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

Eshed Ohn-Bar and Mohan Manubhai Trivedi

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 interact: 1) inside the vehicle cabin, 2) around the vehicle, and 3) inside surrounding 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
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 to 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 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).

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

**Table Color Legend:**
- Studying humans inside cabin.
- Studying humans around vehicles.
- Studying humans in surround vehicles.
- Studying humans inside cabin and in surround vehicles.
- 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).

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

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

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

Table of contents suggests 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 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
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
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.

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...
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 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 handheld 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 annotations 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

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


ficial Intelligence, 2008, pp. 1433–1438.


Pattern Recognition Workshops, 2010.

- Investigating the relationships between gaze patterns, dynamic vehicle surrogates, and driver intentions,” in IEEE Intelligent Vehicles Symposium, June 2009.


- E. Ohn-Bar and M. M. Trivedi, “Are all objects equal? deep spatio-

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 interaction, 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.