^6$)	No. of Layers	Top 5% Error
Inception-ResNet v2 [61]	30	95	4.9
Inception v4 [61]	41	75	5
ResNet 101 [62]	45	100	6.05
DenseNet 201 [63]	18	200	6.34
YOLO v3-608 [64]	63	53+1	6.2
ResNet 50 [62]	26	49	6.7
GoogLeNet [65]	6	22	6.7
VGG-16 [66]	134	13+2	6.8
AlexNet [60]	57	5+2	15.3

Moreover, the progress in object detection research in more recent years has accelerated research focused on the localization of dynamic objects, detection of their pose, and prediction of their short-term future trajectories to enable safe path planning for AVs. Though these dynamic objects – vehicles, pedestrians, bicyclists – can now be easily detected and classified, prediction of their future intention is still not an easy task. State-of-the-art object detection methods based on DL proposed in computer vision literature are highlighted in Table IV (ordered by Top 5% error). All these methods use CNN in some form. The number of parameters and layers is a good indicator of the computational load of the respective architecture. The research indicates that an ego vehicle utilizing one of those architectures [60]–[66] in its vision pipeline can detect and identify an unknown object with an accuracy of around 95%. However, real-time implementation of such heavy networks with online training is still challenging.

All state-of-the-art object detection methods used for AD are based on DL. They work by first detecting and classifying target object(s) and then drawing a bounding box around them to position those objects in the scene. These methods can be categorized as either two-stage or single-stage frameworks [67], with an overview of each category provided below.

1) Two-Stage Framework: the two-stage framework is also known as region proposal object detection. In this framework, general regions of interest are usually targeted in the first neural network. In the second stage, they are classified by a separate classifier network. A few methods belonging to this framework are R-CNN, Fast R-CNN, and Faster R-CNN. A detailed list of such methods and a discussion of them is available in [67]. Overall, two-stage object detection methods are more accurate but are also less computationally efficient, requiring more computational power and inference time.

2) Single-Stage Framework: compared to the previous category, methods in this category are generally faster and more computationally efficient, making them suitable for real-time object detection, but have less accuracy. YOLO [64], [68], [69] and SSD [70] are two examples of such methods.

TABLE V

AP OF COMMON 3D OBJECT DETECTION METHODS ON THE CAR CLASS OF THE KITTI 3D OBJECT DETECTION TEST SET [17]

Algorithm	T (s)	Easy	Moderate	Hard
PointRCNN [73]	0.10	85.9	75.8	68.3
PointPillars [74]	0.02	79.1	75.0	68.3
SECOND [75]	0.04	83.1	73.7	66.2
IPOD [76]	0.20	82.1	72.6	66.3
F-PointNet [77]	0.17	81.2	70.4	62.2
VoxelNet [78]	0.23	77.5	65.1	57.7
MV3D [79]	0.24	66.8	52.8	51.3

C. Deep Learning-Based 3D Object Detection

Lidar outputs 3D point clouds indicating the surfaces of a scene. If the data is sparse, it makes object detection and classification challenging. In general, lidar-based object detection methods consist of three steps: segmentation, clustering, and tracking [71], typically utilizing machine learning techniques such as SVM. The shape of objects and their motion characteristics [48], [72] can also be utilized to identify VRUs and cars. State-of-the-art 3D object detection methods commonly used in AVs are listed in Table V along with their AP on the car class of the KITTI 3D object detection data set. While Table V shows that these detection methods have greatly increased 3D object detection accuracy, convolution complexity still remains a challenge for real-time usage.

D. Pedestrian, Cyclist and Vehicle Detection

A critical task of AVs in a real traffic environment is detecting and tracking other cars and VRUs, the most important dynamic objects on the road. The performance of current state-of-the-art object detection methods for these object classes can be compared through studies like the one proposed by Lang *et al.* [74]. They considered various sensing configurations as well as object detection methods developed by them or other researchers and calculated the detection and classification mAP for each object class. For their first study, the authors used the KITTI BEV benchmark data set, and the comparative results are shown in Table VI. Their second study used the KITTI 3D detection benchmark data set, and the results are presented in Table VII. The tabulated results of case studies with moderate difficulty indicate that the PointPillars method performs better than almost all other methods and also outperforms them when fusion-based methods are applied to detect cars and cyclists. More research is still needed in this

TABLE VI

COMPARATIVE STUDY OF STATE-OF-THE-ART OBJECT DETECTION METHODS [74] (RESULTS ON THE KITTI BEV DETECTION BENCHMARK DATA SET)

Method	M	S	mAP	Car	Pe	C
			Mod.	**Mod.**	**Mod.**	**Mod.**
MV3D [79]		2.80	N/A	76.90	N/A	N/A
Cont-Fuse [80]		16.70	N/A	85.83	N/A	N/A
Roarnet [81]	L&I	10.00	N/A	79.41	N/A	N/A
AVOD-FPN [82]		10.00	64.11	83.79	**51.05**	57.48
F-PointNet [77]		5.90	65.39	84.00	50.22	61.96
HDNET [83]	L&M	20.00	N/A	**86.57**	N/A	N/A
PIXOR++ [83]		25.00	N/A	83.70	N/A	N/A
VoxelNet [78]		4.40	58.25	79.26	40.78	54.76
SECOND [75]	Lidar	20.00	60.56	79.37	46.27	56.04
PointPillars [74]		**62.00**	**66.19**	86.10	50.23	**62.25**

M: Modality; S: Speed (Hz); Pe: Pedestrian; C: Cyclist; L&I: Lidar & Image; L&M: Lidar & Map; mAP: Mean Average Precision

TABLE VII

COMPARATIVE STUDY OF STATE-OF-THE-ART OBJECT DETECTION METHODS [74] (RESULTS ON THE KITTI 3D DETECTION BENCHMARK DATA SET)

Method	M	S	mAP	Car	Pe	C
			Mod.	**Mod.**	**Mod.**	**Mod.**
MV3D [79]		2.8	N/A	62.35	N/A	N/A
Cont-Fuse [80]		16.7	N/A	66.22	N/A	N/A
Roarnet [81]	L&I	10	N/A	73.04	N/A	N/A
AVOD-FPN [82]		10	52.62	71.88	42.81	52.18
F-PointNet [77]		5.9	57.35	70.39	**44.89**	56.77
VoxelNet [78]		4.4	49.05	65.11	33.69	48.36
SECOND [75]	Lidar	20	56.69	73.66	42.56	53.85
PointPillars [74]		**62**	**59.2**	**74.99**	43.53	**59.07**

M: Modality; S: Speed (Hz); Pe: Pedestrian; C: Cyclist; L&I: Lidar & Image

area since the mAP of the current methods, especially when it comes to pedestrians and cyclists that are more vulnerable and frequently disobey traffic laws, is far lower than 90%.

E. Sensor Fusion-Based Object Detection

An accurate fusion of data collected from different sensory sources would dramatically improve object detection effectiveness. It allows different sensing modalities to reinforce each other's strengths and cover individual weaknesses. For sensor fusion, either all sensing modalities perform detection tasks simultaneously and then validate each other's results, or one modality performs the detection while others validate the data [84], [85]. In [86], human sensing performance is compared to ADS, where one of the key findings is that even though human drivers are still better at reasoning overall, the perception capabilities of an ADS utilizing sensor fusion can exceed that of humans, especially in degraded environmental conditions such as low scene lighting [17]. To that end, various sensor combinations commonly used for data fusion are briefly discussed below.

1) Radar-Camera Data Fusion [12]: in this fusion process, radar is mainly used for estimating RoI or distance, while recognition is carried out using cameras [87]–[96]. In two studies, guardrails' locations were determined by radar data, and vehicles were detected using the limited region's vertical symmetry features in image frames [88]. In similar approaches [94], [97], vehicles were detected using symmetry, edge information, and optical flow features of images. Once a vehicle was detected, its distance was calculated using radar-and-camera-fused data. That data was then projected onto a common global occupancy grid, where vehicles were tracked in a global frame of reference using a Kalman filter [89].

2) Lidar-Camera Data Fusion: some approaches have used lidar for reliable object detection while simultaneously using lidar and Camera to perform classification [85], [98], [99]. Others have used a camera for vehicle detection and lidar for ranging [100], [101]. MV3D, AVOD-FPN, and F-PointNet are some of the popular lidar-camera data fusion methods.

3) Radar-Lidar Data Fusion: Data from radar and lidar can be fused to improve the performance of state estimation and tracking of dynamic objects [102]. The state is estimated using Bayesian methods, extended Kalman filter, or particle filter, while data from two independent systems are fused for improved detection and tracking.

4) Radar-Lidar-Camera Data Fusion: through this fusion process, object detection and classification results from the camera are utilized to improve tracking model selection, data association, and movement classification [103], [104].

F. Challenges of Ego-Centric Object Detection

Although dynamic objects such as cars and pedestrians are well-structured and easy to detect, estimation of their dynamics and intent is not a simple task. Therefore, the following challenges have to be addressed to step closer to full autonomy.

1) Physical Limitations of Sensors: compared to camera images, a lidar measurement results in better 3D object detection accuracy and FoV. Motion-based object detection using a camera is sensitive to noise and scene lighting. On the contrary, lidar can work in low visibility environments and is not affected by low light conditions. Compared to radar, however, lidar performs less satisfyingly in rainy and snowy climates [37]. More research is still needed to address challenges arising from sensor physical limitations in scenarios with complicated scene lighting or extreme weather conditions.

2) Accuracy Issues: pedestrian detection accuracy of 2D object detection methods such as YOLO v3 or RetinaNet on some large-scale data sets such as COCO or ImageNet is usually much higher (around 85-95%) than it is on the KITTI 3D object detection data set (lower than 50%) that is much closer to real-world driving conditions. Because of this, a pedestrian may not be detected in some (from a couple to tens of) frames.

3) Reliability and Robustness Issues: despite significant progress in AV research and technology, the reliability and robustness of the perception sensor suite cannot be fully guaranteed. Some sensors may not work as well in low light conditions, while others may be rendered useless by snow or dirt, affecting the AV's performance despite sensor redundancy and sensor fusion. Because of this, finding answers to the following questions is crucial to making progress on sensor

TABLE VIII
OVERVIEW OF HUMAN MOTION PREDICTION METHODS

Category	Based on	Methods		Comments
Modeling approaches	Physics-based	Single-model methods [105]–[112]	Works using Newton's laws of motion	Needs dynamic model
		Multi-model methods [113]–[118]		
	Pattern-based	Sequential models [119]–[126]	Learns prototype of trajectory from observed behavior	Needs past trajectory or behavioral data
		Non-sequential models [122], [127]–[132]		
	Planning-based	Forward planning methods [133]–[141]	Reasons about likely goals and computes possible future paths	Explicit reasoning on long-term motion goals
		Inverse planning methods [142]–[151]		
Contextual cues	Target agent cues	Motion state [109], [117], [124], [127], [131], [136], [144], [145], [151]–[154]	Position and velocity	Needs to account for relevant internal and external stimuli that influence motion behavior
		Articulated pose [123], [155]–[160]	head orientation	
		Semantic attributes [117], [161]–[163]	Age, gender, personality, and awareness	
	Dynamic environment cues	unaware [107], [127], [164]–[169]	Does not care about presence of other agents	
		Individual-aware [108], [117], [119], [126], [131], [145], [153], [154]	Considers the presence of other agents	
		Group-aware [111], [170]–[175]	Considers the presence of other agents and social groupings	
	Static environment cues	Unaware [123], [129], [158], [176]–[182]	Assumes open-space environment	
		Obstacle-aware [120], [126], [131], [152]–[154], [183], [184]	Accounts for the presence of individual static obstacles	
		Map-aware [117], [118], [125], [137]–[139], [142], [147], [151], [185]–[191]	Accounts for the environment geometry and topology	
		Semantics-aware [106], [136], [144], [148], [161], [192]–[195]	Accounts for environment semantics	

reliability: (i) what should be done when a sensor fault occurs? (ii) How can the AV recognize defective data from a sensor? (iii) How to anticipate sensor failure? (iv) How to determine the absolute ground truth during extreme weather conditions?

4) Time Latency Issues: the total time latency from an occurrence in the environment to detection by an AV is dependent on the scan rate of sensors and the AV's computational speed. Processing image frames from cameras and point clouds from lidars require high computational power, without which there would be increased latency. In high-speed driving scenarios where an AV is going upward of 100 km/hr, a 1 second latency means traveling a distance of at least 36 m, significantly reducing the available braking distance in case of an emergency. Therefore, the total time latency should be below 100 ms to ensure safety for fully autonomous driving. What complicates this is that all recent object detection algorithms are DL-based, resulting in a much heavier computational load. Therefore, a trade-off has to be made between speed and accuracy. Some recent high-speed object detection algorithms such as YOLO v3-v5 and Inception v3 are gaining popularity but require a high-performance GPU for real-time application in ADS. Nevertheless, they have shown promising gains in both speed and accuracy. Further research is needed to build upon this progress.

III. STATE ESTIMATION AND MOTION PREDICTION

A. Pedestrian State Estimation and Motion Prediction

Accurate estimation of pedestrian state and future motion is challenging for AVs on the road, especially in heterogeneous traffic environments. So AVs need to analyze a pedestrian's past motion and present state and predict its future path. This is difficult because although most pedestrians frequently move along sidewalks and intersection crossings, they may behave randomly in some instances and not follow traffic rules, perhaps due to an external stimulus. The reaction to that stimulus may or may not be shared with other traffic agents, and those factors may or may not be observable or controllable by an AV. Hence an AV has to consider a multitude of factors, including a pedestrian's pose – standing, starting, walking, stopping – facial expressions, and movement through space to make an effective prediction of that pedestrian's future motion and intention. A summary of human motion prediction methods for AVs developed over the past few decades is presented in Table VIII, broadly categorized by modeling approach and contextual cues. These prediction methods are validated using real-time ground-truth data from data sets collected and standardized by various research and development communities. Table IX provides an overview of popular data sets available and used for human motion prediction and research works performed by researchers in this area.

Despite the progress [20], [21] made in the development of pedestrian state estimation and motion prediction techniques, their accuracy and reliability are still not fully guaranteed. This can be a problem because AVs need to make anticipatory actions for their short-term path plan based on accurate state estimation and motion prediction of the surrounding pedestrians. Furthermore, there is still no model for the prediction of the abnormal behavior of pedestrians walking on the roadside.

TABLE IX
SUMMARY OF EXISTING DATASETS ON HUMAN MOTION TRAJECTORIES

Data set (agent: person)	**Scene description**	**Used by**
ETH [109]	Two pedestrian scenes, top-down view, moderately crowded	[38], [39], [111], [119], [126], [131], [162], [186], [196]–[214]
UCY [215]	Two pedestrian scenes (sparsely populated Zara and crowded students), top-down view	[38], [39], [111], [119], [133], [161], [171], [196], [198]–[214], [216]
Stanford Drone Data Set [173] (with cyclists and vehicles)	Eight urban scenes, 900 m^2 each, top-down view, moderately populated	[106], [180], [196], [203], [205], [217]–[223]
Edinburgh [224]	One pedestrian scene, top-down view, 12×16 m^2, varying density of people	[138], [153], [225]–[228]
Grand Central Station Data Set [229]	Recorded in the crowded New York Grand Central train station	[38], [141], [226], [227], [230], [231]
VIRAT [232] (with cars and other vehicles)	Sixteen urban scenes, 20–50° camera view angle towards the ground plane (homographies included)	[139], [140], [144], [150], [225]
KITTI [233] (with cyclists and vehicles)	Recorded around the mid-size city of Karlsruhe (Germany), in rural areas and on highways	[136], [146], [234]–[236]
Town Center Data Set [237]	Pedestrians moving along a moderately populated street	[155], [161], [204], [231]
ATC [238]	Recording in a shopping center, 900 m^2 coverage, with varying density of people	[190], [239], [240]
Daimler Pedestrian Path Prediction Data Set [181]	Recorded from a moving/standing vehicle, with pedestrians crossing the street, stopping at the curb, or starting to move	[223], [241], [242]
L-CAS [243]	Recorded in a university building from a moving or standing robot	[199], [244]
TrajNet [245]	A superset of data sets, collecting relevant metrics and visualization tools	[231]

While V2X connectivity has been proposed as a solution, its feasibility is still not guaranteed since a pedestrian may not always be online throughout a traffic scenario. Some of the other difficulties in pedestrian state estimation and motion prediction are the following: variation in dimensions of the human body, presence of human pictures on street advertisements, dense or occluded pedestrian detection, and difficulty in real-time robust pedestrian detection.

B. Vehicular State Estimation and Motion Prediction

For any AV, other vehicles on the road are generally the primary concern at any time. Hence, accurate state estimation, tracking, and prediction of other vehicles' near-future paths and understanding their behavior is as important as that of pedestrians. This subsection briefly reviews and summarizes classical vehicle detection, state estimation, tracking, and motion prediction methods.

Vision-based vehicle detection has reached its maturity after decades of research in ML and DL, and the following tables (Tables X - XV) provide an overview of that research. Classic vision-based vehicle detection methods are presented in Table X, and are categorized by their usage of the motion or appearance of vehicles through monocular and stereo cameras. Alongside vehicle detection, state estimation and motion tracking are also essential for predicting the future position of vehicles on the road so that short and long-term path planning and collision avoidance are possible for the ego vehicle. Hence, Table XI highlights application-specific on-road vehicle tracking methods commonly used for monocular and stereo vision setups. Furthermore, Table XII presents the methods utilized for task-specific behavior analysis of on-road vehicles.

Table XIII provides a summary of the existing benchmark data sets for vehicle detection and trajectory prediction, and interested readers can refer to [10] for a detailed analysis and comprehensive review of vision-based vehicle detection, tracking, behavior analysis, and data sets used for this purpose. Alongside detection and tracking, motion prediction and maneuver intention estimation [12] of other vehicles are also equally important for an ego vehicle's safe trajectory planning and execution. Therefore, an overview of current motion prediction methods and their limitations is presented in Table XIV. Finally, methods used for maneuver intention estimation at road intersections are provided in Table XV.

While significant progress has been made in the development of vehicle detection and motion prediction methods, some challenges remain unsolved. Among them is a reduction in the performance of the current methods in extreme weather conditions. Another challenge is identification of abnormal driving behavior of other vehicles in real time. A final challenge is long-term motion prediction of other vehicles irrespective of traffic signals, where a multi-model tracking method is needed.

IV. COOPERATIVE PERCEPTION AND NAVIGATION

CPN refers to the practice of sharing perception and navigation information using V2V and V2X communication [340], [341] in a traffic network to better understand the surrounding environment and increase safety. Receiving perception information from other AVs can help the ego vehicle better understand blind spots or areas blocked by large objects. It can

TABLE X
SUMMARY OF VISION-BASED VEHICLE DETECTION METHODS

Vision type	Characteristic used	Method description
Monocular vision	Motion	Dynamic background modeling of overtaking area [246]
		Optical flow for blind-spot detection [247]
		Optical flow, HMM classification [248]
	Appearance	SVM and NN classification [249], HOG and Gabor features
		Statistical modeling of local features [250]
		Haar-like features, boosted classification, online learning [251]
		Haar-like features, Adaboost classification, active learning [252]
		HOG features, SVM classification. Orientation determined using multiplicative kernel learning [253]
		HOG features, deformable parts-based model [254]
		SURF and edge features, probabilistic classification, blind-spot detection [255]
Stereo vision	Motion	Optical flow [256]
		Occupancy grid, free space computation [257]
		Optical flow, clustering 6D points [258]
		Optical flow, particle-based occupancy grid [259]
		Tracking stixel and fitting probabilistic cuboid model [260]
		Optical flow, spatiotemporally smoothed occupancy grid [261]
	Appearance	Size, width, height, image intensity features, Bayesian classification [262]
		Clustering of 3D points, vehicle orientation estimation [263]
		Color, 3D vertical edges [264]
		V-disparity, clustering in the disparity space [265]

TABLE XI
SUMMARY OF VISION-BASED VEHICLE TRACKING METHODS

Tracking Method		Vision type	Application
Optical flow, geometric constraints, and Kalman filtering [266]		Monocular vision	Tracking and motion estimation
Template matching [267]			Tracking
Feature-based tracking and Kalman filtering [268], [269]			Detection and tracking
Particle filtering	Sivaraman and Trivedi, 2010 [252]		Tracking
	Quan et al., 2011 [253]		Tracking and orientation detection
	Xue and Ling, 2011 [270]		Tracking
	Niknejad et al., 2012 [254]		Detection and tracking
	Danescu et al., 2011 [259]	stereo vision	Position and velocity
Kalman filtering	Rabe et al., 2007 [271]		Motion estimation
	Bota and Nedevschi, 2011 [272]		Position and velocity (tracking)
Extended Kalman filtering	Barth and Franke, 2009 [258]		State and turning behavior estimation
	Lim et al., 2011 [273]		Tracking
Kalman filtering, interacting multiple models [274]			Motion estimation
Mean-shift on 3D points [275]			Tracking

also be an added layer of safety in case of sensor failure. Moreover, sharing trajectory information can help vehicles navigate more seamlessly, for example, by negotiating at intersections or forming highway platoons, or relevant platooning tasks [342]–[347].

The most straightforward approach to CPN is raw (or lightly-processed) information sharing, though this can be challenging due to bandwidth limitations and heavy communication load [348]. Aside from that, both fusing data received from a large variety of sensor arrays of other road users and processing a large volume of raw data can be computationally challenging. Therefore, a more common approach is to share processed perception information, for example an occupancy grid or a real-time map indicating the location, pose, and predicted trajectory of the surrounding objects, vehicles, and VRUs. This section briefly highlights major developments in this area and discusses open challenges facing CPN. Interested readers can visit [27] for a more comprehensive discussion.

A. Recent Progress in CPN

Working cooperatively benefits all vehicles in a network, as it improves every vehicle's understanding of the surrounding environment. In what follows, we list what vehicles stand to gain from cooperative perception and navigation:

1) It extends the LoS and FoV of every vehicle in the network. This, in turn, facilitates detection of traffic

TABLE XII
METHODS FOR ON-ROAD BEHAVIOR ANALYSIS

Specification	Method/classification	Task
Non-context-specific	Template matching score [276]	Detection and tracking of overtaking vehicles
	Dynamic Bayesian network [247]	Dynamic Bayesian network used to predict lane changes of other vehicles
	Optical flow direction, intensity [84]	Optical flows used to detect overtaking vehicles
Context-specific	Neural network [277]	Dynamic visual model of typical on-road behavior, saliency used to detect unusual and critical situations
	Interfacing multiple model likelihood [277]	Velocity and yaw-rate estimation used to infer the turning behavior of oncoming vehicles
	SVM [278]	Histograms of scene flow used to classify intersection vs. non-intersection driving environment
	Trajectory-based augmented particle filter [279]	Vehicle motion is matched to 44 prototypes using QELCS distance
	Trajectory-based HMM [280]	Unsupervised clustering of observed on-road trajectories

TABLE XIII
DATASETS FOR VEHICLE DETECTION AND TRAJECTORY PREDICTION

Data set	Scene description
Caltech 1999, 2001 [281], [282]	Static images of vehicles in a variety of poses
PETS 2001 [283]	Testing set of some 2867 frames from two cameras. Includes videos of preceding vehicles viewed through the front windshield, and a video of following vehicles viewed through the rear windshield
LISA 2010 [252]	Three short videos, 1500, 300, and 300 frames, comprised of highway and urban driving. Monocular detection of preceding vehicles only
Caraffi 2012 [284]	Several videos comprising some 20 minutes of driving on Italian highways
HighD Dataset 2018 [285]	Six different highway locations near Cologne, top-down view, varying densities with light and heavy traffic
Vehicles NGSIM 2006, 2007 [286], [287]	Recording of US Highway 101 and Interstate 80, road segment length 640 and 500 m
KITTI 2012 [233]	Recorded around the mid-sized city of Karlsruhe, Germany, in rural areas and on highways

TABLE XIV
SUMMARY OF VEHICLE MOTION PREDICTION METHODS

Based-on	Broad category	Sub-category	Limitations
Physics-based motion models	Evolution models	Dynamic models [288]–[294]	Limited to short-term motion prediction, unable to anticipate any change in the motion caused by the execution of a particular maneuver or changes caused by external factors
		Kinematic models [116], [288], [295]–[303]	
	Trajectory prediction	Single trajectory simulation [289], [296], [297], [299], [303]	
		Gaussian noise simulation [116], [296], [298], [301], [302], [304]–[306]	
		Monte Carlo simulation [293], [307], [308]	
Maneuver-based motion models [115], [310], [316]–[318]	Prototype trajectories [309]	Representation [120], [310]–[315]	Strictly deterministic representation of time, heavy computational burden, inability to consider physical limitations of a vehicle, and difficult to adapt to different road layouts
		Trajectory prediction [120], [309], [310], [312], [314], [319]–[321]	
	Maneuver intention estimation and maneuver execution	Maneuver intention estimation [105], [310], [316], [318], [322]–[329]	Inter-vehicle dependencies are particularly strong at road intersections, can lead to erroneous interpretations of the situation
		Maneuver execution [105], [310], [322], [330]–[332]	
Interaction-aware motion models	Models based on trajectory prototypes [128], [333]		Computationally expensive and not compatible with real-time risk assessment
	Models based on dynamic Bayesian networks [113], [334]–[339]		

congestion, avoidance of hidden obstacles and hazardous situations [349], safe lane changing/overtaking, and smooth path planning [350].

2) It helps AVs with short-term planning and control, for example, in immediate longitudinal control [351].

3) Speed and heading angle sharing through V2V communication can help with collision avoidance and complement emergency braking systems.

4) Cooperative intersection management through trajectory sharing can improve the safety of intersection naviga-

TABLE XV
SUMMARY OF MANEUVER INTENTION ESTIMATION METHODS AT ROAD INTERSECTION

Maneuvers	**Methods**
Stop, go straight, left turn, right turn [322]	Heuristics
Go straight, left turn, right turn [330]	
Safe errant [105]	SVM-BF
Lane-keeping, lane change left, lane change right [325]	
Lane-keeping, lane change left, lane change right [326]	SVM
Lane-keeping, lane change left, lane change right [324]	
Go straight, turn right, stop [323]	Logistic regression
Stop, brake, keep speed [318]	MLP
Complaint violating [316]	HMM
Go straight, left turn, right turn [310]	Hierarchical HMM
Go straight, left turn, right turn [327]	HMM
Lane-keeping, lane change left, lane change right [329]	
Go straight, turn left, turn right [328]	

tion [27]. This can lead to significant improvements because, for instance, during the ten years from 2005 to 2014, over 20% of the fatalities on EU roads took place at intersections [352] only. Therefore, such cooperative management algorithms, along with rule-based heuristic methods [353], and optimization-based methods [354], could make a noticeable difference in intersection safety.

B. Challenges Facing CPN

Despite recent developments and benefits listed above, CPN faces many challenges that need to be addressed before it can be widely adopted. These challenges include data privacy, data authenticity, handling data from malfunctioning sensors, development of a general architecture for cooperative data fusion, multi-object detection and tracking, and cooperative driving. Some of these challenges are further discussed below.

1) Data Transfer Decision: assuming that V2V communication is established between multiple vehicles for CP, each vehicle has to decide when and how to transmit or receive data:

a) Transmitter: some questions that need to be addressed are the following: what data to send? When and in what situation to send that data? How to assess a hazardous situation? If a nearby vehicle is in a hazardous situation, how to handle it? Among multiple nearby vehicles, how to select a target vehicle to send data? How to be aware of all nearby vehicles' relative positions in real-time?

b) Receiver: what data and how to fuse to extend FoV? Which received data to fuse for object detection if the ego vehicle failed to detect an object? How to select one transmitting vehicle among multiple such vehicles to receive data from? Or should data be received from all such vehicles? Should receiving data be continuous or selective? If continuous, how to handle the increased communication and processing burden? Overall, there needs to be a general framework for CP that defines protocols for data transmission and cooperative behavior. This can enable efficient implementation of CP and reduce potential compatibility issues during data transmission.

2) Data Reliability and Accuracy Issues: an AV connected to a cooperative network perceives the driving environment through several on-board sensors, among which a few are its own, and the rest are located on other vehicles. Therefore, the sensing accuracy is not only dependent on the sensors of an individual vehicle and their accuracy, but also on the performance of the overall network.

3) Data Association Issues: setting aside communication issues, it is still non-trivial to associate the information received from one vehicle with another vehicle's local understanding of the same situation [355]. Further research is needed to understand how the ego vehicle should select from among the data it receives and how that data should be fused with the ego vehicle's own sensory information.

4) Computing Issues: fusing perception data, driving decisions, and future trajectories requires high computational power. A possible solution may be VEC, through which the computational burden is offloaded to nearby edge computing servers, though further research is needed to investigate the viability of this method.

5) Time-Delay and Communication Issues: one area that requires further research is the impact of time delay [356] on the usability of information received through 5G or DSRC V2V and V2X communication. This concerns both information that travels from a single road user to another one and information that travels through a number of intermediaries to reach a road user. Analysis of the technical literature has shown that the lumped communication delay usually ranges from 200 to 800 ms, while the actuation time delay is typically within 20 to 250 ms [357]. According to [358], a lumped actuation delay is the combined result of pure time delays in (i) the engine response, (ii) the throttle actuator, (iii) the brake actuator, and (iv) low-pass filters used for sensors such as engine manifold pressure sensor, wheel speed sensor, etc.

6) Relative Pose and Localization Issues: effective fusion of data from onboard sensors and those obtained through communication requires knowledge of the relative pose and location of the surrounding road users. Determining this can become challenging when a large number of road users are present in a network.

V. FUTURE RESEARCH DIRECTIONS

Up to this point, this survey has presented an overview of research in various areas that enable the development and deployment of CAVs. While significant progess has been made in these areas, many are still facing challenges that require innovative solutions. These challenges and directions of future research are summarized below.

Though an enormous amount of research has been conducted on detection, estimation, and tracking techniques using different sensors for cars, trucks, and VRUs, further research on these methods and sensing modalities is needed so that an AV can confidently identify and predict the behavior of all road users. For vision-based object detection, it is usually difficult to obtain accurate depth information from a single camera, but promising studies to improve monocular camera-based depth estimation are ongoing. Stereo cameras perform much better in this regard, though their performance suffers in poor weather or lighting conditions and future works should address that, bringing their capabilities closer to the human eye. For lidar-based object detection, since sensor cost is a major factor, a future research track can be the study of the use of multiple, low-cost lidars with less dense point clouds instead of one expensive sensor, and how that can affect detection robustness and reliability. Further research is also needed to understand the optimal number, type, and combination of sensors that achieve the best overall perception quality, even in challenging weather and lighting conditions, while maintaining some level of cost-effectiveness.

While current research has made great strides in detecting and classifying vehicles and VRUs, further research is needed to increase object detection accuracy, particularly when it comes to smaller objects. More research is also needed to more accurately predict the intention of different road users and their future trajectories, which should be complemented with advances in computational hardware and software pipelines. This is especially important since VRUs such as pedestrians and cyclists are frequently present in urban traffic environments and may disobey traffic rules or behave unpredictably. While V2X and V2I connectivity have been proposed as means of increasing VRU awareness and enhancing their interaction with AVs, more research is needed to demonstrate the feasibility of this proposal. Future research should also address current challenges in pedestrian state estimation and motion prediction such as variations in human body dimensions, presence of human pictures on street or vehicular advertisements, and dense or occluded pedestrian detection.

While CPN looks like a promising approach for handling a future with a traffic mix of autonomous and manually-driven vehicles, it still faces many challenges that need to be addressed before it can be widely adopted. Some of these challenges are data privacy, data authenticity, data association, handling data from malfunctioning sensors, handling time-delay and communication issues, calculation of relative pose, and cooperative driving.

VI. Conclusion

This survey of the literature on state estimation and motion prediction of vehicles and VRUs summarized the significant progress that has been made in both categories, discussed the most promising results to date, and outlined the areas where further research is needed. In a review of the perception sensors most commonly used in AV research, we described the strengths and weaknesses of cameras, lidars, and radars, reviewed DL algorithms used for 2D and 3D object detection and noted that the most reliable detection results come from a fusion of data from different sensor modalities. We also outlined the areas that need further research, including sensor reliability and performance in extreme weather conditions. In the next section, we surveyed the literature on pedestrian and vehicle state estimation and motion prediction, categorizing existing detection, tracking, behavior analysis, and motion prediction algorithms and available benchmarking data sets. We also reviewed the progress made in the area of cooperative perception and navigation, using V2V and V2X communication to share perception and trajectory information for increased safety and traffic efficiency. While much research is still needed in this area to address several challenges such as data accuracy and association as well as time delay issues, this research can ultimately have a great impact on the widespread adoption of CAVs. Finally, possible future research directions have been proposed that can help address current challenges and accelerate the deployment of AVs on the road.

Acknowledgment

The authors are thankful to the Department of Mechanical Engineering, Virginia Tech, and the Safe-D project (Safety through Disruption, Safe-D National UTC) for supporting this research. In addition, the authors wish to express their gratitude to the anonymous reviewers for their valuable feedback that improved the quality of this paper.

References

[1] J. Stewart. (Aug. 27, 2018). Why Tesla's Autopilot Can't See a Stopped Firetruck. Wired. Accessed: Sep. 19, 2020. [Online]. Available: https://www.wired.com/story/tesla-autopilot-why-crash-radar/

[2] F. Lambert. (Jul. 1, 2016). *Understanding the Fatal Tesla Accident on Autopilot and the NHTSA Probe*. Electrek. Accessed: Sep. 19, 2020. [Online]. Available: https://electrek.co/2016/07/01/understanding-fatal-tesla-accident-autopilot-nhtsa-probe/

[3] *Preliminary Report Hwy18mh010*, National Transportation Safety Board, Washington, DC, USA, 2018.

[4] S. Singh, "Critical reasons for crashes investigated in the national motor vehicle crash causation survey," U.S. Dept. Transp., Nat. Highway Traffic Saf. Admin., Nat. Center Statist. Anal., Tech. Rep. DOT HS 812 115, Feb. 2015. [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/Publication/812506

[5] A. Eskandarian, *Handbook of Intelligent Vehicles*, vol. 2. Cham, Switzerland: Springer, 2012.

[6] X. Wu and A. Eskandarian, "An improved small-scale connected autonomous vehicle platform," in *Proc. ASME Dynamic Syst. Control Conf.*, New York, NY, USA: American Society of Mechanical Engineers Digital Collection, 2019, Art. no. V001T04A003.

[7] Y. Lin, C. Wu, and A. Eskandarian, "Integrating odometry and inter-vehicular communication for adaptive cruise control with target detection loss," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1848–1853.

[8] Y. Lin and A. Eskandarian, "Experimental evaluation of different controllers for cooperative adaptive cruise control," in *Proc. Dyn. Syst. Control Conf.*, vol. 58271. New York, NY, USA: American Society of Mechanical Engineers, 2017, Art. no. V001T44A006.

[9] Y. Lin and A. Eskandarian, "Experimental evaluation of cooperative adaptive cruise control with autonomous mobile robots," in *Proc. IEEE Conf. Control Technol. Appl. (CCTA)*, Aug. 2017, pp. 281–286.

[10] S. Sivaraman and M. M. Trivedi, "Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1773–1795, Dec. 2013.

[11] S. Lefèvre, D. Vasquez, and C. Laugier, "A survey on motion prediction and risk assessment for intelligent vehicles," *ROBOMECH J.*, vol. 1, no. 1, pp. 1–14, 2014.

[12] A. Mukhtar, L. Xia, and T. B. Tang, "Vehicle detection techniques for collision avoidance systems: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2318–2338, May 2015.

[13] B. Paden, M. Čáp, S. Z. Yong, D. Yershov, and E. Frazzoli, "A survey of motion planning and control techniques for self-driving urban vehicles," *IEEE Trans. Intell. Vehicles*, vol. 1, no. 1, pp. 33–55, Jun. 2016.

[14] D. González, J. Pérez, V. Milanés, and F. Nashashibi, "A review of motion planning techniques for automated vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 4, pp. 1135–1145, Apr. 2016.

[15] K. Abboud, H. A. Omar, and W. Zhuang, "Interworking of DSRC and cellular network technologies for V2X communications: A survey," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 9457–9470, Dec. 2016.

[16] G. Bresson, Z. Alsayed, L. Yu, and S. Glaser, "Simultaneous localization and mapping: A survey of current trends in autonomous driving," *IEEE Trans. Intell. Veh.*, vol. 2, no. 3, pp. 194–220, Sep. 2017.

[17] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A survey of autonomous driving: Common practices and emerging technologies," *IEEE Access*, vol. 8, pp. 58443–58469, 2020.

[18] H. Zhu, K.-V. Yuen, L. Mihaylova, and H. Leung, "Overview of environment perception for intelligent vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 10, pp. 2584–2601, Oct. 2017.

[19] W. Schwarting, J. Alonso-Mora, and D. Rus, "Planning and decision-making for autonomous vehicles," *Annu. Rev. Control, Robot., Auto. Syst.*, vol. 1, no. 1, pp. 187–210, May 2018.

[20] A. Rasouli and J. K. Tsotsos, "Autonomous vehicles that interact with pedestrians: A survey of theory and practice," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 900–918, Mar. 2020.

[21] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and K. O. Arras, "Human motion trajectory prediction: A survey," *Int. J. Robot. Res.*, vol. 39, no. 8, pp. 895–935, 2020.

[22] S. Pendleton *et al.*, "Perception, planning, control, and coordination for autonomous vehicles," *Machines*, vol. 5, no. 1, p. 6, Feb. 2017.

[23] S. Kuutti, S. Fallah, K. Katsaros, M. Dianati, F. Mccullough, and A. Mouzakitis, "A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 829–846, Apr. 2018.

[24] J. Van Brummelen, M. O'Brien, D. Gruyer, and H. Najjaran, "Autonomous vehicle perception: The technology of today and tomorrow," *Transp. Res. C, Emerg. Technol.*, vol. 89, pp. 384–406, Apr. 2018.

[25] A. Bighashdel and G. Dubbelman, "A survey on path prediction techniques for vulnerable road users: From traditional to deep-learning approaches," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 1039–1046.

[26] U. Montanaro *et al.*, "Towards connected autonomous driving: Review of use-cases," *Vehicle Syst. Dyn.*, vol. 57, no. 6, pp. 779–814, 2018.

[27] A. Eskandarian, C. Wu, and C. Sun, "Research advances and challenges of autonomous and connected ground vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 2, pp. 683–711, Feb. 2021.

[28] C. Badue *et al.*, "Self-driving cars: A survey," *Expert Syst. Appl.*, vol. 165, Mar. 2020, Art. no. 113816.

[29] D. Feng *et al.*, "Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 3, pp. 1341–1360, Mar. 2021.

[30] X. Yu and M. Marinov, "A study on recent developments and issues with obstacle detection systems for automated vehicles," *Sustainability*, vol. 12, no. 8, p. 3281, Apr. 2020.

[31] W. Wu *et al.*, "A survey of intrusion detection for in-vehicle networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 919–933, Mar. 2020.

[32] J.-W. Hu *et al.*, "A survey on multi-sensor fusion based obstacle detection for intelligent ground vehicles in off-road environments," *J. Frontiers Inf. Technol. Electron. Eng.*, vol. 21, no. 5, pp. 675–692, May 2020.

[33] L. Hu, J. Ou, J. Huang, Y. Chen, and D. Cao, "A review of research on traffic conflicts based on intelligent vehicles," *IEEE Access*, vol. 8, pp. 24471–24483, 2020.

[34] C. Pilz, A. Ulbel, and G. Steinbauer-Wagner, "The components of cooperative perception—A proposal for future works," in *Proc. IEEE Int. Intell. Transp. Syst. Conf. (ITSC)*, Sep. 2021, pp. 7–14.

[35] D. J. Yeong, G. Velasco-Hernandez, J. Barry, and J. Walsh, "Sensor and sensor fusion technology in autonomous vehicles: A review," *Sensors*, vol. 21, no. 6, p. 2140, Mar. 2021.

[36] M. Shan *et al.*, "Demonstrations of cooperative perception: Safety and robustness in connected and automated vehicle operations," *Sensors*, vol. 21, no. 1, p. 200, Dec. 2020.

[37] 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.

[38] Y. Xu, Z. Piao, and S. Gao, "Encoding crowd interaction with deep neural network for pedestrian trajectory prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5275–5284.

[39] Y. Luo, P. Cai, Y. Lee, and D. Hsu, "GAMMA: A general agent motion model for autonomous driving," 2019, *arXiv:1906.01566*.

[40] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 743–761, Apr. 2012.

[41] B. Qin *et al.*, "A general framework for road marking detection and analysis," in *Proc. 16th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2013, pp. 619–625.

[42] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, "Spatial as deep: Spatial CNN for traffic scene understanding," 2017, *arXiv:1712.06080*.

[43] D. Neven, B. D. Brabandere, S. Georgoulis, M. Proesmans, and L. V. Gool, "Towards end-to-end lane detection: An instance segmentation approach," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 286–291.

[44] J. Janai, F. Güney, A. Behl, and A. Geiger, "Computer vision for autonomous vehicles: Problems, datasets and state of the art," *Found. Trends Comput. Graph. Vis.*, vol. 12, nos. 1–3, pp. 1–308, 2020.

[45] C. Fries and H.-J. Wuensche, "Autonomous convoy driving by night: The vehicle tracking system," in *Proc. IEEE Int. Conf. Technol. Practical Robot Appl. (TePRA)*, May 2015, pp. 1–6.

[46] Q. Ha, K. Watanabe, T. Karasawa, Y. Ushiku, and T. Harada, "MFNet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 5108–5115.

[47] P. Hurney, P. Waldron, F. Morgan, E. Jones, and M. Glavin, "Review of pedestrian detection techniques in automotive far-infrared video," *IET Intell. Transp. Syst.*, vol. 9, no. 8, pp. 824–832, 2015.

[48] J. Levinson *et al.*, "Towards fully autonomous driving: Systems and algorithms," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 163–168.

[49] C. Urmson *et al.*, "Autonomous driving in urban environments: Boss and the urban challenge," *J. Field Robot.* vol. 25, no. 8, pp. 425–466, Aug. 2008.

[50] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 year, 1000 km: The Oxford robotcar dataset," *IJ Robot. Res.*, vol. 36, n0. 1, pp. 3–15, 2016.

[51] *Autopilot Press Kit*, Tesla Motors, Austin, TX, USA, Dec. 2018.

[52] H. Somerville, P. Lienert, and A. Sage, *Uber's Use of Fewer Safety Sensors Prompts Questions After Arizona Crash*. London, U.K.: Business News Reuters, 2018.

[53] *Baidu Apollo Auto*. Accessed: May 1, 2019. [Online]. Available: https://github.com/ApolloAuto/apollo

[54] J. Ziegler *et al.*, "Making Bertha drive—An autonomous journey on a historic route," *IEEE Intell. Transp. Syst. Mag.*, vol. 6, no. 2, pp. 8–20, Apr. 2014.

[55] C. Urmson, *Google Self-Driving Car Project*. Austin, TX, USA: South by Southwest (SXSW), 2016.

[56] A. Broggi *et al.*, "Extensive tests of autonomous driving technologies," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1403–1415, Sep. 2013.

[57] E. Guizzo, "How Google's self-driving car works," *IEEE Spectr.*, vol. 18, no. 7, pp. 1132–1141, Oct. 2011.

[58] C. B. Murthy, M. F. Hashmi, N. D. Bokde, and Z. W. Geem, "Investigations of object detection in images/videos using various deep learning techniques and embedded platforms—A comprehensive review," *Appl. Sci.*, vol. 10, no. 9, p. 3280, May 2020.

[59] Z. Zou, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," 2019, *arXiv:1905.05055*.

[60] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 2, pp. 84–90, Jun. 2012.

[61] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," 2016, *arXiv:1602.07261*.

[62] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[63] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[64] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[65] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[66] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[67] L. Liu *et al.*, "Deep learning for generic object detection: A survey," *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 261–318, Oct. 2020.

[68] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[69] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7263–7271.

[70] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, Oct. 2016, pp. 21–37.

[71] R. Dominguez, E. Onieva, J. Alonso, J. Villagra, and C. Gonzalez, "LIDAR based perception solution for autonomous vehicles," in *Proc. 11th Int. Conf. Intell. Syst. Design Appl.*, Nov. 2011, pp. 790–795.

[72] A. Teichman, J. Levinson, and S. Thrun, "Towards 3D object recognition via classification of arbitrary object tracks," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 4034–4041.

[73] S. Shi, X. Wang, and H. Li, "PointRCNN: 3D object proposal generation and detection from point cloud," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 770–779.

[74] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "Pointpillars: Fast encoders for object detection from point clouds," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 12697–12705.

[75] Y. Yan, Y. Mao, and B. Li, "SECOND: Sparsely embedded convolutional detection," *Sensors*, vol. 18, no. 10, p. 3337, 2018.

[76] Z. Yang, Y. Sun, S. Liu, X. Shen, and J. Jia, "IPOD: Intensive point-based object detector for point cloud," 2018, *arXiv:1812.05276*.

[77] C. R. Qi, W. Liu, C. Wu, H. Su, and L. J. Guibas, "Frustum PointNets for 3D object detection from RGB-D data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 918–927.

[78] Y. Zhou and O. Tuzel, "VoxelNet: End-to-end learning for point cloud based 3D object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4490–4499.

[79] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3D object detection network for autonomous driving," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1907–1915.

[80] M. Liang, B. Yang, S. Wang, and R. Urtasun, "Deep continuous fusion for multi-sensor 3D object detection," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 641–656.

[81] K. Shin, Y. P. Kwon, and M. Tomizuka, "RoarNet: A robust 3D object detection based on RegiOn approximation refinement," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 2510–2515.

[82] J. Ku, M. Mozifian, J. Lee, A. Harakeh, and S. L. Waslander, "Joint 3D proposal generation and object detection from view aggregation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 1–8.

[83] B. Yang, M. Liang, and R. Urtasun, "HDNet: Exploiting HD maps for 3D object detection," in *Proc. Conf. Robot Learn.*, 2018, pp. 146–155.

[84] 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 Intell. Vehicles Symp.*, Jun. 2012, pp. 494–499.

[85] S. A. Rodriguez F., V. Fremont, P. Bonnifait, and V. Cherfaoui, "Visual confirmation of mobile objects tracked by a multi-layer lidar," in *Proc. 13th Int. IEEE Conf. Intell. Transp. Syst.*, Sep. 2010, pp. 849–854.

[86] B. Schoettle, *Sensor Fusion: A Comparison of Sensing Capabilities of Human Drivers and Highly Automated Vehicles*. Ann Arbor, MI, USA: Univ. Michigan, 2017.

[87] X. Liu, Z. Sun, and H. He, "On-road vehicle detection fusing radar and vision," in *Proc. IEEE Int. Conf. Veh. Electron. Saf.*, Jul. 2011, pp. 150–154.

[88] 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.

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

[90] 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. Vehicles Symp.*, Jun. 2008, pp. 120–125.

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

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

[93] J. Fritsch *et al.*, "Towards a human-like vision system for driver assistance," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2008, pp. 275–282.

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

[95] Y. Tan, F. Han, and F. Ibrahim, "A radar guided vision system for vehicle validation and vehicle motion characterization," in *Proc. IEEE Intell. Transp. Syst. Conf.*, Sep. 2007, pp. 1059–1066.

[96] 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.

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

[98] 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 Intell. Transp. Syst. Conf.*, Sep. 2007, pp. 1044–1049.

[99] R. Chellappa, G. Qian, and Q. Zheng, "Vehicle detection and tracking using acoustic and video sensors," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2004, p. 793.

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

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

[102] C. Blanc, L. Trassoudaine, and J. Gallice, "EKF and particle filter track-to-track fusion: A quantitative comparison from radar/lidar obstacle tracks," in *Proc. 7th Int. Conf. Inf. Fusion*, Jul. 2005, p. 7.

[103] R. O. Chavez-Garcia and O. Aycard, "Multiple sensor fusion and classification for moving object detection and tracking," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 2, pp. 525–534, Feb. 2016.

[104] H. Cho, Y.-W. Seo, B. V. K. V. Kumar, and R. R. Rajkumar, "A multi-sensor fusion system for moving object detection and tracking in urban driving environments," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 1836–1843.

[105] G. S. Aoude, B. D. Luders, K. K. H. Lee, D. S. Levine, and J. P. How, "Threat assessment design for driver assistance system at intersections," in *Proc. 13th Int. IEEE Conf. Intell. Transp. Syst.*, Sep. 2010, pp. 1855–1862.

[106] P. Coscia, F. Castaldo, F. A. N. Palmieri, A. Alahi, S. Savarese, and L. Ballan, "Long-term path prediction in urban scenarios using circular distributions," *Image Vis. Comput.*, vol. 69, pp. 81–91, Jan. 2018.

[107] A. Elnagar, "Prediction of moving objects in dynamic environments using Kalman filters," in *Proc. IEEE Int. Symp. Comput. Intell. Robot. Autom.*, Jul. 2001, pp. 414–419.

[108] M. Luber, J. A. Stork, G. D. Tipaldi, and K. O. Arras, "People tracking with human motion predictions from social forces," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2010, pp. 464–469.

[109] S. Pellegrini, A. Ess, K. Schindler, and L. van Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 261–268.

[110] D. Petrich, T. Dang, D. Kasper, G. Breuel, and C. Stiller, "Map-based long term motion prediction for vehicles in traffic environments," in *Proc. 16th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2013, pp. 2166–2172.

[111] K. Yamaguchi, A. C. Berg, L. E. Ortiz, and T. L. Berg, "Who are you with and where are you going?" in *Proc. CVPR*, Jun. 2011, pp. 1345–1352.

[112] S. Zernetsch, S. Kohnen, M. Goldhammer, K. Doll, and B. Sick, "Trajectory prediction of cyclists using a physical model and an artificial neural network," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2016, pp. 833–838.

[113] G. Agamennoni, J. I. Nieto, and E. M. Nebot, "Estimation of multi-vehicle dynamics by considering contextual information," *IEEE Trans. Robot.*, vol. 28, no. 4, pp. 855–870, Aug. 2012.

[114] M. Althoff, O. Stursberg, and M. Buss, "Reachability analysis of nonlinear systems with uncertain parameters using conservative linearization," in *Proc. 47th IEEE Conf. Decis. Control*, Dec. 2008, pp. 4042–4048.

[115] 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. Intell. Transp. Syst.*, Sep. 2010, pp. 1625–1631.

[116] N. Kaempchen, K. Weiss, M. Schaefer, and K. C. J. Dietmayer, "IMM object tracking for high dynamic driving maneuvers," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2004, pp. 825–830.

[117] J. Kooij, F. Flohr, E. Pool, and D. Gavrila, "Context-based path prediction for targets with switching dynamics," *Int. J. Comput. Vis.*, vol. 127, no. 3, pp. 239–262, Mar. 2019.

[118] E. A. I. Pool, J. F. P. Kooij, and D. M. Gavrila, "Using road topology to improve cyclist path prediction," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2017, pp. 289–296.

[119] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human trajectory prediction in crowded spaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 961–971.

[120] G. Aoude, J. Joseph, N. Roy, and J. How, *Mobile Agent Trajectory Prediction Using Bayesian Nonparametric Reachability Trees*. St. Louis, MO, USA: American Institute of Aeronautics and Astronautics, 2011, p. 1512. [Online]. Available: https://dspace.mit.edu/bitstream/handle/1721.1/114899/Aoude_Infotech11.pdf?sequence=1&isAllowed=y

[121] M. Goldhammer, K. Doll, U. Brunsmann, A. Gensler, and B. Sick, "Pedestrian's trajectory forecast in public traffic with artificial neural networks," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 4110–4115.

[122] C. G. Keller and D. M. Gavrila, "Will the pedestrian cross? A study on pedestrian path prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 494–506, Apr. 2014.

[123] E. Kruse and F. M. Wahl, "Camera-based observation of obstacle motions to derive statistical data for mobile robot motion planning," in *Proc. IEEE Int. Conf. Robot. Automat.*, vol. 1, May 1998, pp. 662–667.

[124] T. P. Kucner, M. Magnusson, E. Schaffernicht, V. H. Bennetts, and A. J. Lilienthal, "Enabling flow awareness for mobile robots in partially observable environments," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 1093–1100, Apr. 2017.

[125] L. Liao, D. Fox, J. Hightower, H. Kautz, and D. Schulz, "Voronoi tracking: Location estimation using sparse and noisy sensor data," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2003, pp. 723–728.

[126] A. Vemula, K. Muelling, and J. Oh, "Modeling cooperative navigation in dense human crowds," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 1685–1692.

[127] M. Bennewitz, W. Burgard, G. Cielniak, and S. Thrun, "Learning motion patterns of people for compliant robot motion," *Int. J. Robot. Res.*, vol. 24, no. 1, pp. 31–48, 2005.

[128] E. Käfer, C. Hermes, C. Wöhler, H. Ritter, and F. Kummert, "Recognition of situation classes at road intersections," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2010, pp. 3960–3965.

[129] M. Luber, L. Spinello, J. Silva, and K. O. Arras, "Socially-aware robot navigation: A learning approach," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 902–907.

[130] M. K. C. Tay and C. Laugier, "Modelling smooth paths using Gaussian processes," in *Field and Service Robotics*. Berlin, Germany: Springer, 2008, pp. 381–390.

[131] P. Trautman and A. Krause, "Unfreezing the robot: Navigation in dense, interacting crowds," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2010, pp. 797–803.

[132] S. Xiao, Z. Wang, and J. Folkesson, "Unsupervised robot learning to predict person motion," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2015, pp. 691–696.

[133] G. Best and R. Fitch, "Bayesian intention inference for trajectory prediction with an unknown goal destination," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2015, pp. 5817–5823.

[134] A. Bruce and G. Gordon, "Better motion prediction for people-tracking," in *Proc. Int. Conf. Robot. Automat. (ICRA)*, Barcelona, Spain, Apr. 2004, pp. 1–6.

[135] E. Galceran, A. G. Cunningham, R. M. Eustice, and E. Olson, "Multi-policy decision-making for autonomous driving via changepoint-based behavior prediction," in *Proc. Robot., Sci. Syst.*, vol. 1, 2015, p. 6.

[136] V. Karasev, A. Ayvaci, B. Heisele, and S. Soatto, "Intent-aware long-term prediction of pedestrian motion," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2016, pp. 2543–2549.

[137] C. Rosmann, M. Oeljeklaus, F. Hoffmann, and T. Bertram, "Online trajectory prediction and planning for social robot navigation," in *Proc. IEEE Int. Conf. Adv. Intell. Mechatronics (AIM)*, Jul. 2017, pp. 1255–1260.

[138] A. Rudenko, L. Palmieri, and K. O. Arras, "Predictive planning for a mobile robot in human environments," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA), Workshop PlanRob*, Sep. 2017, pp. 1–7.

[139] D. Vasquez, "Novel planning-based algorithms for human motion prediction," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2016, pp. 3317–3322.

[140] D. Xie, S. Todorovic, and S.-C. Zhu, "Inferring," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jan. 2004, pp. 2224–2231.

[141] S. Yi, H. Li, and X. Wang, "Pedestrian behavior modeling from stationary crowds with applications to intelligent surveillance," *IEEE Trans. Image Process.*, vol. 25, no. 9, pp. 4354–4368, Sep. 2016.

[142] S.-Y. Chung and H.-P. Huang, "Incremental learning of human social behaviors with feature-based spatial effects," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 2417–2422.

[143] S. Y. Huang *et al.*, "Deep learning driven visual path prediction from a single image," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5892–5904, Dec. 2016.

[144] K. M. Kitani, B. D. Ziebart, J. A. Bagnell, and M. Hebert, "Activity forecasting," in *Proc. Eur. Conf. Comput. Vis.*, Berlin, Germany: Springer, Oct. 2012, pp. 201–214.

[145] M. Kuderer, H. Kretzschmar, C. Sprunk, and W. Burgard, "Feature-based prediction of trajectories for socially compliant navigation," in *Proc. Robot., Sci. Syst. VIII*, Jul. 2012, pp. 1–8.

[146] N. Lee, W. Choi, P. Vernaza, C. B. Choy, P. H. S. Torr, and M. Chandraker, "DESIRE: Distant future prediction in dynamic scenes with interacting agents," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 336–345.

[147] M. Pfeiffer, U. Schwesinger, H. Sommer, E. Galceran, and R. Siegwart, "Predicting actions to act predictably: Cooperative partial motion planning with maximum entropy models," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2016, pp. 2096–2101.

[148] E. Rehder, F. Wirth, M. Lauer, and C. Stiller, "Pedestrian prediction by planning using deep neural networks," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 1–5.

[149] M. Shen, G. Habibi, and J. P. How, "Transferable pedestrian motion prediction models at intersections," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 4547–4553.

[150] J. Walker, A. Gupta, and M. Hebert, "Patch to the future: Unsupervised visual prediction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3302–3309.

[151] B. D. Ziebart *et al.*, "Planning-based prediction for pedestrians," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2009, pp. 3931–3936.

[152] A. Bera, S. Kim, T. Randhavane, S. Pratapa, and D. Manocha, "GLMP-realtime pedestrian path prediction using global and local movement patterns," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2016, pp. 5528–5535.

[153] J. Elfring, R. van de Molengraft, and M. Steinbuch, "Learning intentions for improved human motion prediction," *Robot. Auto. Syst.*, vol. 62, no. 4, pp. 591–602, Apr. 2014.

[154] G. Ferrer and A. Sanfeliu, "Behavior estimation for a complete framework for human motion prediction in crowded environments," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 5940–5945.

[155] I. Hasan, F. Setti, T. Tsesmelis, A. Del Bue, F. Galasso, and M. Cristani, "MX-LSTM: Mixing tracklets and vislets to jointly forecast trajectories and head poses," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6067–6076.

[156] J. F. P. Kooij, N. Schneider, F. Flohr, and D. M. Gavrila, "Context-based pedestrian path prediction," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, Sep. 2014, pp. 618–633.

[157] M. Roth, F. Flohr, and D. M. Gavrila, "Driver and pedestrian awareness-based collision risk analysis," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2016, pp. 454–459.

[158] V. V. Unhelkar, C. Perez-D'Arpino, L. Stirling, and J. A. Shah, "Human-robot co-navigation using anticipatory indicators of human walking motion," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2015, pp. 6183–6190.

[159] R. Q. Mínguez, I. P. Alonso, D. Fernández-Llorca, and M. Á. Sotelo, "Pedestrian path, pose, and intention prediction through Gaussian process dynamical models and pedestrian activity recognition," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 5, pp. 1803–1814, May 2019.

[160] R. Quintero, J. Almeida, D. F. Llorca, and M. A. Sotelo, "Pedestrian path prediction using body language traits," in *Proc. IEEE Intell. Vehicles Symp. Proc.*, Jun. 2014, pp. 317–323.

[161] W.-C. Ma, D.-A. Huang, N. Lee, and K. M. Kitani, "Forecasting interactive dynamics of pedestrians with fictitious play," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 774–782.

[162] A. Bera, T. Randhavane, and D. Manocha, “Aggressive, tense or shy? Identifying personality traits from crowd videos,” in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 112–118.

[163] S. Oli, B. L’Esperance, and K. Gupta, “Human motion behaviour aware planner (HMBAP) for path planning in dynamic human environments,” in *Proc. 16th Int. Conf. Adv. Robot. (ICAR)*, Nov. 2013, pp. 1–7.

[164] A. Elnagar and K. Gupta, “Motion prediction of moving objects based on autoregressive model,” *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 28, no. 6, pp. 803–810, Nov. 1998.

[165] K. Kim, D. Lee, and I. Essa, “Gaussian process regression flow for analysis of motion trajectories,” in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1164–1171.

[166] T. Kucner, J. Saarinen, M. Magnusson, and A. J. Lilienthal, “Conditional transition maps: Learning motion patterns in dynamic environments,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nov. 2013, pp. 1196–1201.

[167] S. Thompson, T. Horiuchi, and S. Kagami, “A probabilistic model of human motion and navigation intent for mobile robot path planning,” in *Proc. 4th Int. Conf. Auto. Robots Agents*, Feb. 2009, pp. 663–668.

[168] Z. Wang, P. Jensfelt, and J. Folkesson, “Building a human behavior map from local observations,” in *Proc. 25th IEEE Int. Symp. Robot Human Interact. Commun. (RO-MAN)*, Aug. 2016, pp. 64–70.

[169] Q. Zhu, “Hidden Markov model for dynamic obstacle avoidance of mobile robot navigation,” *IEEE Trans. Robot. Autom.*, vol. 7, no. 3, pp. 390–397, Jun. 2016.

[170] I. Karamouzas and M. Overmars, “Simulating and evaluating the local behavior of small pedestrian groups,” *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 3, pp. 394–406, Mar. 2012.

[171] S. Pellegrini, A. Ess, and L. Van Gool, “Improving data association by joint modeling of pedestrian trajectories and groupings,” in *Proc. Eur. Conf. Comput. Vis.*, Berlin, Germany: Springer, Sep. 2010, pp. 452–465.

[172] F. Qiu and X. Hu, “Modeling group structures in pedestrian crowd simulation,” *Simul. Model. Pract. Theory*, vol. 18, pp. 190–205, Feb. 2010.

[173] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, “Learning social etiquette: Human trajectory understanding in crowded scenes,” in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2016, pp. 549–565.

[174] M. Seitz, G. Köster, and A. Pfaffinger, *Pedestrian Group Behavior in a Cellular Automaton*. Cham, Switzerland: Springer, 2014, pp. 807–814.

[175] H. Singh, R. Arter, L. Dodd, P. Langston, E. Lester, and J. Drury, “Modelling subgroup behaviour in crowd dynamics DEM simulation,” *Appl. Math. Model.*, vol. 33, no. 12, pp. 4408–4423, Dec. 2009.

[176] M. Bennewitz, W. Burgard, and S. Thrun, “Using EM to learn motion behaviors of persons with mobile robots,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2002, pp. 502–507.

[177] D. Ellis, E. Sommerlade, and I. Reid, “Modelling pedestrian trajectory patterns with Gaussian processes,” in *Proc. IEEE 12th Int. Conf. Comput. Vis. Workshops, (ICCV) Workshops*, Sep. 2009, pp. 1229–1234.

[178] S. Ferguson, B. Luders, R. C. Grande, and J. P. How, *Real-Time Predictive Modeling and Robust Avoidance of Pedestrians With Uncertain, Changing Intentions*. Cham, Switzerland: Springer, 2015, pp. 161–177.

[179] A. F. Foka and P. E. Trahanias, “Probabilistic autonomous robot navigation in dynamic environments with human motion prediction,” *Int. J. Social Robot.*, vol. 2, no. 1, pp. 79–94, Mar. 2010.

[180] H. O. Jacobs, O. K. Hughes, M. Johnson-Roberson, and R. Vasudevan, “Real-time certified probabilistic pedestrian forecasting,” *IEEE Robot. Autom. Lett.*, vol. 2, no. 4, pp. 2064–2071, Oct. 2017.

[181] N. Schneider and D. M. Gavrila, “Pedestrian path prediction with recursive Bayesian filters: A comparative study,” in *Proc. German Conf. Pattern Recognit.*, Cham, Switzerland: Springer, 2013, pp. 174–183.

[182] D. Vasquez, T. Fraichard, O. Aycard, and C. Laugier, “Intentional motion on-line learning and prediction,” *Mach. Vis. Appl.*, vol. 19, nos. 5–6, pp. 411–425, Oct. 2008.

[183] M. Althoff, O. Stursberg, and M. Buss, “Stochastic reachable sets of interacting traffic participants,” in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2008, pp. 1086–1092.

[184] E. Rehder and H. Kloeden, “Goal-directed pedestrian prediction,” in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Dec. 2015, pp. 50–58.

[185] Y. F. Chen, M. Liu, M. Everett, and J. P. How, “Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 285–292.

[186] S.-Y. Chung and H.-P. Huang, “A mobile robot that understands pedestrian spatial behaviors,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2010, pp. 5861–5866.

[187] H. Gong, J. Sim, M. Likhachev, and J. Shi, “Multi-hypothesis motion planning for visual object tracking,” in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 619–626.

[188] P. Henry, C. Vollmer, B. Ferris, and D. Fox, “Learning to navigate through crowded environments,” in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2010, pp. 981–986.

[189] T. Ikeda, Y. Chigodo, D. Rea, F. Zanlungo, M. Shiomi, and T. Kanda, “Modeling and prediction of pedestrian behavior based on the sub-goal concept,” *Robotics*, vol. 10, pp. 137–144, Jul. 2013.

[190] A. Rudenko, L. Palmieri, A. J. Lilienthal, and K. O. Arras, “Human motion prediction under social grouping constraints,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 3358–3364.

[191] H. Chieh Yen, H. Pang Huang, and S. Yun Chung, “Goal-directed pedestrian model for long-term motion prediction with application to robot motion planning,” in *Proc. IEEE Workshop Adv. Robot. Social Impacts*, Aug. 2008, pp. 1–6.

[192] L. Ballan, F. Castaldo, A. Alahi, F. Palmieri, and S. Savarese, “Knowledge transfer for scene-specific motion prediction,” in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Cham, Switzerland: Springer, 2016, pp. 697–713.

[193] F. Kuhnt, J. Schulz, T. Schamm, and J. M. Zollner, “Understanding interactions between traffic participants based on learned behaviors,” in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2016, pp. 1271–1278.

[194] N. Lee and K. M. Kitani, “Predicting wide receiver trajectories in American football,” in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2016, pp. 1–9.

[195] S. Zheng, Y. Yue, and J. Hobbs, “Generating long-term trajectories using deep hierarchical networks,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1543–1551.

[196] D. Varshneya and G. Srinivasaraghavan, “Human trajectory prediction using spatially aware deep attention models,” 2017, *arXiv:1705.09436*.

[197] S. Kim *et al.*, “BRVO: Predicting pedestrian trajectories using velocity-space reasoning,” *Int. J. Robot. Res.*, vol. 34, no. 2, pp. 201–217, Feb. 2015.

[198] A. Vemula, K. Muelling, and J. Oh, “Social attention: Modeling attention in human crowds,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 1–7.

[199] N. Radwan, W. Burgard, and A. Valada, “Multimodal interaction-aware motion prediction for autonomous street crossing,” *Int. J. Robot. Res.*, vol. 39, no. 13, pp. 1567–1598, Nov. 2020.

[200] M. Pfeiffer, G. Paolo, H. Sommer, J. Nieto, R. Siegwart, and C. Cadena, “A data-driven model for interaction-aware pedestrian motion prediction in object cluttered environments,” in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 1–8.

[201] N. Bisagno, B. Zhang, and N. Conci, “Group LSTM: Group trajectory prediction in crowded scenarios,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 1–12.

[202] P. Zhang, W. Ouyang, P. Zhang, J. Xue, and N. Zheng, “SR-LSTM: State refinement for LSTM towards pedestrian trajectory prediction,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12085–12094.

[203] T. Zhao *et al.*, “Multi-agent tensor fusion for contextual trajectory prediction,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 12126–12134.

[204] H. Xue, D. Q. Huynh, and M. Reynolds, “SS-LSTM: A hierarchical LSTM model for pedestrian trajectory prediction,” in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 1186–1194.

[205] A. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, H. Rezatofighi, and S. Savarese, “SoPhie: An attentive GAN for predicting paths compliant to social and physical constraints,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1349–1358.

[206] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, “Social GAN: Socially acceptable trajectories with generative adversarial networks,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2255–2264.

[207] M. Huynh and G. Alaghband, “Trajectory prediction by coupling scene-LSTM with human movement LSTM,” in *Proc. Int. Symp. Vis. Comput.*, Springer, Oct. 2019, pp. 244–259.

[208] N. Nikhil and B. T. Morris, “Convolutional neural network for trajectory prediction,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 186–196.

[209] Z. Pei, X. Qi, Y. Zhang, M. Ma, and Y.-H. Yang, "Human trajectory prediction in crowded scene using social-affinity long short-term memory," *Pattern Recognit.*, vol. 93, pp. 273–282, Sep. 2019.

[210] Y. Huang, H. Bi, Z. Li, T. Mao, and Z. Wang, "STGAT: Modeling spatial-temporal interactions for human trajectory prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6272–6281.

[211] J. Amirian, J.-B. Hayet, and J. Pettre, "Social ways: Learning multi-modal distributions of pedestrian trajectories with GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 2964–2972.

[212] C. Blaiotta, "Learning generative socially aware models of pedestrian motion," *IEEE Robot. Autom. Lett.*, vol. 4, no. 4, pp. 3433–3440, Oct. 2019.

[213] V. Kosaraju, A. Sadeghian, R. Martín-Martín, I. Reid, H. Rezatofighi, and S. Savarese, "Social-BiGAT: Multimodal trajectory forecasting using bicycle-GAN and graph attention networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 137–146.

[214] B. Ivanovic and M. Pavone, "The trajectron: Probabilistic multi-agent trajectory modeling with dynamic spatiotemporal graphs," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2375–2384.

[215] A. Lerner, Y. Chrysanthou, and D. Lischinski, "Crowds by example," *Comput. Graph. Forum*, vol. 26, no. 3, pp. 655–664, 2007.

[216] F. Bartoli, G. Lisanti, L. Ballan, and A. D. Bimbo, "Context-aware trajectory prediction," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 1941–1946.

[217] T. van der Heiden, N. S. Nagaraja, C. Weiss, and E. Gavves, "Safe-Critic: Collision-aware trajectory prediction," 2019, *arXiv:1910.06673*.

[218] Y. Chai, B. Sapp, M. Bansal, and D. Anguelov, "MultiPath: Multiple probabilistic anchor trajectory hypotheses for behavior prediction," 2019, *arXiv:1910.05449*.

[219] T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Neighbourhood context embeddings in deep inverse reinforcement learning for predicting pedestrian motion over long time horizons," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 1179–1187.

[220] O. Makansi, E. Ilg, O. Cicek, and T. Brox, "Overcoming limitations of mixture density networks: A sampling and fitting framework for multimodal future prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7144–7153.

[221] S. Eiffert and S. Sukkarieh, "Predicting responses to a robot's future motion using generative recurrent neural networks," 2019, *arXiv:1909.13486*.

[222] D. Ridel, N. Deo, D. Wolf, and M. Trivedi, "Scene compliant trajectory forecast with agent-centric spatio-temporal grids," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 2816–2823, Apr. 2020.

[223] K. Saleh, M. Hossny, and S. Nahavandi, "Intent prediction of pedestrians via motion trajectories using stacked recurrent neural networks," *IEEE Trans. Intell. Veh.*, vol. 3, no. 4, pp. 414–424, Dec. 2018.

[224] B. Majecka, "Statistical models of pedestrian behaviour in the forum," M.S. thesis, School Inform., Univ. Edinburgh, Edinburgh, Scotland, 2009.

[225] F. Previtali, A. Bordallo, L. Iocchi, and S. Ramamoorthy, "Predicting future agent motions for dynamic environments," in *Proc. 15th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2016, pp. 94–99.

[226] H. Xue, D. Q. Huynh, and M. Reynolds, "Bi-prediction: Pedestrian trajectory prediction based on bidirectional LSTM classification," in *Proc. Int. Conf. Digit. Image Comput., Techn. Appl. (DICTA)*, Nov. 2017, pp. 1–8.

[227] T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Soft+ hardwired attention: An LSTM framework for human trajectory prediction and abnormal event detection," *Neural Netw.*, vol. 108, pp. 466–478, Dec. 2018.

[228] J. F. Carvalho, M. Vejdemo-Johansson, F. T. Pokorny, and D. Kragic, "Long-term prediction of motion trajectories using path homology clusters," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Nov. 2019, pp. 3–8.

[229] B. Zhou, X. Wang, and X. Tang, "Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2871–2878.

[230] H. Su, J. Zhu, Y. Dong, and B. Zhang, "Forecast the plausible paths in crowd scenes," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, p. 2.

[231] H. Xue, D. Huynh, and M. Reynolds, "Location-velocity attention for pedestrian trajectory prediction," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2019, pp. 2038–2047.

[232] S. Oh *et al.*, "A large-scale benchmark dataset for event recognition in surveillance video," in *Proc. CVPR*, Jun. 2011, pp. 3153–3160.

[233] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361.

[234] J. Wu, J. Ruenz, and M. Althoff, "Probabilistic map-based pedestrian motion prediction taking traffic participants into consideration," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1285–1292.

[235] N. Rhinehart, K. M. Kitani, and P. Vernaza, "R2P2: A reparameterized pushforward policy for diverse, precise generative path forecasting," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 772–788.

[236] S. Srikanth, J. A. Ansari, R. K. Ram, S. Sharma, J. K. Murthy, and K. M. Krishna, "INFER: INtermediate representations for FuturE pRediction," 2019, *arXiv:1903.10641*.

[237] B. Benfold and I. Reid, "Stable multi-target tracking in real-time surveillance video," in *Proc. CVPR*, Jun. 2011, pp. 3457–3464.

[238] D. Brscic, T. Kanda, T. Ikeda, and T. Miyashita, "Person tracking in large public spaces using 3-D range sensors," *IEEE Trans. Human-Mach. Syst.*, vol. 43, no. 6, pp. 522–534, Nov. 2013.

[239] A. Rudenko, L. Palmieri, and K. O. Arras, "Joint long-term prediction of human motion using a planning-based social force approach," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 1–7.

[240] S. Molina *et al.*, "Modelling and predicting rhythmic flow patterns in dynamic environments," in *Proc. U. K.-RAS Conf., Robots Work. Among Us*, Oct. 2018, pp. 135–146.

[241] K. Saleh, M. Hossny, and S. Nahavandi, "Contextual recurrent predictive model for long-term intent prediction of vulnerable road users," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 8, pp. 3398–3408, Aug. 2020.

[242] A. T. Schulz and R. Stiefelhagen, "A controlled interactive multiple model filter for combined pedestrian intention recognition and path prediction," in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst.*, Sep. 2015, pp. 173–178.

[243] Z. Yan, T. Duckett, and N. Bellotto, "Online learning for human classification in 3D LiDAR-based tracking," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 864–871.

[244] L. Sun, Z. Yan, S. M. Mellado, M. Hanheide, and T. Duckett, "3DOF pedestrian trajectory prediction learned from long-term autonomous mobile robot deployment data," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 1–7.

[245] A. Sadeghian, V. Kosaraju, A. Gupta, S. Savarese, and A. Alahi, "TrajNet: Towards a benchmark for human trajectory prediction," 2018, *arXiv:1805.07663*.

[246] 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.

[247] D. Kasper *et al.*, "Object-oriented Bayesian networks for detection of lane change maneuvers," *IEEE Intell. Transp. Syst. Mag.*, vol. 4, no. 3, pp. 19–31, Aug. 2012.

[248] 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.

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

[250] C.-C.-R. Wang and J.-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.

[251] 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.

[252] S. Sivaraman and M. 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.

[253] 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.

[254] H. T. 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.

[255] B. F. Lin *et al.*, "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, Feb. 2012.

[256] U. Franke, C. Rabe, H. Badino, and S. Gehrig, "6D-vision: Fusion of stereo and motion for robust environment perception," in *Proc. Joint Pattern Recognit. Symp.*, Berlin, Germany: Springer, Aug. 2005, pp. 216–223.

[257] H. Badino, U. Franke, and R. Mester, "Free space computation using stochastic occupancy grids and dynamic programming," in *Proc. Workshop Dyn. Vis., (ICCV)*, vol. 20. Rio de Janeiro, Brazil: Citeseer, Oct. 2007, p. 73.

[258] 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.

[259] R. Danescu, F. Oniga, and S. Nedevschi, "Modeling and tracking the driving environment with a particle-based occupancy grid," *IEEE Trans. Intell. Transp.*, vol. 12, no. 4, pp. 1331–1342, Jan. 2011.

[260] F. Erbs, A. Barth, and U. Franke, "Moving vehicle detection by optimal segmentation of the dynamic stixel world," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 951–956.

[261] 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, Mar. 2012.

[262] P. Chang, D. Hirvonen, T. Camus, and B. Southall, "Stereo-based object detection, classi-cation, and quantitative evaluation with automotive applications," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, Sep. 2005, p. 62.

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

[264] I. Cabani, G. Toulminet, and A. Bensrhair, "Contrast-invariant obstacle detection system using color stereo vision," in *Proc. 11th Int. IEEE Conf. Intell. Transp. Syst.*, Oct. 2008, pp. 1032–1037.

[265] A. Broggi *et al.*, "Terramax vision at the urban challenge 2007," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 1, pp. 194–205, Mar. 2010.

[266] 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. Vehicles Symp.*, Jun. 2005, pp. 199–204.

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

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

[269] S. Sridhar and A. Eskandarian, "Visual object tracking on the inverse perspective map for autonomous vehicles," in *Proc. ASME Dyn. Syst. Control Conf.*, New York, NY, USA: American Society of Mechanical Engineers Digital Collection, 2017, Art. no. V002T17A010.

[270] 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.

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

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

[273] Y.-C. Lim, C.-H. Lee, S. Kwon, and J. Kim, "Event-driven track management method for robust multi-vehicle tracking," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 189–194.

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

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

[276] J. D. Alonso, E. R. 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, May 2008.

[277] 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.

[278] A. Geiger and B. Kitt, "Object flow: A descriptor for classifying traffic motion," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2010, pp. 287–293.

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

[280] S. Sivaraman, B. Morris, and M. Trivedi, "Learning multi-lane trajectories using vehicle-based vision," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 2070–2076.

[281] (2001). *Caltech Computational Vision Caltech Cars*. [Online]. Available: http://www.vision.caltech.edu/html-files/archive.html

[282] (1999). *Caltech Computational Vision Caltech Cars*. [Online]. Available: http://www.vision.caltech.edu/html-files/archive.html

[283] (2001). *Performance Evaluation of Tracking and Surveillance, Pets*. [Online]. Available: http://www.cvg.cs.rdg.ac.U.K./PETS2001/pets2001-dataset.html

[284] C. Caraffi, T. Vojir, 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. Intell. Transp. Syst.*, Sep. 2012, pp. 975–982.

[285] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highD dataset: A drone dataset of naturalistic vehicle trajectories on German highways for validation of highly automated driving systems," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2118–2125.

[286] J. Colyar and J. Halkias, "Interstate 80 freeway dataset," Federal Highway Administration (FHWA), Washington, DC, USA, Tech. Rep. FHWA-HRT-06-137, 2006. [Online]. Available: https://www.fhwa.dot.gov/publications/research/operations/06137/06137.pdf

[287] J. Colyar and J. Halkias, "Us highway 101 dataset," Federal Highway Administration (FHWA), Washington, DC, USA, Tech. Rep. FHWA-HRT-07-030, 2007.

[288] R. Rajamani, *Vehicle Dynamics and Control*. Cham, Switzerland: Springer, 2011.

[289] M. Brännström, E. Coelingh, and J. Sjöberg, "Model-based threat assessment for avoiding arbitrary vehicle collisions," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 658–669, Sep. 2010.

[290] C.-F. Lin, A. G. Ulsoy, and D. J. LeBlanc, "Vehicle dynamics and external disturbance estimation for vehicle path prediction," *IEEE Trans. Control Syst. Technol.*, vol. 8, no. 3, pp. 508–518, May 2000.

[291] J. Huang and H.-S. Tan, "Vehicle future trajectory prediction with a DGPS/INS-based positioning system," in *Proc. Amer. Control Conf.*, 2006, p. 6.

[292] R. Pepy, A. Lambert, and H. Mounier, "Reducing navigation errors by planning with realistic vehicle model," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2006, pp. 300–307.

[293] A. Eidehall and L. Petersson, "Statistical threat assessment for general road scenes using Monte Carlo sampling," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 1, pp. 137–147, Mar. 2008.

[294] N. Kaempchen, B. Schiele, and K. Dietmayer, "Situation assessment of an autonomous emergency brake for arbitrary vehicle-to-vehicle collision scenarios," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 4, pp. 678–687, Dec. 2009.

[295] N. Brouwer, H. Kloeden, and C. Stiller, "Comparison and evaluation of pedestrian motion models for vehicle safety systems," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 1–6.

[296] S. Ammoun and F. Nashashibi, "Real time trajectory prediction for collision risk estimation between vehicles," in *Proc. IEEE 5th Int. Conf. Intell. Comput. Commun. Process.*, Aug. 2009, pp. 417–422.

[297] J. Hillenbrand, A. M. Spieker, and K. Kroschel, "A multilevel collision mitigation approach—Its situation assessment, decision making, and performance tradeoffs," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 4, pp. 528–540, Dec. 2006.

[298] A. Polychronopoulos, M. Tsogas, A. J. Amditis, and L. Andreone, "Sensor fusion for predicting vehicles' path for collision avoidance systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 3, pp. 549–562, Sep. 2007.

[299] R. Miller and Q. Huang, "An adaptive peer-to-peer collision warning system," in *Proc. Veh. Technol. Conf. IEEE 55th Veh. Technol. Conf. VTC Spring*, May 2002, pp. 317–321.

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

[301] H.-S. Tan and J. Huang, "DGPS-based vehicle-to-vehicle cooperative collision warning: Engineering feasibility viewpoints," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 4, pp. 415–428, Dec. 2006.

[302] T. Batz, K. Watson, and J. Beyerer, "Recognition of dangerous situations within a cooperative group of vehicles," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2009, pp. 907–912.

[303] P. Lytrivis, G. Thomaidis, and A. Amditis, "Cooperative path prediction in vehicular environments," in *Proc. 11th Int. IEEE Conf. Intell. Transp. Syst.*, Oct. 2008, pp. 803–808.

[304] K. P. Murphy, "Dynamic Bayesian networks: Representation, inference, and learning," Ph.D. dissertation, Dept. Comput. Sci., Graduate Division, Univ. California, Berkeley, Berkeley, CA, USA, Fall 2002. [Online]. Available: https://ibug.doc.ic.ac.uk/media/uploads/documents/courses/DBN-PhDthesis-LongTutorail-Murphy.pdf

[305] H. Veeraraghavan, N. Papanikolopoulos, and P. Schrater, "Deterministic sampling-based switching Kalman filtering for vehicle tracking," in *Proc. IEEE Intell. Transp. Syst. Conf.*, Sep. 2006, pp. 1340–1345.

[306] H. Dyckmanns, R. Matthaei, M. Maurer, B. Lichte, J. Effertz, and D. Stiker, "Object tracking in urban intersections based on active use of *a priori* knowledge: Active interacting multi model filter," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 625–630.

[307] A. Broadhurst, S. Baker, and T. Kanade, "Monte Carlo road safety reasoning," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2005, pp. 319–324.

[308] M. Althoff and A. Mergel, "Comparison of Markov chain abstraction and Monte Carlo simulation for the safety assessment of autonomous cars," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1237–1247, Dec. 2011.

[309] S. Atev, G. Miller, and N. P. Papanikolopoulos, "Clustering of vehicle trajectories," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 647–657, Sep. 2010.

[310] T. Christopher, "Analysis of dynamic scenes: Application to driving assistance," Ph.D. dissertation, Inst. Polytechn. de Grenoble, Grenoble, France, Sep. 2009. [Online]. Available: https://tel.archives-ouvertes.fr/tel-00530679/document

[311] D. Vasquez and T. Fraichard, "Motion prediction for moving objects: A statistical approach," in *Proc. Intl. Conf. Robot. Autom.*, vol. 4, Apr./May 2004, pp. 3931–3936.

[312] C. Hermes, C. Wohler, K. Schenk, and F. Kummert, "Long-term vehicle motion prediction," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2009, pp. 652–657.

[313] D. Vasquez, T. Fraichard, and C. Laugier, "Growing hidden Markov models: An incremental tool for learning and predicting human and vehicle motion," *Int. J. Robot. Res.*, vol. 28, nos. 11–12, pp. 1486–1506, 2009.

[314] J. Joseph, F. Doshi-Velez, A. S. Huang, and N. Roy, "A Bayesian nonparametric approach to modeling motion patterns," *Auton. Robot.*, vol. 31, no. 4, p. 383, Nov. 2011.

[315] Q. Tran and J. Firl, "Online maneuver recognition and multimodal trajectory prediction for intersection assistance using nonparametric regression," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2014, pp. 918–923.

[316] G. S. Aoude, V. R. Desaraju, L. H. Stephens, and J. P. How, "Driver behavior classification at intersections and validation on large naturalistic data set," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 724–736, Jun. 2012.

[317] I. Dagli and D. Reichardt, "Motivation-based approach to behavior prediction," in *Proc. Intell. Vehicle Symp.*, Jun. 2002, pp. 227–233.

[318] M. G. Ortiz, J. Fritsch, F. Kummert, and A. Gepperth, "Behavior prediction at multiple time-scales in inner-city scenarios," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 1068–1073.

[319] W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, and S. Maybank, "A system for learning statistical motion patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1450–1464, Sep. 2006.

[320] D. Buzan, S. Sclaroff, and G. Kollios, "Extraction and clustering of motion trajectories in video," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2004, pp. 521–524.

[321] J. Wiest, F. Kunz, U. Kressel, and K. Dietmayer, "Incorporating categorical information for enhanced probabilistic trajectory prediction," in *Proc. 12th Int. Conf. Mach. Learn. Appl. (ICMLA)*, vol. 1, Dec. 2013, pp. 402–407.

[322] D. Greene *et al.*, "An efficient computational architecture for a collision early-warning system for vehicles, pedestrians, and bicyclists," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 942–953, Dec. 2011.

[323] S. Klingelschmitt, M. Platho, H.-M. Gros, V. Willert, and J. Eggert, "Combining behavior and situation information for reliably estimating multiple intentions," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2014, pp. 388–393.

[324] B. Morris, A. Doshi, and M. Trivedi, "Lane change intent prediction for driver assistance: On-road design and evaluation," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 895–901.

[325] P. Kumar, M. Perrollaz, S. Lefevre, and C. Laugier, "Learning-based approach for online lane change intention prediction," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2013, pp. 797–802.

[326] H. M. Mandalia and M. D. D. Salvucci, "Using support vector machines for lane-change detection," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, vol. 49, no. 22. Newbury Park, CA, USA: Sage, 2005, pp. 1965–1969.

[327] H. Berndt, J. Emmert, and K. Dietmayer, "Continuous driver intention recognition with hidden Markov models," in *Proc. 11th Int. IEEE Conf. Intell. Transp. Syst.*, Oct. 2008, pp. 1189–1194.

[328] T. Streubel and K. H. Hoffmann, "Prediction of driver intended path at intersections," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2014, pp. 134–139.

[329] S. Lefevre, Y. Gao, D. Vasquez, H. E. Tseng, R. Bajcsy, and F. Borrelli, "Lane keeping assistance with learning-based driver model and model predictive control," in *Proc. 12th Int. Symp. Adv. Vehicle Control*, 2014, pp. 1–8.

[330] A. Tamke, T. Dang, and G. Breuel, "A flexible method for criticality assessment in driver assistance systems," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 697–702.

[331] C. Laugier *et al.*, "Probabilistic analysis of dynamic scenes and collision risks assessment to improve driving safety," *IEEE Intell. Transp. Syst. Mag.*, vol. 3, no. 4, pp. 4–19, Oct. 2011.

[332] M. Althoff, O. Stursberg, and M. Buss, "Model-based probabilistic collision detection in autonomous driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 2, pp. 299–310, Jun. 2009.

[333] A. Lawitzky, D. Althoff, C. F. Passenberg, G. Tanzmeister, D. Wollherr, and M. Buss, "Interactive scene prediction for automotive applications," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2013, pp. 1028–1033.

[334] M. Brand, N. Oliver, and A. Pentland, "Coupled hidden Markov models for complex action recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1997, pp. 994–999.

[335] N. Oliver and A. P. Pentland, "Graphical models for driver behavior recognition in a SmartCar," in *Proc. IEEE Intell. Vehicles Symp.*, Oct. 2000, pp. 7–12.

[336] M. Liebner, M. Baumann, F. Klanner, and C. Stiller, "Driver intent inference at urban intersections using the intelligent driver model," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2012, pp. 1162–1167.

[337] G. Agamennoni, J. I. Nieto, and E. M. Nebot, "A Bayesian approach for driving behavior inference," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 595–600.

[338] S. Lefevre, C. Laugier, and J. Ibanez-Guzman, "Risk assessment at road intersections: Comparing intention and expectation," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2012, pp. 165–171.

[339] S. Lefevre, C. Laugier, and J. Ibanez-Guzman, "Evaluating risk at road intersections by detecting conflicting intentions," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2012, pp. 4841–4846.

[340] I. Rasheed, F. Hu, Y.-K. Hong, and B. Balasubramanian, "Intelligent vehicle network routing with adaptive 3D beam alignment for mmWave 5G-based V2X communications," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 5, pp. 2706–2718, May 2020.

[341] F. A. Schiegg, I. Llatser, D. Bischoff, and G. Volk, "Collective perception: A safety perspective," *Sensors*, vol. 21, no. 1, p. 159, Dec. 2020.

[342] M. Goli and A. Eskandarian, "Merging strategies, trajectory planning and controls for platoon of connected, and autonomous vehicles," *Int. J. Intell. Transp. Syst. Res.*, vol. 18, no. 1, pp. 153–173, Jan. 2020.

[343] M. Goli and A. Eskandarian, "MPC-based lateral controller with look-ahead design for autonomous multi-vehicle merging into platoon," in *Proc. Amer. Control Conf. (ACC)*, Jul. 2019, pp. 5284–5291.

[344] M. Goli and A. Eskandarian, "Evaluation of a multi-vehicle merging strategy under different lateral maneuvers in the presence of sudden braking," in *Proc. Dyn. Syst. Control Conf.*, vol. 57267, 2015, Art. no. V003T50A011.

[345] M. Goli and A. Eskandarian, "A systematic multi-vehicle platooning and platoon merging: Strategy, control, and trajectory generation," in *Proc. ASME Dyn. Syst. Control Conf.*, vol. 46193. New York, NY, USA: American Society of Mechanical Engineers Digital Collection, 2014, p. V002T25A006. [Online]. Available: https://asmedigitalcollection.asme.org/DSCC/proceedings/DSCC2014/46193/V002T25A006/228914

[346] M. Goli and A. Eskandarian, "Mobile robot coordinated platooning: A small-scale experimental evaluation to emulate connected vehicles," in *Proc. Dyn., Vibrat., Control*, vol. 4, Nov. 2015, Art. no. V04AT04A004.

[347] M. Goli and A. Eskandarian, "The effect of information, and communication topologies on input-to-state stability of platoon," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 540–544.

[348] P. Lv, Y. He, J. Han, and J. Xu, "Objects perceptibility prediction model based on machine learning for V2I communication load reduction," in *Proc. Int. Conf. Wireless Algorithms, Syst., Appl.*, Cham, Switzerland: Springer, Jun. 2021, pp. 521–528.

[349] S. Sridhar and A. Eskandarian, "Cooperative perception in autonomous ground vehicles using a mobile-robot testbed," *IET Intell. Transp. Syst.*, vol. 13, no. 10, pp. 1545–1556, Oct. 2019.

[350] S.-W. Kim, W. Liu, M. H. Ang, E. Frazzoli, and D. Rus, "The impact of cooperative perception on decision making and planning of autonomous vehicles," *IEEE Intell. Transp. Syst. Mag.*, vol. 7, no. 3, pp. 39–50, Jul. 2015.

[351] P. Ghorai and A. Eskandarian, "Longitudinal control algorithm for cooperative autonomous vehicles to avoid accident with vulnerable road users," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2020, pp. 1–6.

[352] *Traffic Safety Basic Facts on Junctions*, European Commission-Directorate General for Transport, Brussels, Belgium, 2016.

[353] J. Rios-Torres and A. A. Malikopoulos, "A survey on the coordination of connected and automated vehicles at intersections and merging at highway on-ramps," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 5, pp. 1066–1077, May 2017.

[354] L. Chen and C. Englund, "Cooperative intersection management: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 2, pp. 570–586, Feb. 2016.

[355] J. B. Collins and J. K. Uhlmann, "Efficient gating in data association with multivariate Gaussian distributed states," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 3, pp. 909–916, Jul. 1992.

[356] P. Ghorai, A. Eskandarian, and Y.-K. Kim, "Study the effect of communication delay for perception and collision avoidance in cooperative autonomous driving," in *Proc. Dyn., Vibrat., Control*, vol. 7, Nov. 2020, Art. no. V07BT07A015.

[357] A. M. H. Al-Jhayyish and A. W. Schmidt, "Feedforward strategies for cooperative adaptive cruise control in heterogeneous vehicle strings," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 113–122, Jan. 2018.

[358] L. Xiao and F. Gao, "Practical string stability of platoon of adaptive cruise control vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1184–1194, Dec. 2011.

[359] C. Wu, Y. Lin, and A. Eskandarian, "Cooperative adaptive cruise control with adaptive Kalman filter subject to temporary communication loss," *IEEE Access*, vol. 7, pp. 93558–93568, 2019.

[360] P. Zhou *et al.*, "AICP: Augmented informative cooperative perception," 2021, *arXiv:2101.05508*.

Prasenjit Ghorai received the B.Tech. degree in electronics and instrumentation engineering from the Maulana Abul Kalam Azad University of Technology (formerly West Bengal University of Technology), India, the M.Tech. degree in control and instrumentation engineering from the University of Calcutta, and the Ph.D. degree in engineering from the National Institute of Technology (NIT) Agartala (in collaboration with Indian Institute of Technology, Guwahati, India). He was an Assistant Professor of electronics and instrumentation engineering with NIT Agartala from 2011 to 2019. He is currently working as a Post-Doctoral Associate with the Autonomous Systems and Intelligent Machines Laboratory, Virginia Tech, where he conducts the research on cooperative and connected autonomous vehicles.

Azim Eskandarian (Senior Member, IEEE) has been a Professor and the Head of the Mechanical Engineering Department, Virginia Tech, since August 2015. He became the Nicholas and Rebecca Des Champs Chaired Professor in April 2018. He has been a Professor with the Electrical and Computer Engineering Department since 2021. He established the Autonomous Systems and Intelligent Machines laboratory, Virginia Tech, and has conducted pioneering researches on autonomous vehicles, human/driver cognition and vehicle interface, advanced driver assistance systems, and robotics. Before joining Virginia Tech, he was a Professor of engineering and applied science at George Washington University (GWU) and the Founding Director of the Center for Intelligent Systems Research from 1996 to 2015, the Director of the Transportation Safety and Security University Area of Excellence from 2002 to 2015, and the Co-Founder of the National Crash Analysis Center in 1992 and its Director from 1998 to 2002 and from 2013 to 2015. From 1989 to 1992, he was an Assistant Professor at Pennsylvania State University, York, PA, USA, and an Engineer/a Project Manager in the industry from 1983 to 1989. He is a fellow of ASME and a member of SAE. He received the SAE's Vincent 2021 Bendix Automotive Electronics Engineering Award, the IEEE ITS Society's Outstanding Researcher Award in 2017, and GWU's School of Engineering Outstanding Researcher Award in 2013. (For more information see https://www.asim.me.vt.edu)

Young-Keun Kim received the B.E. degree from Handong University, Pohang, South Korea, in 2008, and the M.Sc.Eng. and Ph.D. degrees in mechanical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2014. Since 2014, he has been with the Department of Mechanical and Control Engineering, Handong Global University. He is currently an Associate Professor and the Head of the Smart Sensor Systems Laboratory, to conduct researches on intelligent machine vision for automated systems, deep learning for perception, and LiDAR signal processing.

Goodarz Mehr received the B.Sc. degree in mechanical engineering from the Sharif University of Technology, Tehran, Iran. He is currently pursuing the Ph.D. degree in mechanical engineering from Virginia Tech. His research interests include robotics and control, stochastic planning models, machine learning, and cooperative perception.