Looking at Humans in the Age of Self-Driving and Highly Automated Vehicles

Eshed Ohn-Bar and Mohan Manubhai Trivedi[1]

***Abstract*—This paper highlights the role of humans in the next generation of driver assistance and intelligent vehicles. Understanding, modeling, and predicting human agents are discussed in three domains where humans and highly automated or self-driving vehicles interact: 1) inside the vehicle cabin, 2) around the vehicle, and 3) inside surrounding vehicles. Efforts within each domain, integrative frameworks across domains, and scientific tools required for future developments are discussed to provide a human-centered perspective on research in intelligent vehicles.**

***Index Terms*—Intelligent vehicles, human intent and behavior analysis, human-robot interaction, driver assistance, highly autonomous vehicles, vehicle-driver hand-off, risk forecasting, pedestrian/vehicle tracking, cognitive engineering.**

I. Introduction

THERE is an unprecedented interest, activity, and excitement in the field of intelligent vehicles. In a great technological milestone, the culmination of research efforts of the past decades in a broad range of disciplines, including vehicle control, robotics, sensing, machine perception, navigation, mapping, machine learning, embedded systems, human-machine interactivity, and human factors, has realized practical and affordable systems for various automated features in automobiles [114]. This advancement is opening doors to possibilities only thought to be fictional a few decades ago. The aim of this work is to recognize the next set of research challenges required to be addressed for achieving highly reliable, fail-safe, intelligent vehicles which can earn the trust of humans who would ultimately purchase and use these vehicles.

It is clear that automobile industry has made a firm commitment to support developments towards what can be seen as "disruptive" transformation of automobiles driven by human drivers to intelligent robots who transport humans on the roads. What will then be the role of humans in such a rapidly approaching future? Would they seat as passive occupants, who fully trust their vehicles? Would there be a need for humans to "take over" control in some situations either triggered by the need perceived by the autonomous vehicle or desired by someone in the cabin? How should these autonomous vehicles interact with humans outside the vehicle (either as drivers of non-autonomous vehicles, pedestrians, emergency workers, etc.)? Because the future of intelligent vehicles lies in the collaboration of two intelligent systems, one robot and another human, this study aims to present core research ideas as they relate to humans in and around vehicles. In this collaboration of human and robot, the need for intelligent vehicles to observe, understand, model, infer and anticipate human behavior is necessary now more than ever.

This paper follows three main domains where humans and highly automated or self-driving vehicles interact (illustrated in Fig. 1):

- **Humans in vehicle cabin:** Whether the humans in the vehicle cabin are active drivers, passengers, or passive drivers, they may still be required to "take over" control in some situations triggered by the perceived need of the autonomous vehicle (for instance, under rare situations such as construction zones or police controlled intersections). In such situations, looking at the humans inside the vehicle cabin is necessary to access readiness to take over. If active drivers, are they distracted, did they pay attention to objects of interest (e.g. traffic signs, pedestrians), are they fatigued? If passengers, are they sitting properly (e.g. for proper airbag deployment in case of emergency), are they giving directions, are they distracting the driver? If passive drivers, in the case of automated vehicles requiring take over at crucial moments, are they engaged in a secondary task, are their hands free, have they been alert to the changing driving environment?
- **Humans around the vehicle:** In addition to monitoring humans inside the vehicle cabin, observing humans in the vicinity of the intelligent vehicles is also essential for safe and smooth navigation. Because the road is shared with pedestrians, both an automobile driven by humans or intelligent robots who transport humans must be able to sense pedestrian intent and communicate with pedestrians. Where and how are humans around vehicle interacting with the vehicle? These include pedestrians, bike riders, skate boarders, traffic controllers, construction workers, emergency responders, etc. Are they in the path of the vehicle? Are they communicating their intent via body gestures? Are they distracted? Addressing such research issues can result in improved quality of navigation and assistance.
- **Humans in surrounding vehicles:** Intelligent vehicles must take into consideration humans in surrounding vehicles. Activity analysis and observation of intent applies to such humans as well, which operate under specific experience level, aggressiveness, style, age, distraction-level, etc. For instance, imagine two intelligent vehicles arriving at a stop-controlled intersection. In such a situation, both vehicles may be fully autonomous, only one

[1]Laboratory for Intelligent and Safe Automobiles, University of California San Diego, La Jolla CA 92092, USA {eohnbar, mtrivedi}@ucsd.edu

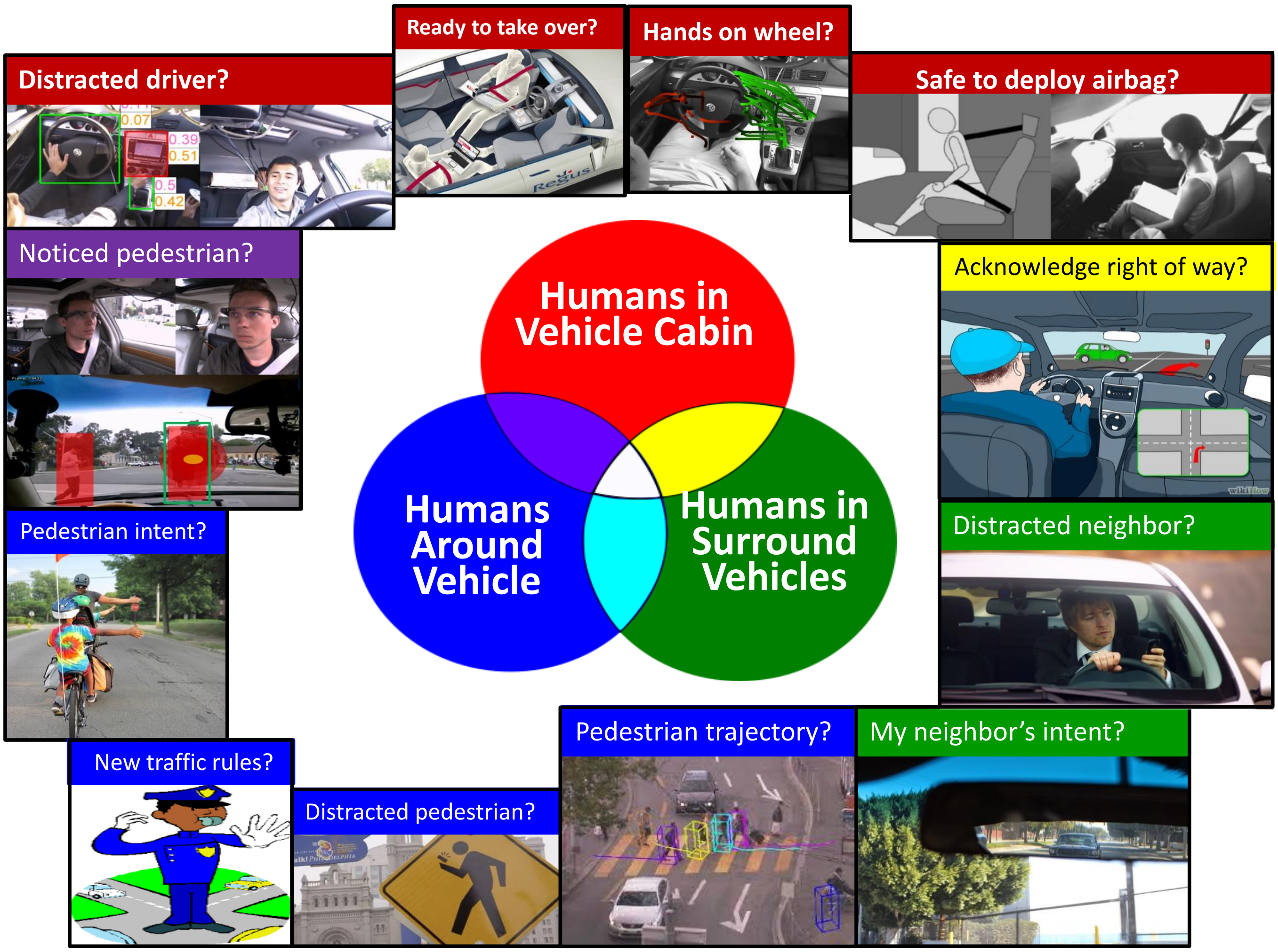


Fig. 1: Intricate roles of humans to be considered in the development of highly automated and self-driving vehicles. For a safe and comfortable ride, intelligent vehicles must observe, understand, model, infer, and predict behavior of occupants inside the vehicle cabin, pedestrians around the vehicle, and humans in surrounding vehicles.

of the vehicles may be fully autonomous, or both may be human-operated. Observing the humans by direct or indirect observation is necessary to acknowledge or give right of way. Are the humans in other vehicles driving in a risky manner? Is their behavior normal or abnormal? What will they do next, and what general and user-specific cues can be leveraged towards this identification? Are they acknowledging right of way at stop-controlled intersection? Are they engaged in secondary tasks, which motivates the ego-vehicle to avoid its vicinity?

We continue by providing an overview of relevant research studies. The studies are categorized in Section II for providing a highlight of the current research landscape. Section II studies emerging research topics in vision-based intelligent vehicles for each of the domains where humans and highly automated or self-driving vehicles interact. Section III follows with an analysis of the publicly available vision tools required for addressing the highlighted research issues. Finally, summary and conclusions are provided in Section III.

II. Looking at Humans in and Around the Vehicle: Research Landscape and Accomplishments

The study of human-centric cues for driver assistance is an active research topic in intelligent vehicles, machine learning, and computer vision. Therefore, an extensive amount of work has been done in the field, from analysis of driver goals and intentions, human-machine interface design and customization, pedestrian activity classification, and up to identification of surrounding aggressive drivers (Fig. 1).

As means of identifying research trends, our first step is to give an overview of selected studies employing computer vision and machine learning techniques for intelligent vehicles applications. In order to maintain focus over the a large research landscape, the following approach for clustering research studies is pursued:

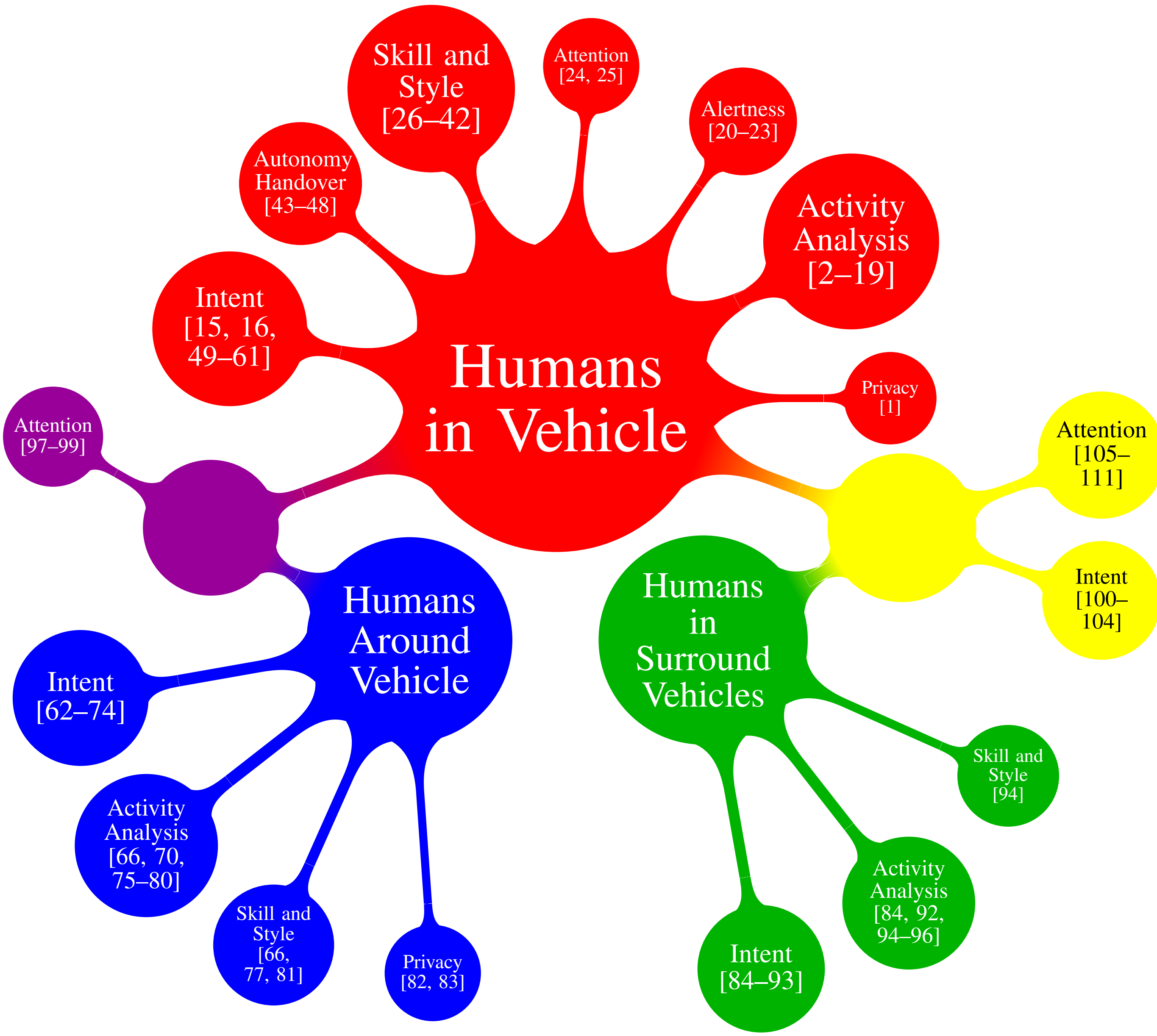


Fig. 2: Trends in human-centric intelligent vehicle research. The figure visualizes related research studies discussed in this paper as they relate to different semantic goals, from maneuver analysis and prediction, to style modeling. Each topic size is proportional the count of studies surveyed it contains.

- **Domain clustering**: Throughout the paper we partition the research space based on the three domains in Fig. 1, of humans inside the vehicle, around, and in surrounding vehicles. Although all three domains share the human agent, the domain-based clustering is useful because studies tend to focus on one of the three domains. From a vision perspective, methodologies and research goals among papers within the same domain tend to be more similar. Domain clustering also allows comparing and contrasting the domains in terms of what has been done and what has yet to be achieved.
- **Research goal clustering**: Related studies generally attempt to analyze, model, classify, and/or predict activities. This suggests a clustering based on the research task, whether humans inside or outside of a vehicle are concerned. We select seven types of overall research goals found in the surveyed studies. This clustering is employed for gaining a deeper understanding of the research landscape and discussing potential future research directions. Research goals include agent intent analysis and activity prediction (what will happen next?), attention model (where and what is the focus of the agent?), skill and style (what type of agent?), alertness and distraction (what is the state of the agent?), and general activity

TABLE I: Overview of human-centric related research studies by research goal and human-centric cues employed. Goal types follow Table II, with [I] - intent and prediction, [Ac] - activity and behavior understanding, [D] - distraction and alertness, [At] - attention, and [S] - skill and style. VD refers to Vehicle Dynamics. PD refers to Pedestrian Dynamics (i.e. position, velocity).

Study	Type	Goal Detail	Cue Type
Jain et al. [51, 112], 2016	I	Lane Change Prediction	Head, Lane, VD, GPS, Map
Tran et al. [15], 2012	I,Ac	Brake	Foot, VD
Lefèvre et al. [49], 2011	I	Intent at Intersections	Map, VD
Molchanov et al. [13, 17], 2015	Ac	Secondary Tasks/Infotainment	Hand, Video
Ohn-Bar et al. [9] [16] 2014	Ac	Secondary Tasks/Infotainment	Head, Hand, Eye, Image
Tawari et al. [14] [25], 2014	Ac,At	Gaze Zone	Head, Eye
Toma et al. [2], 2012	Ac	Secondary Tasks/Phone	Head, Image
Ahlstrom et al. [11], 2012	Ac	Gaze Zone	Head, Eye
Cheng and Trivedi [18], 2010	Ac	Driver/Passenger Classification	Hand, Image
Vicente et al. [24], 2015	At	Gaze Zone	Head, Eye, Image
Liu et al. [23], 2015	D	Distraction Detection	Head, Eye
Jimnez et al. [21], 2012	D	Gaze Zone	Head, Eye
Wllmer et al. [20], 2011	D	Distraction Detection	Head
Lefèvre et al. [30], 2015	S	Style	VD
Schulz et al. [70, 71], 2015	I,Ac	Pedestrian Intent Recognition	PD, Head
Møgelmose et al. [67], 2015	I	Pedestrian Risk Estimation	PD, GPS, Map
Madrigal et al. [65], 2014	I	Intention-Aware Pedestrian Tracking	PD, Social Context
Kooij et al. [73], 2014	I	Pedestrian Path Prediction	PD, Head, Situation Criticality, Scene Layout
Quintero et al. [66], 2014	I,Ac,S	Pedestrian Path Prediction	PD, Body Pose, Subject Style
Goldhammer et al. [63, 77], 2014	I,S	Pedestrian Path and Gait Analysis	PD, Head
Pellegrini et al. [113], 2009	I	Pedestrian Path Prediction	PD, Social Context
Kooij et al. [75], 2016	Ac	Pedestrian Behavior Patterns	PD
Kataoka et al. [79], 2015	Ac	Pedestrian Activity Classification	PD, Video
Choi and Savarese [76], 2014	Ac	Pedestrian Activity Classification	PD, Social Context
Li et al. [84], 2016	I,Ac	Car Fluents	Video, Vehicle Part State
Laugier et al. [91], 2011	I	Behavior and Risk Assessment	VD, Lane, Turn Signal, GPS
Fröhlich et al. [88], 2014	I	Lane Change Intent	Turn Signal
Graf et al. [89], 2014	I	Turn Intent	VD, GPS, Map
Bahram et al. [104], 2016	I	Interaction-Aware Maneuver Prediction	VD, GPS, Map
Ohn-Bar et al. [102], 2015	I	Overtake and Brake Prediction	Head, Hand, Foot, VD, Lane
Jahangiri et al. [85], 2015	I	Intent to Run Redlight	VD, Scene Layout
Gindele et al. [87], 2013	I	Contextual Path Prediction	VD, Map, Lanes
Doshi et al. [101], 2011	I	Lane Change Forecasting	Head, Lane, VD
Aoude et al. [90], 2010	I	Threat Assessment	VD, GPS, Map, Lanes
Tawari et al. [111], 2014	At	Attention and Surround Criticality	Head, VD, Lane
Bar et al. [107], 2013	At	Seen/Missed Objects	Head, Eye, VD, Image
Mori et al. [108], 2012	At	Surround Awareness	Head, Eye, VD
Takagi et al. [110], 2011	At	Gaze Target	Head, Eye, VD
Doshi and Trivedi [105], 2010	At	Attention Focus	Head, Video
Phan et al. [97], 2014	At	Awareness of Pedestrians	VD
Tanishige et al. [98], 2014	At	Pedestrian Detectability	Head, Eye, PD, Video
Tawari et al. [99], 2014	At	Driver and Pedestrian Attention	Head, Eye, PD

Color codes:

Red	Studying humans inside cabin.
Blue	Studying humans around vehicles.
Green	Studying humans in surround vehicles.
Yellow	Studying humans inside cabin and in surround vehicles.
Purple	Studying humans inside and around vehicles.

classification and behavior analysis (how is the agent operating?). Two additional goals not falling into the previous categories are autonomy handover and privacy-related tasks. We emphasize that the chosen research goals are closely related to each other and that there are other potential choices for research goal clustering [117]. Depending on the study, it may fall into one or multiple of the research goals. The research goals are consistent with topics in machine vision and learning-based studies as related to the type of data, methodologies, and metrics employed.

- **Cue type analysis**: A third type of analysis for highlighting trends in related studies can be made based on the type of cues employed in the study. We make a distinction between studies employing direct human-observing cues (e.g. body pose) and indirect cues (e.g. vehicle dynamics, GPS). This is shown in Table II. Furthermore, we detail the specific type of cues employed by selected studies in Table I, which complements the other two clustering techniques described above.

Fig. 2 shows a domain-based and research goal-based clustering of the papers listed in the corresponding Table II. An emphasis is put on recent studies (mostly after 2008). In Fig. 2, the size of the node is proportional to the number of studies it contains. Fig. 2 can be used to draw several conclusions. We first identify trends, and then discuss further detail of the

TABLE II: Overview of selected studies discussing different aspects of humans on the road. Methods are categorized according to task and whether humans were observed directly (e.g. body pose cues) or indirectly (e.g. pedal press, GPS/Map, vehicle trajectory).

Goal	Direct	Indirect
Intent and Prediction		
- In Vehicle	[15, 16, 50, 51]	[49, 52–61]
- Around Vehicle	[62–74]	-
- Surrounding Vehicles	-	[84–91, 93]
- In+Surrounding Vehicles	[100–103]	[92, 104]
Activity		
- In Vehicle	[2–4, 9–19, 44, 48, 115]	[5–8, 28]
- Around Vehicle	[66, 70, 75–80]	-
- In Surrounding Vehicles	-	[84, 92, 95, 96]
Distraction and Alertness		
- In Vehicle	[20–23]	-
Attention		
- In Vehicle	[24, 25]	-
- In+Around Vehicle	[97–99]	-
- In+Surrounding Vehicles	[105–108, 110, 111]	-
Skill and Style		
- In Vehicle	[27]	[26, 28–42, 116]
- Around Vehicle	[66, 77, 81]	-
- In Surrounding Vehicles	-	[94]

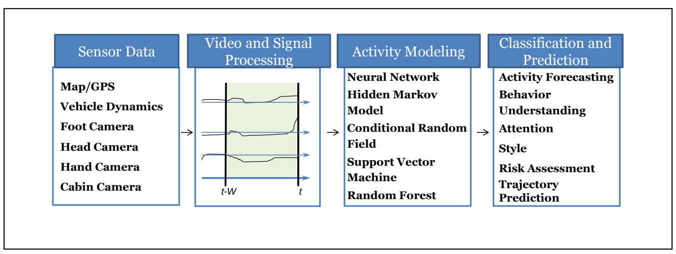


Fig. 3: Overview of the sensing and learning pipeline commonly used to study humans in the cabin.

studies in each domain in the following sections (Section II-A, II-B, II-C).

As might be expected, a large number of human-centric studies emphasize humans inside the vehicle. This domain also contains most of the diversity in terms of research goals, but research efforts are not distributed equally. A large number of behavior and activity analysis studies on driver gestures, secondary tasks, distraction, and maneuver classification and prediction have been performed. In-vehicle study of activities allows for a fine sensor resolution of the human agent, from vehicle dynamic sensors and up to eye and gaze analysis. The studies in this cluster still vary drastically in terms of the type of cues and vision techniques employed, as shown in Table I. Certain research tasks, such as skill and style of humans, in-vehicle occupant interaction, and activity analysis of passengers, has seen less attention.

Fig. 2 allows for a high-level comparison between the domain of looking at humans inside the vehicle and the other two

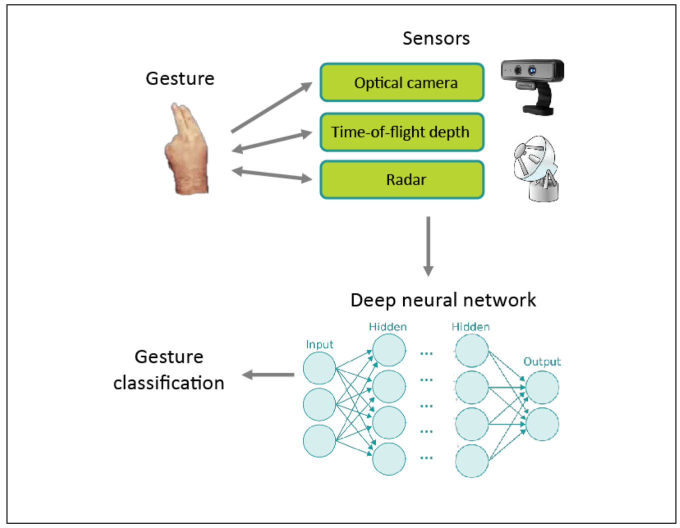
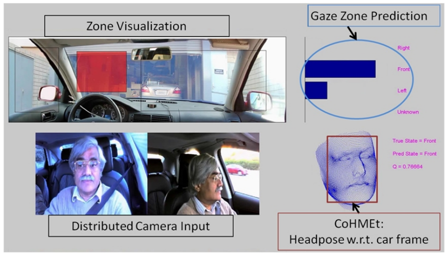


Fig. 4: A multi-sensor driver gesture recognition system with a deep neural network [13].

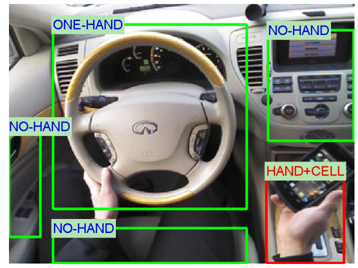
domains. Although human drivers can analyze fine-grained pose, style, and activity cues for identification of agent intent in all three domains (see Fig. 1), fine-grained semantic analysis around and in surrounding vehicles is still in early stages. Looking at humans around the vehicle commonly involves path prediction and to a lesser extent activity classification. Trajectory level path prediction is often done with little notion of skill, style, social cues, or distraction. Future improvement in camera and sensing modalities would provide access to better and larger datasets. Consequently, we expect research tasks in the less studied two domains to become more diverse as in the looking inside the vehicle domain. Direct observation of humans in surrounding vehicles has not been done, although humans employ it everyday on the road.

Another main conclusion that can be drawn relates to integrative schemes, which are also shown to be studied to a lesser extent. The studies are limited to attention-related studies as these reason over objects around the vehicle in order to infer surround awareness and gaze target. On the road, holistic understanding of both humans inside, around the ego-vehicle, and in surround vehicles is essential for effective driver assistance and higher vehicle autonomy. Holistic understanding of all three domains is a task performed by everyday human drivers while inferring intents, analyzing potential risk, and smoothly navigating a vehicle [119, 120]. Another relevant research topic is the modeling of social relationships among agents, which are employed by drivers in order to recognize and communicate intents. More specific examples can be found in Section II-D.

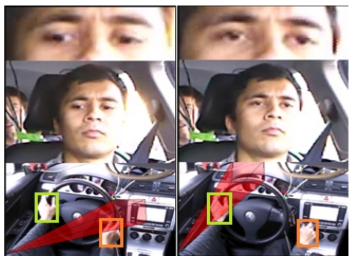
Fig. 2 and Table II provide a high-level analysis of trends in related research studies within domains and research goals. Certain research goals are shown to be highly represented in one domain, but almost none existent in another. Nonetheless, even within a certain domain of human study, large variations



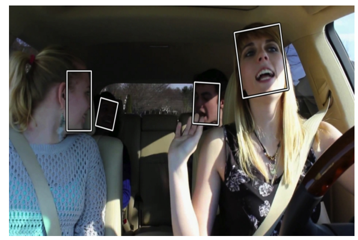
(a) Gaze zone classification using head cues [25].



(b) Object interaction analysis
and secondary task classification with hand cues [118].



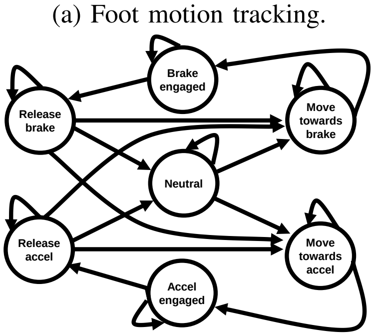
(c) Head, hand, and eye cue integration
for secondary task activity analysis [9, 115].



(d) Cabin occupant activity and interaction analysis.

Fig. 5: Emerging research topics for studying humans inside the vehicle.

(a) Foot motion tracking.



(b) Activity state model.

Fig. 6: Foot gesture recognition and prediction using a motion tracker and a temporal state model, such as a Hidden Markov Model [15].

exist in the types of cues employed for a specific task. Table I provides a closer look to the type of human-observing cues employed in the surveyed studies.

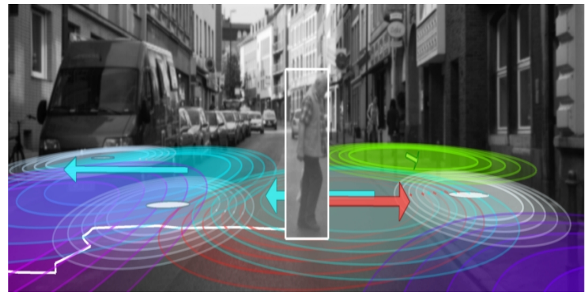
Next, we provide a deeper discussion for each domain as well as integrative frameworks below.

A. *Looking at Humans in the Cabin*

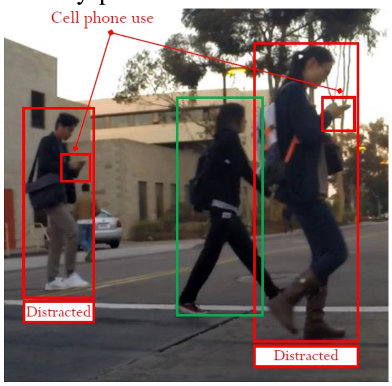
The surveyed papers in Fig. 2 show large diversity in terms of the research tasks for studying humans inside the vehicle. Further detail is provided in Table I in terms of study details and cue analyzed. A highlight of the research tasks is shown in Fig. 5, with an example research pipeline in Figs. 3 and 4. Dynamics of driver body pose, such as head [25], hand [16], eye [21], and foot [15] (Fig. 6) can be employed for in-cabin analysis of secondary tasks [2, 9, 11, 14, 24, 123, 124] and intent modeling and maneuver prediction [16, 49–51, 103]. Certain types of secondary tasks, such as gaze zone estimation and head gesture analysis, are more commonly studied than others, such as driver-object interaction (e.g. infotainment analysis [9] and cell-phone use [2]). Although passenger-related secondary tasks were shown to be critical for driver state monitoring from naturalistic driving studies [125], there are very few vision and learning studies on such tasks. Driver and passenger hand gesture and user identification have been studied in [18, 126, 127], but a large number of research tasks relating to interaction activity analysis has not been pursued. Fig. 5 highlights the need for the understanding and integration of multiple cues at different levels of representation. Such holistic modeling is essential for accurate, robust, and natural human-machine interaction. In particular, for studying humans in the cabin under semi-autonomy and control hand off [43, 45–47]. Depth sensors may also be used for improved activity recognition [118, 128–130].

Looking inside the vehicle often involves multiple types

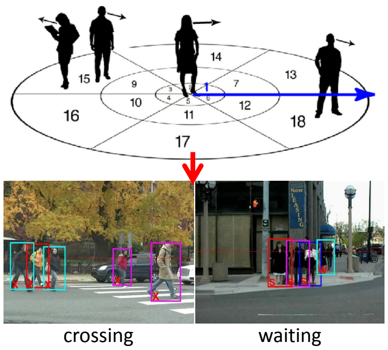
(a) Pedestrian body pose and attribute classification [121].



(b) Pedestrian path and intent prediction [72, 74].



(c) Fine-grained pedestrian activity classification.



(d) Social relationships modeling [76, 78, 122].

Fig. 7: Emerging research topics for studying people around the vehicle.

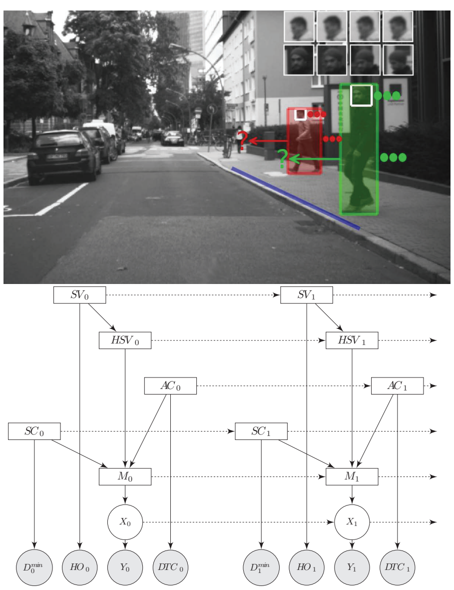


Fig. 8: Pedestrian path prediction using a Dynamic Bayesian Network for incorporating contextual cues of pedestrian head orientation and situational awareness, situation criticality, and spatial layout cues [73].

of on-board sensors in addition to a camera, such as vehicle dynamics [5–7, 31–33], phone [8, 29, 34–41], or GPS [26, 28, 52, 53, 55–59]. These provide another useful modality for analyzing the behavior of humans inside the vehicle, such as skill and style recognition from inertial sensors [28]. Velocity, yaw-rate, and other vehicle parameters provide a signal useful for intent and maneuver recognition [52, 53, 56, 57]. GPS and map data can provide scene context (e.g. intersection vs. highway), strategic maneuver analysis [131, 132], or be used in tactic and operation prediction models [52, 133]. In Liebner *et al.* [52] turn and stop maneuvers at intersections are predicted using GPS trajectories and a Bayesian Network for modeling driver intent.

B. Looking at Humans Around the Vehicle

Humans around the vehicle can be sensed with a variety of vision sensors, including color, thermal, and range sensors. Table I demonstrates a variety of research goals and cues employed to study pedestrians, with a highlight of research tasks shown in Fig. 7. The task of analyzing surround pedestrians is related to the heavily-studied visual surveillance tasks of scene and activity modeling [122]. In this work, we emphasize studies performed from movable platforms and leverage the specific geometrical and contextual cues induced by on-road settings. Here, scene information such as lane and road information can be combined with pedestrian detection and tracking for performing intent-aware path prediction and

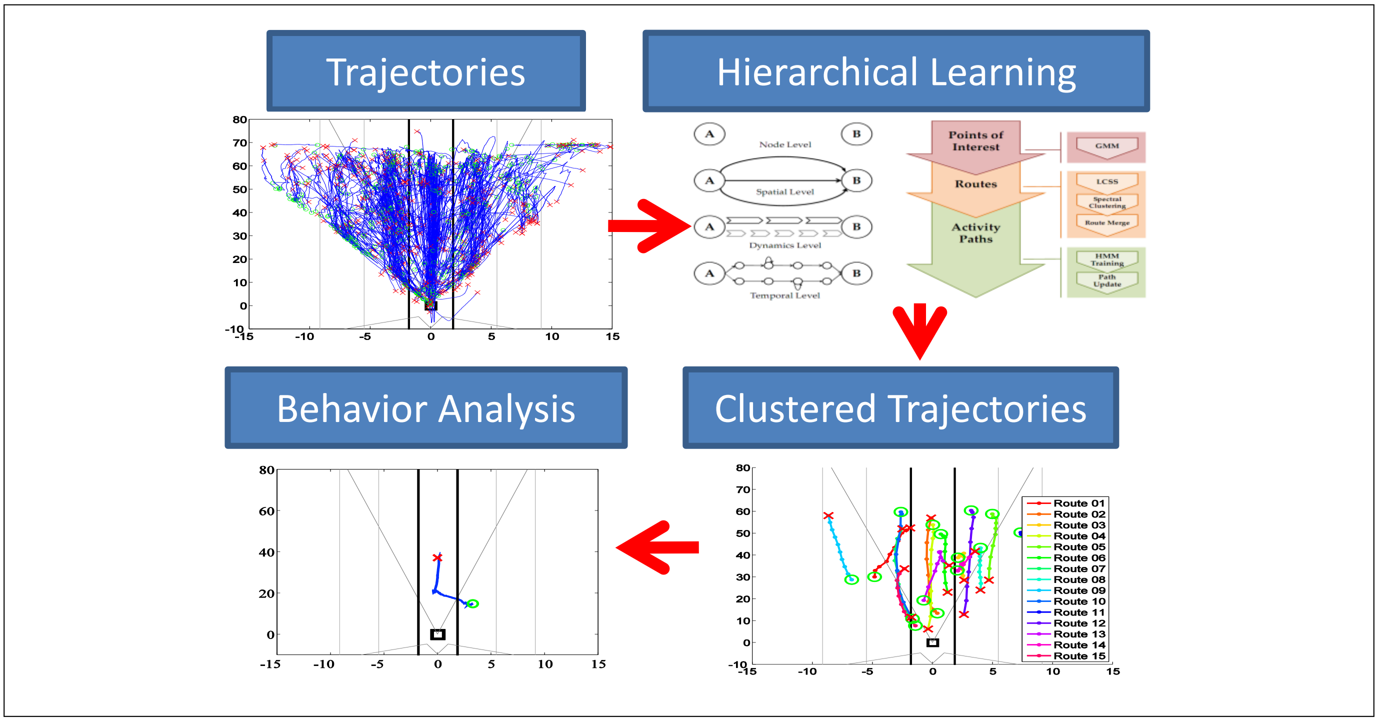


Fig. 9: Activity analysis of people in surrounding vehicles. In [94], a hierarchical representation of the trajectory dynamics is used to perform behavior analysis of vehicle motion patterns. A Hidden Markov Model is used to perform trajectory classification and detect abnormal trajectory events.

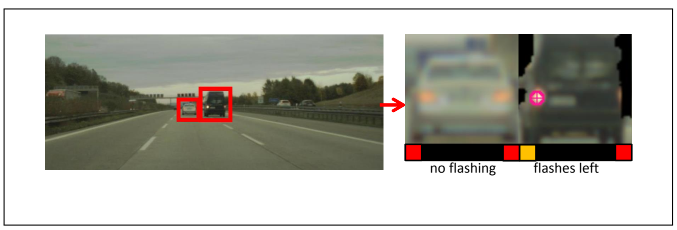


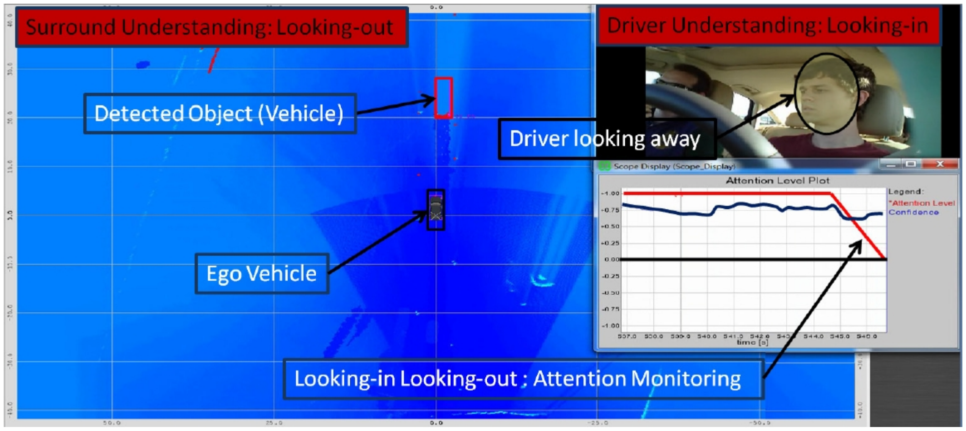
Fig. 10: Intent detection using turn signal analysis [88]. First, vehicles are detected and tracked using a Mixture-of-Experts model and a Kanade-Lucas-Tomasi tracker. Consequently, light spots are detected, and classification of events is performed with an AdaBoost classifier over frequency-domain features.

activity classification [63, 65–67, 70, 71, 73–78]. Map information and vision-based pedestrian tracking are employed in [67] for risk estimation of pedestrians around a vehicle. Body pose and head pose cues can be used to infer pedestrian intent to cross and predict path [68, 73, 74, 134–136]. In Kooij *et al.* [73] pedestrian situation awareness (head orientation), distance-based situation criticality, and spatial layout (curb cues) are employed on top of a Switching Linear Dynamical System to anticipate pedestrian crossing (Fig. 8). Gait analysis using body pose for walking activity classification has been studied in [77, 79]. Spatio-temporal relationships between people have been incorporated in [78] for activity classification. As shown in Table II, finer-grained semantic analysis of skill, style, attention, distraction, and social interaction inference of people around the vehicle is in its early stages. Several recent naturalistic driving datasets with additional modalities, fine-grained attribute and pose information [137–140] will help to further push the richness of analysis provided by algorithms looking at humans around the vehicle. Increased resolution of the sensing modules will play a key role in advances for intricate analysis of pedestrian state, intent, and social relationship modeling [78, 122]. Because smooth and safe driving often involves navigation around humans (e.g. construction zones) and interaction with pedestrians (Fig. 7 depicts some of the relevant research tasks), this domain of human analysis for intelligent vehicles is expected to have high research and commercial activity.

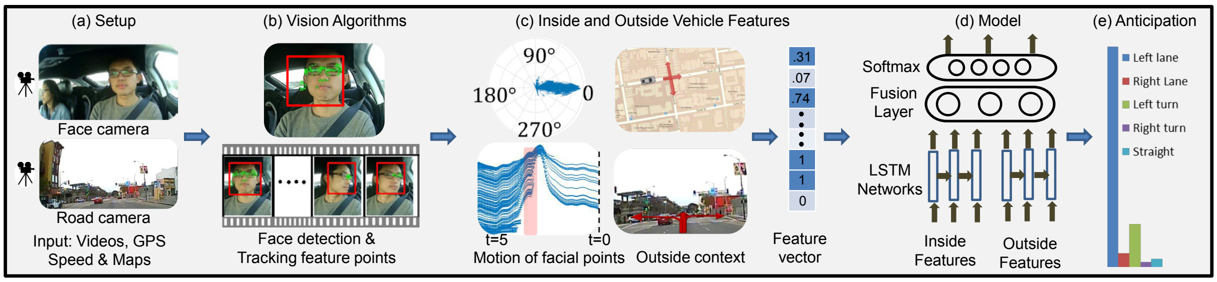
C. Looking at Humans in Surround Vehicles

Understanding intent of drivers in surround vehicles, a task continuously performed by human drivers, is also useful for machine drivers. The research tasks are therefore shared across the three domains of humans in intelligent vehicles. When looking at humans in surround vehicles, vision-based algorithms can be applied to understand behavior and intent, predict maneuvers, and recognize skill, style, and attention.

Understanding activity and modeling intent of other vehicles is widely researched for path prediction and activity classification [85–87, 141]. Intent modeling is a critical step towards risk assessment [55, 89–92]. Lefèvre *et al.* [54] employs a



(a) Helping a distracted driver by sensing situational need and driver alertness levels [111].



(b) Lane change maneuver prediction using driver and surround cues [112]. A Recurrent Neural Network with Long Short-Term Memory (LSTM) units is employed to fuse cue modalities and capture temporal dependencies.

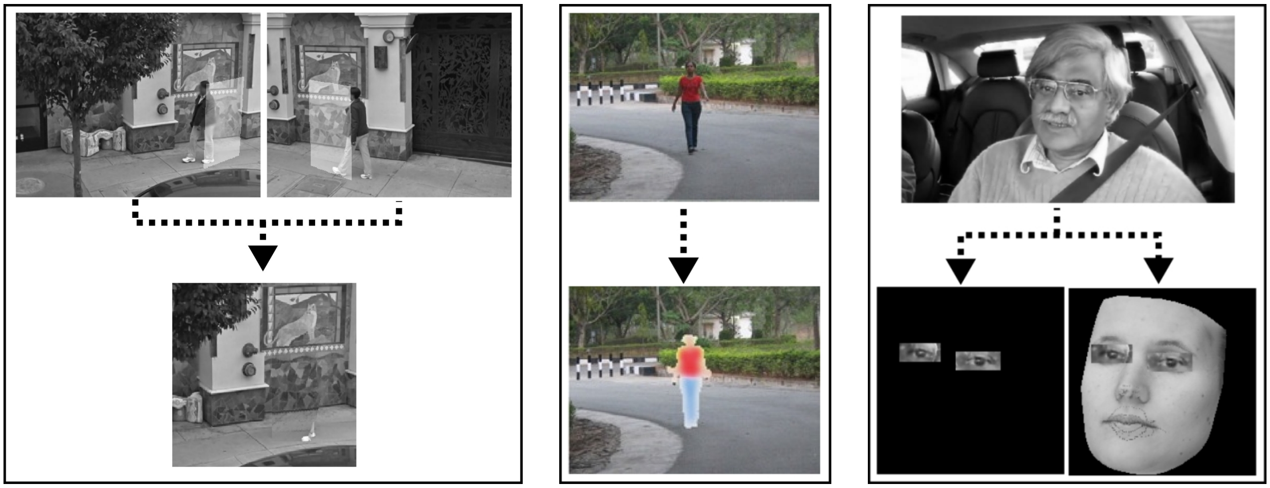
Fig. 11: Emerging research topics in integrative frameworks for on-road activity analysis.

Dynamic Bayesian model over spatial layout and vehicles state (position, orientation, and speed) cues for detecting conflicting intentions and estimating risk at intersections. In Zhang *et al.* [96], a generative model for modeling traffic patterns at intersections is proposed using vehicle trajectory, orientation, and scene cues. Sivaraman *et al.* [94] proposes learning trajectory patterns of surround vehicles with a hierarchical representation of trajectory dynamics and a Hidden Markov Model. The trajectory patterns are employed for surround vehicles behavior analysis, including detection of abnormal events. Detection of turn signals [84, 88, 93] is also useful in understanding the intent of humans in surround vehicles (Fig. 10). In Fröhlich *et al.* [88], vehicles are detected using a Mixture-of-Experts model and tracked with a Kanade-Lucas-Tomasi tracker. After background segmentation and light spot detection, an AdaBoost classifier is employed over frequency-domain features for performing turn signal analysis. Because predicting intents of other vehicles is crucial to safe driving, a robotic driving system should capture subtle cues of aggressiveness, skill, style, attention, and distraction of humans in surround vehicles. It is known that age, gender, and other properties of the human driver influence driver behavior [85], so that vision-based observation of humans in other vehicles (e.g. body pose cues, preparatory movement of other drivers, age classification, etc.) can be useful when working towards aforementioned research tasks.

D. *Integrative Frameworks*

On the road, humans inside vehicles, around vehicles, and in surround vehicles all interact together. Therefore, intelligent vehicles are vehicles that can integrate information coming from multiple domains for better scene understanding and improved forecasting [142]. Holistic understanding is useful for effective and appropriately engaged driver assistance system, successful human-robot communication, and autonomous driving. Example integrative systems are shown in Fig. 9.

As drivers interact with their surrounding continuously, driver activities are often related to surrounding agent cues (e.g. other vehicles and pedestrians). Maneuver prediction [101–103, 143] often requires integrating surround and cabin cues for an improved model of the driver state and consequently better early event detection with lower false positive rates. In Ohn-Bar *et al.* [102], both driver observing cues (head, hand, and foot) and surround agent cues (distance and locations to other vehicles) are integrated with Multiple



(a) Preserving scene [82] (b) Preserving action [83] (c) Preserving gaze direction [1]

Fig. 12: Comparison of selected works in de-identification from different applications: (a) Google street view: removing pedestrians and preserving scene using multiple views, (b) Surveillance: Obscuring identity of actor and preserving action and (c) Intelligent vehicles: Protecting driver's identity and preserving driver's gaze.

Kernel Learning to identify intent of the ego-vehicle driver to overtake. Driver attention estimation is another common research theme in integrative frameworks, where driver cues and surround object cues, such as pedestrian detection [99] or salient objects [105], are integrated to estimate attentiveness to surround objects. In Tawari *et al.* [111], situational need assessment and driver alertness levels are employed as cues for an assistive braking system (Fig. 11). Jain *et al.* [112] employs multi-modal Long Short-Term Memory networks for maneuver anticipation.

III. Naturalistic Datasets and Analysis Tools

The survey of related research studies in Section II captured the research landscape in terms of what has been done, and what still needs to be done. As in all science and engineering fields, a key component in future research relies on access to naturalistic, high-quality, large datasets which can provide insights into better algorithmic and system designs. Studying user-specific nuances and achieving better situational awareness in autonomous systems all require standardized metrics and benchmarks. Furthermore, data accessibility issues are a main reason why integrative frameworks are still little developed and understood on a principled manner. We therefore mention current tools and datasets available to the scientific community for the study of humans in and around vehicles. The discussion further raises issues as to requirements for further progress in the field.

A. Towards Privacy Protecting Safety Systems

The development of intelligent vehicles requires careful consideration of safety and security of people in and around the vehicle. This article has touched upon the fundamentals needed to deal with safety issues but as naturalistic datasets are developed there are important questions about security and identity.

There is a trade-off between privacy and extracting driver behavior. Many existing state-of-the-art algorithms on driver behavior are able to achieve their purpose due to analysis of raw signal and video input, with possible privacy implications. Privacy preserving considerations may play a role in the construction of publicly available large-scale datasets, especially as current state-of-the-art algorithms for intelligent vehicles require large amounts of data for training and evaluation. Therefore, as a community, it is important to raise the standards of both safety and security in the development on intelligent vehicles.

B. Naturalistic Driving Datasets

Table III lists recent datasets which are publicly available for the study of humans inside and around the vehicle. As can be seen, only a handful of such standardized datasets currently exist. Because pedestrian detection and tracking is a well-studied problem, such tasks have several publicly available benchmarks, including Caltech pedestrians [144], Daimler [145], KITTI [138], and Cityscapes [146, 147]. The Caltech roadside pedestrians dataset [137] includes body pose and fine-grained pedestrian attribute information. Other datasets are not generally captured in driving settings (e.g. surveillance applications [148], static camera [78], and stroller or hand-held camera [149–151]).

The datasets are visualized in Fig. 14, demonstrating the progress that has been made in the field so far. Face and hand detection and analysis can now be measured in harsh occlusion and illumination settings in the vehicle. Similarly, challenging datasets observing surround agents continuously push the field further with comparative evaluations. As can be seen in Fig. 14, the majority of the dataset emphasizes basic vision tasks of detection, segmentation, or pose estimation. On exception is the Brain4Cars dataset [51] which provides anno-

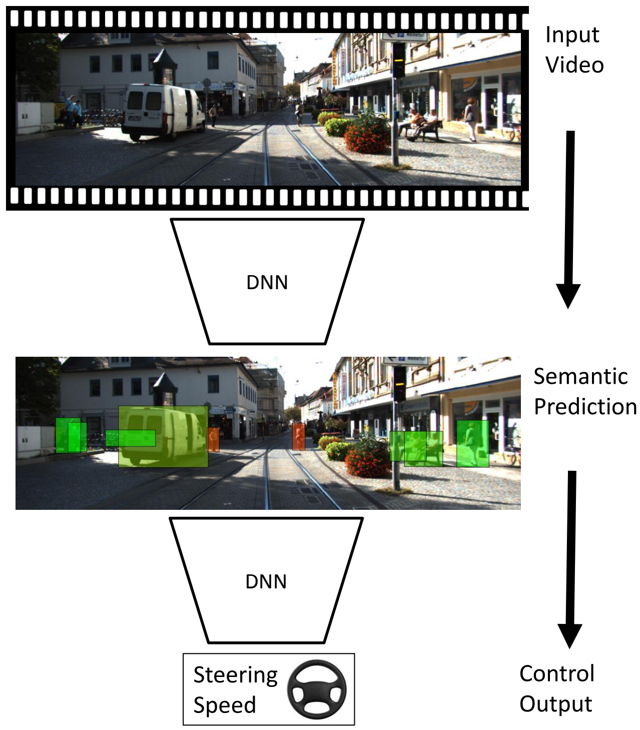


Fig. 13: Example image-to-control policy pipeline (mediated-semantic perception [152, 153]) with deep networks (DNN), where initial prediction of semantic scene elements is followed by a control policy algorithm.

TABLE III: Overview of selected publicly available naturalistic datasets from a mobile vehicle platform.

Dataset	Description
Studying humans inside cabin	
VIVA-Hands [118, 173] (2014)	Detection, tracking, and gestures of driver and passenger hands in video.
VIVA-Faces [174] (2014)	Detection and pose estimation of in-vehicle occupants' faces.
Studying humans inside cabin and in surround vehicles.	
Brain4Cars [51]	Lane change maneuver prediction with cabin-view camera, scene-view camera, GPS, and vehicle dynamics.
Studying humans around vehicles.	
Caltech [137] (2015)	Body pose and fine-grained classification of pedestrians, including age, gender, and activity.
Studying surround vehicles and humans around vehicles.	
KITTI [138] (2012)	Vehicle and pedestrian 3D tracklets annotated with stereo imagery, GPS, lidar, and vehicle dynamics.
Cityscapes [146] (2015)	On-road object segmentation with stereo video, vehicle dynamics, and GPS.

tations for activity anticipation. As methods further progress on such recent benchmarks, additional higher-level semantic tasks such as activity understanding and forecasting could be introduced and evaluated.

IV. Concluding Remarks

Intelligent vehicles are at the core of transforming how people and goods are transported. As technology takes a step closer towards self-driving with recent advances in machine sensing, learning, and planning, many issues are still left unresolved. In particular, we highlight research issues as they relate to the understanding of human agents which interact with the automated vehicle. Self-driving and highly automated vehicles are required to navigate smoothly while avoiding obstacles and understanding high levels of scene semantics. For achieving such goals, further developments in perception (e.g. driveable paths), 3D scene understanding, and policy planning are needed. The current surge of interest in intelligent vehicle technologies is related to recent progress and increased maturity in image recognition techniques [154–157] and, in particular, to the successful application of deep learning to image and signal recognition tasks [158–162]. Deep temporal reasoning approaches [112, 163] have also shown similarly impressive performance, and are useful for a variety of learning tasks (e.g. distraction detection [20]). Furthermore, control policy for self-driving, both mediated-semantic perception approaches [152] and behavior reflex, end-to-end, image to control space approaches [164–172] (e.g. Fig. 13) have been making major strides. The exciting and expanding research frontiers raise additional questions regarding the ability of techniques to capture context in a holistic manner, handle many atypical scenarios and objects, perform analysis of fine-grained short-term and long-term activity information regarding observed agents, forecast activity events and make decisions while being surrounded by human agents, and interact with humans.

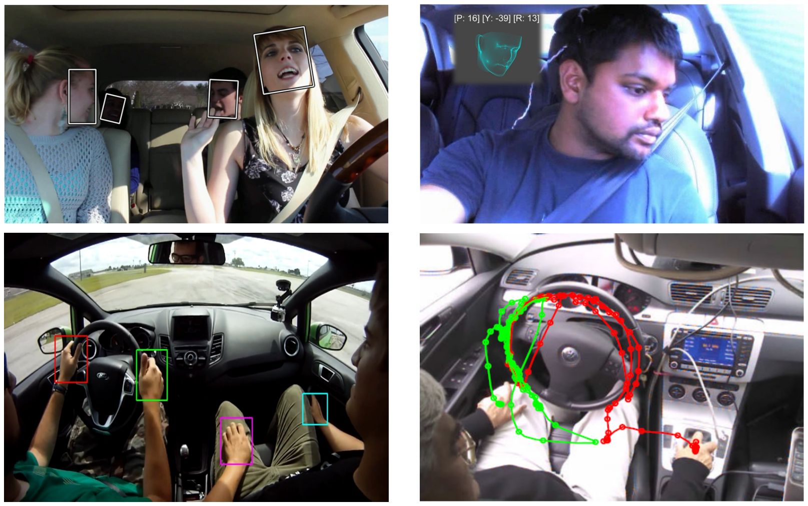
Moving towards vehicles with higher autonomy opens new research avenues in dealing with learning, modeling, active control, perception of dynamic events, and novel architectures for distributed cognitive systems. Furthermore, these challenges must be addressed in a safety-time critical context. We hope that this paper serves as an invitation to pursue exciting multidisciplinary research leading towards a safer, smoother, efficient, and enjoyable driving experience.

V. Acknowledgments

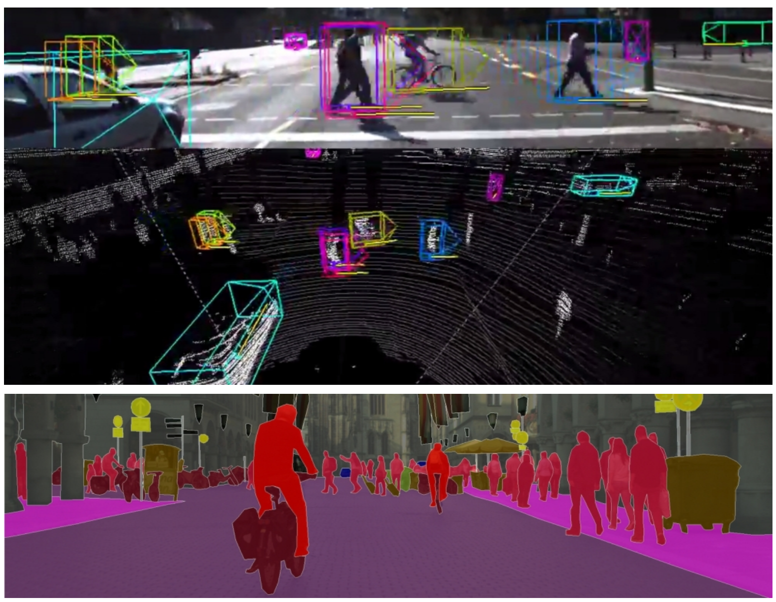
The invaluable contributions of over 80 dedicated scholars of the UCSD LISA team who pursued a research agenda on human-centered intelligent vehicles over the past fifteen years is greatly acknowledged. Their contributions have been at the core of the ideas presented in this paper. We thank Sujitha Martin for her assistance and comments. We are also grateful to the generous support of our sponsors that has allowed us to sustain our scientific and engineering pursuits. We sincerely appreciate the invitation by the Editor-in-Chief of the new *IEEE Transactions on Intelligent Vehicles* for giving an opportunity to prepare this manuscript. We thank the reviewers and associate editor for their valuable comments.

References

[1] S. Martin, A. Tawari, and M. M. Trivedi, "Toward privacy-protecting safety systems for naturalistic driving videos," *IEEE Trans. Intelligent Transportation Systems*, 2014.



(a) Naturalistic datasets for looking inside the vehicle for face and hand analysis (VIVA dataset).



(b) Naturalistic datasets for looking at humans around a vehicle and in other vehicles (top - KITTI, bottom - Cityscapes).

Fig. 14: Example images from publicly available datasets (Table III) for analysis of humans inside and outside of the vehicle.

[2] M.-I. Toma, L. J. Rothkrantz, and C. Antonya, "Driver cell phone usage detection on strategic highway research program (SHRP2) face view videos," in *IEEE Intl. Conf. on Cognitive Infocommunications*, 2012.

[3] K. Behn, A. Pavelkov, and A. Herout, "Implicit hand gestures in aeronautics cockpit as a cue for crew state and workload inference," in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2015.

[4] A. Fuentes, R. Fuentes, E. Cabello, C. Conde, and I. Martin, "Videosensor for the detection of unsafe driving behavior in the proximity of black spots," *Sensors*, vol. 14, no. 11, 2014.

[5] F. Attal, A. Boubezoul, L. Oukhellou, and S. Espi, "Riding patterns recognition for powered two-wheelers users' behaviors analysis," in *IEEE Conf. Intelligent Transprtation Systems*, 2013.

[6] A. Bender, G. Agamennoni, J. R. Ward, S. Worrall, and E. M. Nebot, "An unsupervised approach for inferring driver behavior from naturalistic driving data," *IEEE Trans. Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3325–3336, 2015.

[7] A. Sathyanarayana, S. O. Sadjadi, and J. H. L. Hansen, "Leveraging sensor information from portable devices towards automatic driving maneuver recognition," in *IEEE Conf. Intelligent Transportation Systems*, 2012.

[8] L. M. Bergasa, D. Almera, J. Almazn, J. J. Yebes, and R. Arroyo, "Drivesafe: An app for alerting inattentive drivers and scoring driving behaviors," in *IEEE Intelligent Vehicles Symposium*, 2014.

[9] E. Ohn-Bar, S. Martin, A. Tawari, and M. M. Trivedi, "Head, eye, and hand patterns for driver activity recognition," in *IEEE Intl. Conf. Pattern Recognition*, 2014.

[10] F. Parada-Loira, E. Gonzlez-Agulla, and J. L. Alba-Castro, "Hand gestures to control infotainment equipment in cars," in *IEEE Intelligent Vehicles Symposium*, 2014.

[11] C. Ahlstrom, T. Victor, C. Wege, and E. Steinmetz, "Processing of eye/head-tracking data in large-scale naturalistic driving data sets," *IEEE Trans. Intelligent Transportation Systems*, 2012.

[12] A. Rangesh, E. Ohn-Bar, and M. M. Trivedi, "Hidden hands: Tracking hands with an occlusion aware tracker," in *IEEE Conf. Computer Vision and Pattern Recognition Workshops-HANDS*, 2016.

[13] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Multi-sensor system for drivers hand-gesture recognition," in *IEEE Intl. Conf. Automatic Face and Gesture Recognition*, 2015.

[14] A. Tawari, K. H. Chen, and M. M. Trivedi, "Where is the driver looking: Analysis of head, eye and iris for robust gaze zone estimation," in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2014.

[15] C. Tran, A. Doshi, and M. M. Trivedi, "Modeling and prediction of driver behavior by foot gesture analysis," *Computer Vision and Image Understanding*, vol. 116, pp. 435–445, 2012.

[16] E. Ohn-Bar and M. M. Trivedi, "Beyond just keeping hands on the wheel: Towards visual interpretation of driver hand motion patterns," in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2014.

[17] P. Molchanov, S. Gupta, K. Kim, and J. Kautz, "Hand gesture recognition with 3D convolutional neural networks," in *IEEE Conf. Computer Recognition and Pattern Recognition-HANDS*, 2015.

[18] S. Cheng and M. M. Trivedi, "Vision-based infotainment user determination by hand recognition for driver assistance," *IEEE Trans. Intelligent Transportation Systems*, 2010.

[19] A. D. lvarez, Francisco, S. Garca, J. E. Naranjo, J. J. Anaya, and F. Jimnez, "Modeling the driving behavior of electric vehicles using smartphones and neural networks," *IEEE Trans. Intelligent Transportation Systems Magazine*, 2012.

[20] M. Wllmer, C. Blaschke, T. Schindl, B. Schuller, B. Frber, S. Mayer, and B. Trefflich, "Online driver distraction detection using long short-term memory," *IEEE Trans. Intelligent Transportation Systems*, 2011.

[21] P. Jimnez, L. M. Bergasa, J. Nuevo, N. Hernndez, and I. G. Daza, "Gaze fixation system for the evaluation of driver distractions induced by ivis," *IEEE Trans. Intelligent Transportation Systems*, 2012.

[22] R. O. Mbouna, S. G. Kong, and M.-G. Chun, "Visual analysis of eye state and head pose for driver alertness monitoring," *IEEE Trans. Intelligent Transportation Systems*, 2013.

[23] T. Liu, Y. Yang, G.-B. Huang, Y. K. Yeo, and Z. Lin, "Driver distraction detection using semi-supervised machine learning," *IEEE Trans. Intelligent Transportation Systems*, 2015.

[24] F. Vicente, Z. Huang, X. Xiong, F. D. la Torre, W. Zhang, and D. Levi, "Driver gaze tracking and eyes off the road detection system," *IEEE Trans. Intelligent Transportation Systems*, 2015.

[25] A. Tawari and M. M. Trivedi, "Robust and continuous estimation of driver gaze zone by dynamic analysis of multiple face videos," in *IEEE Intelligent Vehicles Symposium*, 2014.

[26] A. Witayangkurn, T. Horanont, Y. Sekimoto, and R. Shibasaki, "Anomalous event detection on large-scale gps data from mobile phones using hidden markov model and cloud platform," in *Pervasive and Ubiquitous Computing*, 2013.

[27] M.-I. Toma, L. J. Rothkrantz, and C. Antonya, "Car driver skills assessment based on driving postures recognition," in *IEEE Intl. Conf. on Cognitive Infocommunications*, 2012.

[28] M. V. Ly, S. Martin, and M. Trivedi, "Driver classification and driving style recognition using inertial sensors," in *IEEE Intelligent Vehicles Symposium*, 2013.

[29] D. Johnson and M. Trivedi, "Driving style recognition using a smartphone as a sensor platform," in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2011.

[30] S. Lefèvre, A. Carvalho, Y. Gao, H. E. Tseng, and F. Borrelli, "Driver models for personalised driving assistance," *Vehicle System Dynamics*, vol. 53, no. 12, pp. 1705–1720, 2015.

[31] D. Drr, D. Grabengiesser, and F. Gauterin, "Online driving style recognition using fuzzy logic," in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2014.

[32] A. S. Zeeman and M. J. Booysen, "Combining speed and acceleration to detect reckless driving in the informal public transport industry," in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2013.

[33] A. Aljaafreh, N. Alshabatat, and M. S. N. Al-Din, "Driving style recognition using fuzzy logic," in *IEEE Intl. Conf. Vehicular Electronics and Safety*, 2012.

[34] H. Eren, S. Makinist, E. Akin, and A. Yilmaz, "Estimating driving behavior by a smartphone," in *IEEE Intelligent Vehicles Symposium*, 2012.

[35] J. Dai, J. Teng, X. Bai, Z. Shen, and D. Xuan, "Mobile phone based drunk driving detection," in *Intl. Conf. Pervasive Computing Technologies for Healthcare*, 2010.

[36] D. W. Koh and H. B. Kang, "Smartphone-based modeling and detection of aggressiveness reactions in senior drivers," in *IEEE Intelligent Vehicles Symposium*, 2015.

[37] R. Arajo, . Igreja, R. de Castro, and R. E. Arajo, "Driving coach: A smartphone application to evaluate driving efficient patterns," in *IEEE Intelligent Vehicles Symposium*, 2012.

[38] G. Castignani, R. Frank, and T. Engel, "Driver behavior profiling using smartphones," in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2013.

[39] erman Castignani, T. Derrmann, R. Frank, and T. Enge, "Driver behavior profiling using smartphones: A low-cost platform for driver monitoring," *IEEE Intelligent Transportation Systems Magazine*, 2015.

[40] J.-H. Hong, B. Margines, and A. K. Dey, "A smartphone-based sensing platform to model aggressive driving behaviors," in *ACM Conf. Human Factors in Computing Systems*, 2014.

[41] H. Eren, S. Makinist, E. Akin, and A. Yilmaz, "Estimating driving behavior by a smartphone," in *IEEE Intelligent Vehicles Symposium*, 2012.

[42] J. Gonalves, J. S. V. Gonalves, R. J. F. Rossetti, and C. Olaverri-Monreal, "Smartphone sensor platform to study traffic conditions and assess driving performance," in *IEEE Conf. on Intelligent Transportation Systems*, 2014.

[43] C. Gold, D. Dambck, L. Lorenz, and K. Bengler, "take over! how long does it take to get the driver back into the loop?" *Human Factors and Ergonomics*, vol. 57, no. 1, pp. 1938–1942, 2013.

[44] V. A. Shia, Y. Gao, R. Vasudevan, K. D. Campbell, T. Lin, F. Borrelli, and R. Bajcsy, "Semiautonomous vehicular control using driver modeling," *IEEE Trans. Intelligent Transportation Systems*, vol. 15, no. 6, pp. 2696–2709, 2014.

[45] V. A. Banks and N. A. Stanton, "Keep the driver in control: Automating automobiles of the future," *Applied Ergonomics*, vol. 53, Part B, pp. 389–395, 2016.

[46] J. Koo, J. Kwac, W. Ju, M. Steinert, L. Leifer, and C. Nass, "Why did my car just do that? explaining semi-autonomous driving actions to improve driver understanding, trust, and performance," *Interactive Design and Manufacturing*, vol. 9, no. 4, pp. 269–275, 2014.

[47] M. Walch, K. Lange, M. Baumann, and M. Weber, "Autonomous driving: Investigating the feasibility of car-driver handover assistance," in *Intl. Conf. AutomotiveUI*, 2015.

[48] C. Braunagel, W. Stolzmann, E. Kasneci, and W. Rosenstiel, "Driver-activity recognition in the context of conditionally autonomous driving," in *IEEE Conf. Intelligent Transportation Systems*, 2015.

[49] S. Lefèvre, J. Ibañez-Guzmán, and C. Laugier, "Context-based estimation of driver intent at road intersections," in *IEEE Intelligent Vehicles Symposium*, 2011.

[50] A. Nakano, H. Okuda, T. Suzuki, S. Inagaki, and S. Hayakawa, "Symbolic modeling of driving behavior based on hierarchical segmentation and formal grammar," *Intl. Conf. Intelligent Robots and Systems*, pp. 5516–5521, 2009.

[51] A. Jain, H. S. Koppula, B. Raghavan, S. Soh, and A. Saxena, "Car that knows before you do: Anticipating maneuvers via learning temporal driving models," in *IEEE Intl. Conf. Computer Vision*, 2015.

[52] M. Liebner, M. Baumann, F. Klanner, and C. Stiller, "Driver intent inference at urban intersections using the intelligent driver model," in *IEEE Intelligent Vehicles Symposium*, 2012.

[53] H. Berndt and K. Dietmayer, "Driver intention inference with vehicle onboard sensors," in *IEEE Conf. Vehicular Electronics and Safety*, 2009.

[54] S. Lefèvre, C. Laugier, and J. Ibañez-Guzmán, "Evaluating risk at road intersections by detecting conflicting intentions," in *IEEE Conf. Intelligent Robots and Systems*, 2012.

[55] T. Streubel and K. H. Hoffmann, "Prediction of driver intended path at intersections," in *IEEE Intelligent Vehicles Symposium*, 2014, pp. 134–139.

[56] J. Krumm, "A markov model for driver turn prediction," *SAE World Congress*, 2008.

[57] H. Berndt, J. Emmert, and K. Dietmayer, "Continuous driver intent recognition with hidden markov models," *IEEE Intl. Conf. Intelligent*

Transportation Systems, pp. 1189–1194, 2008.

[58] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, "Maximum entropy inverse reinforcement learning," in *AAAI Conference on Artificial Intelligence*, 2008, pp. 1433–1438.

[59] B. D. Ziebart, A. Maas, J. A. Bagnell, and A. K. Dey, "Human behavior modeling with maximum entropy inverse optimal control," *AAAI Conference on Artificial Intelligence*, 2009.

[60] N. Pugeault and R. Bowden, "Learning pre-attentive driving behaviour from holistic visual features," in *European Conf. Computer Vision*, 2010.

[61] B. Tang, S. Khokhar, and R. Gupta, "Turn prediction at generalized intersections," in *IEEE Intelligent Vehicles Symposium*, 2015.

[62] S. Ferguson, B. Luders, R. C. Grande, and J. P. How, "Real-time predictive modeling and robust avoidance of pedestrians with uncertain, changing intentions," in *Intl. Workshop Algorithmic Foundations of Robotics*, 2014.

[63] M. Goldhammer, M. Gerhard, S. Zernetsch, K. Doll, and U. Brunsmann, "Early prediction of a pedestrian's trajectory at intersections," in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2013.

[64] S. Khler, M. Goldhammer, S. Bauer, K. Doll, U. Brunsmann, and K. Dietmayer, "Early detection of the pedestrian's intention to cross the street," in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2012.

[65] F. Madrigal, J.-B. Hayet, and F. Lerasle, "Intention-aware multiple pedestrian tracking," in *IEEE Intl. Conf. Pattern Recognition*, 2014.

[66] R. Quintero, I. Parra, D. Llorca, and M. Sotelo, "Pedestrian path prediction based on body language and action classification," in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2014.

[67] A. Møgelmose, M. Trivedi, and T. Moeslund, "Trajectory analysis and prediction for improved pedestrian safety: Integrated framework and evaluations," in *IEEE Intelligent Vehicles Symposium*, 2015.

[68] C. Keller and D. Gavrila, "Will the pedestrian cross? a study on pedestrian path prediction," *IEEE Trans. Intelligent Transportation Systems*, vol. 15, no. 2, 2014.

[69] T. Gandhi and M. Trivedi, "Image based estimation of pedestrian orientation for improving path prediction," in *IEEE Intelligent Vehicles Symposium*, 2008, pp. 506–511.

[70] A. T. Schulz and R. Stiefelhagen, "Pedestrian intention recognition using latent-dynamic conditional random fields," in *IEEE Intelligent Vehicles Symposium*, 2015.

[71] ——, "A controlled interactive multiple model filter for combined pedestrian intention recognition and path prediction," in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2015.

[72] T. Bandyopadhyay, C. Z. Jie, D. Hsu, M. H. Ang, D. Rus, and E. Frazzoli, "Intention-aware pedestrian avoidance," in *Intl. Symposium on Experimental Robotics*, P. J. Desai, G. Dudek, O. Khatib, and V. Kumar, Eds., 2013.

[73] J. F. P. Kooij, N. Schneider, F. Flohr, and D. M. Gavrila, "Context-based pedestrian path prediction," in *European Conf. Computer Vision*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., 2014.

[74] J. F. P. Kooij, N. Schneider, and D. M. Gavrila, "Analysis of pedestrian dynamics from a vehicle perspective," in *IEEE Intelligent Vehicles Symposium*, 2014.

[75] J. F. P. Kooij, G. Englebienne, and D. M. Gavrila, "Mixture of switching linear dynamics to discover behavior patterns in object tracks," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 322–334, 2016.

[76] W. Choi and S. Savarese, "Understanding collective activities of people from videos," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 36, no. 6, pp. 1242–1257, 2014.

[77] M. Goldhammer, A. Hubert, S. Koehler, K. Zindler, U. Brunsmann, K. Doll, and B. Sick, "Analysis on termination of pedestrians gait at urban intersections," in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2014.

[78] W. Choi, K. Shahid, and S. Savarese, "What are they doing? : Collective activity classification using spatio-temporal relationship among people," in *IEEE Intl. Conf. Computer Vision Workshops*, 2009.

[79] H. Kataoka, Y. Aoki, Y. Satoh, S. Oikawa, and Y. Matsui, "Fine-grained walking activity recognition via driving recorder dataset," in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2015.

[80] J. Kooij, M. Liem, J. Krijnders, T. Andringa, and D. Gavrila, "Multi-modal human aggression detection," *Computer Vision and Image Understanding*, vol. 144, pp. 106–120, 2016.

[81] D. Llorca, R. Quintero, I. Parra, R. Izquierdo, C. Fernandez, and M. Sotelo, "Assistive pedestrian crossings by means of stereo localization and rfid anonymous disability identification," in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2015.

[82] A. Flores and S. Belongie, "Removing pedestrians from google street view images," in *IEEE Conf. Computer Vision and Pattern Recognition Workshops*, 2010.

[83] P. Agrawal and P. Narayanan, "Person de-identification in videos," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 21, 2011.

[84] B. Li, T. Wu, C. Xiong, and S.-C. Zhu, "Recognizing car fluents from video," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2016.

[85] A. Jahangiri, H. A. Rakha, and T. A. Dingus, "Adopting machine learning methods to predict redlight running violations," in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2015.

[86] C. L. Azevedo and H. Farah, "Using extreme value theory for the prediction of head-on collisions during passing maneuvres," in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2015.

[87] T. Gindele, S. Brechtel, and R. Dillmann, "Learning context sensitive behavior models from observations for predicting traffic situations," in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2013.

[88] B. Fröhlich, M. Enzweiler, and U. Franke, "Will this car change the lane? - turn signal recognition in the frequency domain," in *IEEE Intelligent Vehicles Symposium*, 2014.

[89] R. Graf, H. Deusch, F. Seeliger, M. Fritzsche, and K. Dietmayer, "A learning concept for behavior prediction at intersections," in *IEEE Intelligent Vehicles Symposium*, 2014.

[90] 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 *IEEE Conf. Intelligent Transportation Systems*, 2010.

[91] C. Laugier, I. E. Paromtchik, M. Perrollaz, M. Yong, J. D. Yoder, C. Tay, K. Mekhnacha, and A. Ngre, "Probabilistic analysis of dynamic scenes and collision risks assessment to improve driving safety," *IEEE Intelligent Transportation Systems Magazine*, vol. 3, no. 4, pp. 4–19, 2011.

[92] G. Agamennoni, J. I. Nieto, and E. M. Nebot, "A bayesian approach for driving behavior inference," in *IEEE Intelligent Vehicles Symposium*, 2011.

[93] R. K. Satzoda and M. M. Trivedi, "Looking at vehicles in the night: Detection dynamics of rear lights," *IEEE Trans. Intelligent Transportation Systems*, 2016.

[94] S. Sivaraman, B. Morris, and M. M. Trivedi, "Learning multi-lane trajectories using vehicle-based vision," in *IEEE Intl. Conf. Computer Vision Workshops-CVVT*, 2011.

[95] M. P. Philipsen, M. B. Jensen, R. K. Satzoda, M. M. Trivedi, A. Møgelmose, and T. B. Moeslund, "Day and night-time drive analysis using stereo vision for naturalistic driving studies," in *IEEE Intelligent Vehicles Symposium*, 2015.

[96] H. Zhang, A. Geiger, and R. Urtasun, "Understanding high-level semantics by modeling traffic patterns," in *IEEE Intl. Conf. on Computer Vision*, 2013.

[97] M. T. Phan, V. Fremont, I. Thouvenin, M. Sallak, and V. Cherfaoui, "Recognizing driver awareness of pedestrian," in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2014, pp. 1027–1032.

[98] R. Tanishige, D. Deguchi, K. Doman, Y. Mekada, I. Ide, and H. Murase, "Prediction of driver's pedestrian detectability by image processing adaptive to visual fields of view," in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2014.

[99] A. Tawari, A. Mogelmose, S. Martin, T. Moeslund, and M. Trivedi, "Attention estimation by simultaneous analysis of viewer and view," in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2014.

[100] B. Morris, A. Doshi, and M. Trivedi, "Lane change intent prediction for driver assistance: On-road design and evaluation," in *IEEE Intelligent Vehicles Symposium*, 2011.

[101] A. Doshi, B. T. Morris, and M. M. Trivedi, "On-road prediction of driver's intent with multimodal sensory cues," *IEEE Pervasive Computing*, vol. 10, pp. 22–34, 2011.

[102] E. Ohn-Bar, A. Tawari, S. Martin, and M. M. Trivedi, "On surveillance for safety critical events: In-vehicle video networks for predictive driver assistance systems," *Computer Vision and Image Understanding*, vol. 134, pp. 130–140, 2015.

[103] J. McCall and M. M. Trivedi, "Driver behavior and situation aware brake assistance for intelligent vehicles," *Proceedings of the IEEE*, vol. 95, pp. 374–387, 2007.

[104] M. Bahram, C. Hubmann, A. Lawitzky, M. Aeberhard, and D. Wollherr, "A combined model- and learning-based framework for interaction-aware maneuver prediction," *IEEE Trans. Intelligent Transportation*

Systems, 2016.

[105] A. Doshi and M. M. Trivedi, “Attention estimation by simultaneous observation of viewer and view,” in *IEEE Conf. Computer Vision and Pattern Recognition Workshops*, 2010.

[106] ——, “Investigating the relationships between gaze patterns, dynamic vehicle surround analysis, and driver intentions,” *IEEE Intelligent Vehicles Symposium*, June 2009.

[107] T. Bar, D. Linke, D. Nienhuser, and J. Zollner, “Seen and missed traffic objects: A traffic object-specific awareness estimation,” in *IEEE Intelligent Vehicles Symposium*, 2013.

[108] M. Mori, C. Miyajima, P. Angkititrakul, T. Hirayama, Y. Li, N. Kitaoka, and K. Takeda, “Measuring driver awareness based on correlation between gaze behavior and risks of surrounding vehicles,” in *IEEE Conf. Intelligent Transportation Systems*, 2012.

[109] M. Rezaei and R. Klette, “Look at the driver, look at the road: No distraction! no accident!” in *IEEE Conf. Computer Vision and Pattern Recognition*, 2014.

[110] K. Takagi, H. Kawanaka, M. Bhuiyan, and K. Oguri, “Estimation of a three-dimensional gaze point and the gaze target from the road images,” in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2011.

[111] A. Tawari, S. Sivaraman, M. M. Trivedi, T. Shannon, and M. Tippelhofer, “Looking-in and looking-out vision for urban intelligent assistance: Estimation of driver attentive state and dynamic surround for safe merging and braking,” in *IEEE Intelligent Vehicles Symposium*, 2014.

[112] A. Jain, H. S. Koppula, S. Soh, B. Raghavan, A. Singh, and A. Saxena, “Brain4cars: Car that knows before you do via sensory-fusion deep learning architecture,” *CoRR*, vol. abs/1601.00740, 2016.

[113] S. Pellegrini, A. Ess, K. Schindler, and L. van Gool, “You’ll never walk alone: Modeling social behavior for multi-target tracking,” in *IEEE Intl. Conf. on Computer Vision*, 2009.

[114] U. Ozguner, T. Acarman, and K. Redmill, *Autonomous ground vehicles*. Artech House, 2011.

[115] S. Martin, E. Ohn-Bar, and M. M. Trivedi, “Automatic critical event extraction and semantic interpretation by looking-inside,” in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2015.

[116] O. Kumtepe, G. B. Akar, and E. Yuncu, “Driver aggressiveness detection using visual information from forward camera,” in *IEEE Intl. Conf. Advanced Video and Signal Based Surveillance*, 2015, pp. 1–6.

[117] S. Hamdar, “Driver behavior modeling,” in *Handbook of Intelligent Vehicles*, 2012, pp. 537–558.

[118] E. Ohn-Bar and M. M. Trivedi, “The power is in your hands: 3D analysis of hand gestures in naturalistic video,” in *IEEE Conf. Computer Vision and Pattern Recognition Workshops-AMFG*, 2013.

[119] M. Sivak and B. Schoettle, “Road safety with self-driving vehicles: General limitations and road sharing with conventional vehicles,” University of Michigan Transportation Research Institute, Tech. Rep. UMTRI-2015-2, 2015.

[120] K. Bengler, K. Dietmayer, B. Farber, M. Maurer, C. Stiller, and H. Winner, “Three decades of driver assistance systems: Review and future perspectives,” *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 4, pp. 6–22, 2014.

[121] A. Ess, B. Leibe, K. Schindler, and L. Van Gool, “Robust multiperson tracking from a mobile platform,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 10, pp. 1831–1846, 2009.

[122] A. Robicquet, A. Alahi, A. Sadeghian, B. Anenberg, J. Doherty, E. Wu, and S. Savarese, “Forecasting social navigation in crowded complex scenes,” *CoRR*, vol. abs/1601.00998, 2016.

[123] D. W. Hansen and Q. Ji, “In the eye of the beholder: A survey of models for eyes and gaze,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 478–500, 2010.

[124] E. Ohn-Bar, S. Martin, and M. M. Trivedi, “Driver hand activity analysis in naturalistic driving studies: Issues, algorithms and experimental studies,” *Journal of Electronic Imaging*, vol. 22, pp. 1–10, 2013.

[125] S. G. Klauer, T. A. Dingus, V. L. Neale, J. D. Sudweeks, and D. J. Ramsey, “Road safety with self-driving vehicles: General limitations and road sharing with conventional vehicles,” Virginia Tech Transportation Institute, Tech. Rep. DOT HS 810 594, 2006.

[126] A. Rangesh, E. Ohn-Bar, and M. M. Trivedi, “Long-term, multi-cue tracking of hands in vehicles,” *IEEE Trans. Intelligent Transportation Systems*, 2016.

[127] E. Ohn-Bar and M. M. Trivedi, “Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations,” *IEEE Trans. Intelligent Transportation Systems*, vol. 15, no. 6, pp. 2368–2377, Dec 2014.

[128] D. Tang, H. J. Chang, A. Tejani, and T. K. Kim, “Latent regression forest: Structured estimation of 3d articulated hand posture,” in *IEEE Conf. Computer Vision and Pattern Recognition*, 2014.

[129] J. S. Supancic, G. Rogez, Y. Yang, J. Shotton, and D. Ramanan, “Depth-based hand pose estimation: Data, methods, and challenges,” in *IEEE Intl. Conf. on Computer Vision*, 2015.

[130] E. Ohn-Bar and M. M. Trivedi, “In-vehicle hand activity recognition using integration of regions,” in *IEEE Intelligent Vehicles Symposium*, 2013.

[131] B. D. Ziebart, A. Maas, A. K. Dey, and J. A. Bagnell, “Navigate like a cabbie: probabilistic reasoning from observed context-aware behavior,” *in Proceedings of the 10th International Conference on Ubiquitous Computing*, 2008.

[132] I. Nizetic, K. Fertalj, and D. Kalpic, “A prototype for the short-term prediction of moving object’s movement using markov chains,” *Intl. Conf. Information Technology Interfaces*, pp. 559–564, 2009.

[133] J. Krumm, “Where will they turn: predicting turn proportions at intersections,” *Personal and Ubiquitous Computing*, vol. 14, pp. 591–599, 2010.

[134] F. Flohr, M. Dumitru-Guzu, J. Kooij, and D. Gavrila, “Joint probabilistic pedestrian head and body orientation estimation,” in *IEEE Intelligent Vehicles Symposium*, 2014.

[135] ——, “A probabilistic framework for joint pedestrian head and body orientation estimation,” *IEEE Trans. Intelligent Transportation Systems*, vol. 16, no. 4, pp. 1872–1882, 2015.

[136] E. Rehder, H. Kloeden, and C. Stiller, “Head detection and orientation estimation for pedestrian safety,” in *IEEE Intl. Conf. Intelligent Transportation Systems*, 2014.

[137] D. Hall and P. Perona, “Fine-grained classification of pedestrians in video: Benchmark and state of the art,” in *IEEE Conf. Computer Vision and Pattern Recognition*, 2015.

[138] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the KITTI vision benchmark suite,” in *IEEE Conf. Computer Vision and Pattern Recognition*, 2012.

[139] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: The kitti dataset,” in *Intl. Journal of Robotics Research*, 2013.

[140] S. Hwang, J. Park, N. Kim, Y. Choi, and I. S. Kweon, “Multispectral pedestrian detection: Benchmark dataset and baseline,” in *IEEE Conf. Computer Vision and Pattern Recognition*, 2015.

[141] M. S. Kristoffersen, J. V. Dueholm, R. Satzoda, M. Trivedi, A. Møgelmose, and T. Moeslund, “Understanding surrounding vehicular maneuvers: A panoramic vision-based framework for real-world highway studies,” in *IEEE Conf. Computer Vision and Pattern Recognition Workshops-ATS*, 2016.

[142] A. Carvalho, S. Lefèvre, G. Schildbach, J. Kong, and F. Borrelli, “Automated driving: The role of forecasts and uncertaintya control perspective,” *European Journal of Control*, vol. 24, pp. 14–32, 2015.

[143] S. Y. Cheng, S. Park, and M. M. Trivedi, “Multi-spectral and multi-perspective video arrays for driver body tracking and activity analysis,” *Computer Vision and Image Understanding*, vol. 106, pp. 245–257, 2007.

[144] P. Dollár, C. Wojek, B. Schiele, and P. Perona, “Pedestrian detection: An evaluation of the state of the art,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 743–761, 2012.

[145] M. Enzweiler and D. M. Gavrila, “Monocular pedestrian detection: Survey and experiments,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2009.

[146] M. Cordts, M. Omran, S. Ramos, T. Scharwächter, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The Cityscapes dataset,” in *CVPR Workshop on The Future of Datasets in Vision*, 2015.

[147] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The Cityscapes dataset for semantic urban scene understanding,” in *IEEE Conf. Computer Vision and Pattern Recognition*, 2016.

[148] Y. Deng, P. Luo, C. C. Loy, and X. Tang, “Pedestrian attribute recognition at far distance,” in *Intl. Conf. Multimedia*, 2009.

[149] A. Ess, B. Leibe, and L. V. Gool, “Depth and appearance for mobile scene analysis,” in *IEEE Intl. Conf. on Computer Vision*, 2007.

[150] L. Leal-Taixé, A. Milan, I. Reid, S. Roth, and K. Schindler, “Motchallenge 2015: Towards a benchmark for multi-target tracking,” *arXiv:1504.01942 [cs]*, 2015.

[151] M. Andriluka, S. Roth, and B. Schiele, “Monocular 3D pose estimation and tracking by detection,” in *IEEE Conf. Computer Vision and Pattern*

Recognition, 2010.

[152] S. Ullman, "Against direct perception," *Behavioral and Brain Sciences*, vol. 3, no. 03, pp. 373–381, 1980.

[153] E. Ohn-Bar and M. M. Trivedi, "Are all objects equal? deep spatio-temporal importance prediction in driving videos," *Under Review*, 2016.

[154] ——, "Learning to detect vehicles by clustering appearance patterns," *IEEE Trans. Intelligent Transportation Systems*, 2015.

[155] Y. Xiang, W. Choi, Y. Lin, and S. Savarese, "Data-driven 3D voxel patterns for object category recognition," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2015.

[156] T. Wu, B. Li, and S. C. Zhu, "Learning and-or models to represent context and occlusion for car detection and viewpoint estimation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2015.

[157] Q. Hu, S. Paisitkriangkrai, C. Shen, and A. van den Hengel, "Fast detection of multiple objects in traffic scenes with a common detection framework," *IEEE Trans. Intell. Transp. Syst.*, 2015.

[158] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *CVPR*, 2014.

[159] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," in *Intl. Conf. Learning Representations*, 2014.

[160] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Neural Information Processing Systems*, 2012.

[161] X. Chen, K. Kundu, Y. Zhu, A. Berneshawi, H. Ma, S. Fidler, and R. Urtasun, "3D object proposals for accurate object class detection," in *NIPS*, 2015.

[162] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2016.

[163] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, 1997.

[164] D. Pomerleau, "Alvinn: an autonomous land vehicle in a neural network," in *Neural Information Processing Systems*, 2016.

[165] T. Jochem, D. Pomerleau, B. Kumar, and J. Armstrong, "Pans: A portable navigation platform," in *IEEE Intelligent Vehicles Symposium*, 1995.

[166] D. Pomerleau, "Neural network vision for robot driving," 1995.

[167] C. J. C. H. Watkins, "Learning from delayed rewards," Ph.D. dissertation, King's College, Cambridge, 1989.

[168] M. Bojarski, D. D. Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba, "End to end learning for self-driving cars," *CoRR*, vol. arXiv:1604.07316, 2016.

[169] U. Muller, J. Ben, E. Cosatto, B. Flepp, and Y. L. Cun, "Off-road obstacle avoidance through end-to-end learning," in *Neural Information Processing Systems*, 2006.

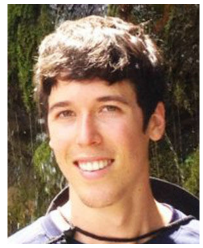
[170] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[171] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," in *Intl. Conf. Learning Representations*, 2016.

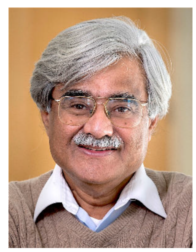
[172] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *Intl. Conf. Machine Learning*, 2016.

[173] N. Das, E. Ohn-Bar, and M. M. Trivedi, "On performance evaluation of driver hand detection algorithms: Challenges, dataset, and metrics," in *IEEE Conf. Intelligent Transportation Systems*, 2015.

[174] "VIVA: Vision for intelligent vehicles and applications challenge," http://cvrr.ucsd.edu/vivachallenge/.



Eshed Ohn-Bar received the B.S. degree in mathematics from the University of California, Los Angeles, and the M.S. degree in electrical engineering from the University of California, San Diego, where he is currently pursuing a Ph.D. degree with a focus on signal and image processing in the Laboratory for Safe and Intelligent Automobiles. His research interests include vision for intelligent vehicles, driver assistance and safety systems, computer vision, object detection and tracking, and multi-modal behavior recognition. He has been the recipient of the best paper award at the IEEE Workshop on Analysis and Modeling of Faces and Gestures and best industry related paper runner up at the Intentional Conference on Pattern Recognition.



Mohan Manubhai Trivedi is a distinguished Professor of Electrical and Computer Engineering and the founding director of LISA: Laboratory for Intelligent and Safe Automobiles and the Computer Vision and Robotics Research Laboratory at the University of California, San Diego. He received the B.E.(Hons.) degree from Birla Institute of Technology and Science, Pilani, India, and the Ph.D. degree from Utah State University, Logan, UT, USA. He and his team are currently engaged in machine and human perception, machine learning, human-robot interactivity, multi-modal effect and gesture analysis, human-centered interfaces, intelligent vehicles, and active safety and embedded systems.

Trivedi regularly serves as a consultant to industry and government agencies in the U.S. and abroad. He has given over 100 keynote/plenary talks. Prof. Trivedi is a Fellow of the IEEE, The International Association for Pattern Recognition, and The International Society for Optical Engineers. He has served as the Editor-in-Chief of the *Machine Vision and Applications*, editor of *Image and Vision Computing*, and is currently a Senior Editor of the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS. He served as the General Chair of the *IEEE Intelligent Vehicles Symposium* 2010. His team designed and deployed the Eagle Eyes system on the U.S.Mexico border in 2006. He served as a Charter Member and the Vice Chair of the Executive Committee of the University of California (UC) system-wide UC Discovery Program and as an Expert Panelist for the Strategic Highway Research Program of the Transportation Research Board of the National Academies. He has been a recipient of the Distinguished Alumnus Award from Utah State University, the Pioneer Award and the Meritorious Service Award from the IEEE Computer Society, the IEEE ITSS Lead Institution Award and the IEEE ITS Society's highest honor, the "Outstanding Research Award", and several best paper awards.