IEEE Open Journal of
Intelligent Transportation Systems

Received 11 March 2021; revised 11 June 2021 and 20 August 2021; accepted 22 August 2021. Date of publication 31 August 2021; date of current version 20 September 2021.

Digital Object Identifier 10.1109/OJITS.2021.3109039

A Review of UTDrive Studies: Learning Driver Behavior From Naturalistic Driving Data

YONGKANG LIU (Student Member, IEEE), AND JOHN H. L. HANSEN (Fellow, IEEE)

CRSS-UTDrive Lab, Department of Electrical and Computer Engineering, University of Texas at Dallas, Richardson, TX 75080, USA

CORRESPONDING AUTHOR: J. H. L. HANSEN (e-mail: john.hansen@utdallas.edu)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by IRB.

ABSTRACT Intelligent vehicles and Advanced Driver Assistance Systems (ADAS) are being developed rapidly over the past few years. Many applications such as vehicle localization, environment perception, and path planning have shown promising potentialities. While there is great interest in migrating from complete human-controlled vehicles towards fully autonomous vehicles, it is natural that researchers spending more effort trying to understand the interaction between vehicles with various levels of automation in large-scale traffic scenarios. Next-generation vehicles are expected to have the capacity of evaluating driver conditions, vehicle capabilities, surrounding traffic contexts, and take advantage of such knowledge to ensure safe and efficient driving. Three general research questions are raised to achieve this goal, which are (i) how can we acquire sufficient data, (ii) how to evaluate and understand driving behavior, and (iii) how to deliver information effectively to drivers. In this article, we present a review of previous studies from the UTDrive project attempts to answer above questions.

INDEX TERMS Driver behavior modeling, driving performance assessment, mobile platform advancement, naturalistic driving, intelligent vehicle.

I. INTRODUCTION

DRIVING is a complex multi-task process that involves extensive human-machine interaction such as monitoring the environment and surrounding vehicle conditions, predicting other driver's movement and potential risks, determining the best action of their vehicle, and executing the maneuver by controlling the gas/brake pedals and steering wheels to ensure safety. This results in an increased high standard for drivers to operate a vehicle safely on the road. With the goal of improving road safety and establishing an efficient intelligent transportation system, significant efforts have been made across different research focuses and fields. Therefore, a number of autonomous applications have shown promising potentialities such as ego-vehicle localization, trajectory prediction, object detection, end-to-end learning, and connected mobility systems.

According to a National Highway Traffic Safety Administration (NHTSA) report, human errors are the most common factor in road accidents [1]. Modeling and understanding driver conditions through their driving behavior are, therefore, key elements to ensure safe driving. Especially in this transition era, vehicles have gradually been equipped with various levels of automation systems. In the near future, intelligent vehicles with different automation levels and driver engagements will likely be mixed to form large-scale diverse traffic environments. Greater research and understanding are needed regarding the vehicle and driver monitoring in these mixed assistive driving scenarios to improve driving safety. By using the vehicle's onboard sensors or combine information through vehicle communication, next-generation intelligent vehicles should have the ability to evaluate and understand the driver's status, performance, and driving behavior. As a result, such systems could warn of protentional risks, provide guidance when necessary (e.g., lane level guidance), and make essential adjustments or actions when critical.

To achieve this, three general research questions could be raised. Which are (i) how can we acquire sufficient data, (ii) how to evaluate and understand driving behavior, and (iii) how to deliver information effectively to drivers without introducing added distraction. The purpose of this article is not to provide a comprehensive presentation of all significant

The review of this article was arranged by Associate Editor Chongfeng Wei.

progress on the topic of driver behavior analysis. Instead, we present a review of studies from the UTDrive project that attempts to answer above questions.

The rest of the paper is organized as follows: Section II briefly introduces efforts in the naturalistic driving data collection. Section III reviews several researches from the UTDrive project aiming to evaluate and understand driving behavior, while a study of providing visual guidance to the driver is presented in Section IV. Section V summaries the paper with future study directions.

II. DATASETS FOR DRIVER BEHAVIOR LEARNING

Naturalistic driving data are important and indispensable resources for driver behavior learning and understanding. Unlike other autonomous driving tasks (e.g., object detection, vehicle tracking, trajectory prediction, etc.) that have many well-annotated open-source datasets such as KITTI [3], nuScenes [4], Argoverse [5], Waymo [6], Lyft L5 [7] and BDD100K [8]. The number of datasets specifically designed for driver behavior learning/understanding is limited. To the best of the authors' knowledge, there is no clear definition of naturalistic driving data. A strict definition of naturalistic driving data is that data should be collected using participants' familiar vehicles with several cameras and sensors installed, which capture vehicle maneuvers and driver behaviors in an unobtrusive style. However, such data collection will be costly and time-consuming to practice, as well as complex to coordinate. Therefore, only a few datasets fit this definition such as Strategic Highway Research Program 2 (SHRP 2) data [9], 100-car study data [10], and Candrive study data [11]. More datasets will be in line with naturalistic driving data if we define it in a more general way, in which the participants were performing naturalistic driving and data were collected using instrument vehicles. For example, the Brain4Cars [12] dataset consists of natural driving videos with both inside and outside views of the car, its speed, and the Global Position System (GPS) coordinates collected by 10 drivers, the Berkeley DeepDrive Video dataset (BDDV) [13] comprised of real driving videos and GPS/Inertial Measurement Unit (IMU) data, the Honda Research Institute Driving Dataset (HDD) [14] includes 104 hours of real human driving in the San Francisco Bay Area collected using an instrumented vehicle equipped with different sensors, and the Drive&Act [15] dataset focused on driver behavior recognition.

Collected by the Center for Robust Speech Systems (CRSS) - UTDrive lab since 2006, The UTDrive naturalistic driving dataset is under an international collaboration to build a large-scale driving data corpus that can be used for a wide range of research activities related to driving behavior [2]. In total, more than 500 drivers have been collected in three different countries, using very similar data collection vehicles, sensors, routes, and secondary tasks, which for the first time would allow cross-continent-based comparisons of driver distraction/behavior research advancements.

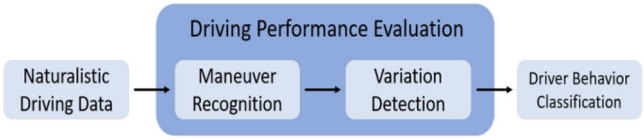

FIGURE 1. A generalized system framework for driving behavior classification.

Studies of the UTDrive naturalistic driving dataset have resulted in a large number of researches focused on driving behavior learning and evaluation. Examples of such studies from the UTDrive project will be reviewed in the following section.

III. UNDERSTANDING DRIVER BEHAVIOR AND EVALUATE DRIVING PERFORMANCE FROM VEHICLE DYNAMIC SIGNALS

Driving maneuvers are not only important components in understanding driver behavior but also basic units of a completed driving session. Therefore, maneuver analysis is essential when processing large amounts of naturalistic driving data. Understanding how and why these maneuvers are performed can provide knowledge about how well the driver controls the vehicle, how driving performance varies over time, or the driver's mental workload and status.

From the vehicle control's perspective, a driver's primary physical contacts in the vehicle are the steering wheel and gas/brake pedals. Any factors that may influence their driving performance (e.g., distraction, driving experience, environmental context, etc.) has a direct impact on body movements which will then affect the control of the vehicle. Hence, one typical approach for the development of ADAS is to identify risky driving maneuvers by analyzing abrupt variations in vehicle dynamic signals. Fig. 1 demonstrates a generalized framework of the driving classification system. For every driving session, maneuvers are first detected and identified. Next, variations against normal patterns (i.e., neutral driving) for each maneuver are calculated. These variations can be utilized to classify the driver behavior (e.g., normal driving vs distracted driving) or evaluate the driving performance. Naturalistic driving data is preferred because it is important to analyze driver's driving maneuvers and reactions in real-world traffic scenarios.

In the remaining of this section, five studies from the UTDrive project will be introduced. Two of them are focused on lane-change activity, the third one is exploring driving event recognition using data collected from the smartphone, followed by a driving performance analysis study based on different levels of driving experience, vehicle familiarity, and route familiarity. In the end, we present the latest effort on improving our Mobile-UTDrive App functionality to support real-time driver/driving behavior measurement and visualization. Please note that only highlighted aspects of our studies are introduced, we encourage readers to refer to our previous paper [16], [17], [19], [41] for more detailed explanations.

TABLE 1. Results of lane-change detection.

Ground Truth \ Predicted	*LCL*	*LK*	*LCR*
Left Lane-Change (LCL)	**80.36%**	8.04%	11.61%
Lane Keeping (LK)	1.41%	**91.90%**	6.69%
Right Lane-Change (LCR)	6.04%	10.74%	**83.22%**

A. LANE-CHANGE DETECTION FROM STEERING SIGNAL

Lane changing is an essential driving action that is executed enormous times every day and has been one of the most common reasons for accidents. The objective of this study is to detect lane-change maneuvers using vehicle dynamic signals (i.e., CAN-Bus) [16]. For studies related to lane-change, most early efforts either focused on computer vision algorithms to detect and track lane markings on the road [20], [21], or employed LiDAR and map information for sensor fusion. Curve estimation models [22] and lane type classification [23] have been explored by utilizing LiDAR signals along with images, which provide a better contextual understanding of the road. Digital map data is also be used to render virtual images and achieve lane-level vehicle localization [24]. With the development of deep learning models, many autonomous driving tasks have shown great progress. Reference [25] reviewed recent lane-change detection methods and reporting new results based on deep learning.

The lane-change detection framework introduced in this study is constructed by three stages: (1) data pre-processing for noise reduction and subtraction of turning events, (2) driving maneuver segmentation to divide a long driving data into a sequence of time-variant events, and (3) lane-change maneuver detection using a double-layered HMM architecture.

The purpose of the pre-processing stage is to reduce the effect of noise and to exclude turning events. Next, a time-frequency spectral analysis approach (filterbank analysis) is employed to segment driving maneuvers. The intuition is that variations observed in the frequency domain of CAN-Bus signals can help determine maneuver boundaries [26].

The overall lane-change detection framework is designed as a double-layered HMM [27] architecture. The upper layer will take the complete driving route and consecutive maneuvers into consideration, while the lower layer is focused on the vehicle's dynamical movements inside the individual driving event. Two Gaussian HMMs, one for lane-change detection and the other for lane-keeping detection respectively, will be trained. During testing, the result with higher possibilities will be treated as the detection result of each incoming test segment at the lower layer. The upper layer will then view this result as a hidden state to yield the state sequence in a Discrete HMM as shown in Fig. 2.

Lane-change detection results using this proposed framework are summarized in Table 1. Even with unbalanced distributions in our data set, an 80.36% detection accuracy for left lane-change (LCL) and 83.22% for right lane-change (LCR) was achieved. Most of the detection failures happen when the vehicle speed is low and near intersections, the reasonable explanation is that dynamic characteristics extracted from CAN-Bus signal are not obvious in those scenarios, with additional feature sources (e.g., road information), the detect accuracy under such scenarios are expected to be improved.

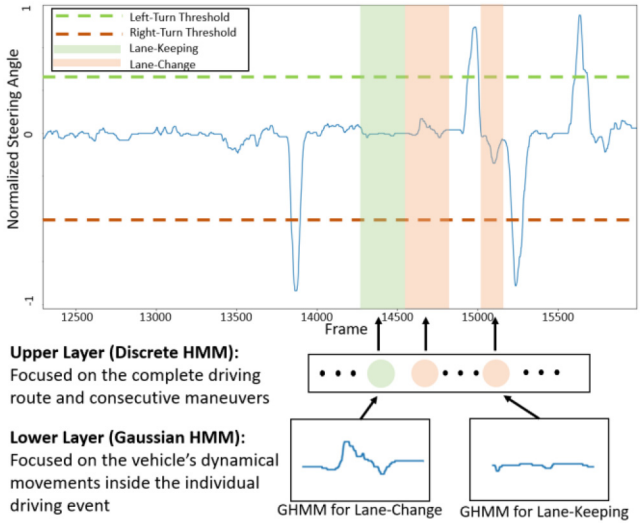

FIGURE 2. Double-layered HMM framework for lane-change detection. The upper layer yields the state sequence in a Discrete HMM, lower layer is two Gaussian HMMs trained for lane-change detection and lane-keeping detection respectively.

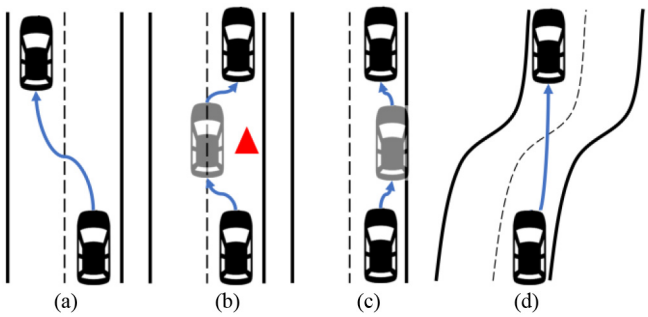

FIGURE 3. Examples of various lane-change related events appeared in naturalistic driving scenarios. (a) a typical lane-change event, (b) a swerve shift which is to avoid obstacles on road, (c) a lane-drift event followed by a correction, (d) a lane-change event on a curve road.

Although contextual information could be obtained and fused from various sensors, the accurate detection of lane-change maneuvers remains a challenging research problem. Difficulties that limit the performance of lane-change detection in realistic driving scenarios could be the presence of road curve, vehicle judder due to the road condition, or fluctuation in drivers. As illustrated in Fig. 3, case (a) represents an example of a typical left lane-change (LCL) event. In case (b) (i.e., avoidance lane-keeping), although appears to be a lane-change event from vehicle dynamic data, the vehicle is actually making a move to avoid obstacles on the

FIGURE 4. Classifying short Lane Change (LC) video clips with transferred segmentation masks as risky or safe. The average duration of these clips is ∼ 10 seconds. 0.937 AUC score was achieved with the proposed method [17].

road. Case (c) usually happens when the driver is fatigued or distracted, the vehicle will drift towards another lane followed by a correction maneuver. In case (d), the driver may perform a lane change while driving on a curve road, the steering angle will hardly change in this case, making it almost impossible to detect if using CAN-Bus signal alone. Therefore, sensor fusion and multimodality detection under such cases will be a valuable future research direction.

B. RISKY ACTION RECOGNITION IN LANE CHANGE VIDEO CLIPS

In this work, we designed a deep spatiotemporal classification network that uses pre-trained state-of-the-art segmentation network Mask R-CNN [28] as its spatial feature extractor for classifying dangerous lane change behavior in short video clips captured by a monocular camera [17]. The results demonstrate the adaptive capabilities of deep learning, reinforce the claim that with the increasing availability of pre-trained high-performance deep learning models, new problems can be addressed without collecting extensive dedicated datasets for them.

We proposed a novel framework for binary video classification in this work. First, masked images obtained by the Mask R-CNN network were passed through convolutional layers to extract abstract features from the contrasted compositions created by the masks. These high-level features were then fed into Long Short-Term Memory (LSTM) cells to depict temporal relationships. The intuition of using such overlaid image as input is that a temporal composition with highly contrasting elements can be more helpful to recognize distinct elements. As shown in Fig. 4, the masked image sequence relays a more striking version of the lane change action than the raw sequence. Therefore, features extracted from such images are expected to contain more information to help detect risky lane-change maneuvers.

To evaluate the performance of our proposed framework, a subset of the NUDrive [2] dataset consists of 860 lane change video clips was used in the experiment. Ten annotators were asked to watch the video clips and give a risk level rate of each instance subjectively. Risk ratings were first normalized for each annotator, then, the normalized scores were averaged to obtain a single score per lane change. The riskiest 5% of the lane change instance was labeled as risky while the rest was assumed to be safe. The final distribution is 43 to 817 for the positive and negative classes respectively. More details of the experiments and different frameworks tested can be found in [17].

The best evaluation result (0.937 AUC score) was obtained with our proposed framework which used a Mask R-CNN semantic mask extractor. We believe this result was because of the masked-contrasted temporal compositions' aptitude for relaying semantic information. Our experiments also demonstrate the adaptive capabilities of deep learning. With the increasing availability of high-performance deep learning models, new problems can be tackled without collecting huge datasets for them.

C. DRIVING EVENT RECOGNITION USING FREE-POSITIONED PORTABLE DEVICE

There is a great influence on the automotive industry with the rapid development of portable device applications. Various sensors such as cameras, microphones, IMU (e.g., accelerometer, gyroscope, etc.), and GPS are integrated into smartphones. Therefore, data collected and processed by such sensors could be utilized to deliver a comprehensive description of driving scenarios. A lot of efforts have been made to transfer portable devices to an alternative platform capable of naturalistic driving data collection [29], distraction and drowsy alert [30], driver behavior analyzing [31], [32], aggressive/dangerous driving detection [33], [34] or to provide auxiliary driving guidance [35]. Moreover, researchers also employ such a device's ability to connect to the network, providing services like traffic notification [36], remote diagnostics [37], [38], vehicle-to-vehicle (V2V) communication [39], and other applications in intelligent transportation system field [40].

In the past few years, the Mobile-UTDrive App designed as a multi-modal data collection platform has been developed by our lab, which is aiming to collect driver, vehicle, and environmental context that describing the comprehensive driving scenario. To demonstrate the potential mobile platform advancements, a driving event recognition experiment with our previous proposed data pre-processing framework, which allows for a free-positioned device for onboard sensing will be introduced [18], [41].

One challenge for in-vehicle sensing platform development is the relative movements and orientation difference when the smartphone is located inside the vehicle. Our previous study [41] proposed a framework to overcome this challenge. The first step is a geometry coordinate transformation step taking raw smartphone data as input. This step will rotate/re-orientate the smartphone-referenced acceleration data to the vehicle-referenced coordinate system. Since the smartphone pose is unknown, it is difficult to determine which axis is corresponded with the vehicle's longitudinal, lateral, or vertical movement. Therefore, an additional axis

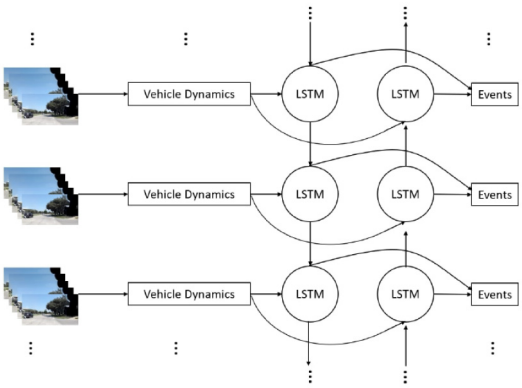

FIGURE 5. Inspired by the sequence-tagging problem in the NLP field, we adopted the Bi-LSTM network for driving event reconnection. A sequence of vehicle dynamic data from the mobile platform will replace the word embedding vector as the input, and their word contextual representation output will correspond with our driving event labels.

TABLE 2. Driving event recognition results using raw vehicle sensor data.

Predicted / Ground Truth	*LCL*	*LK*	*LCR*	*TNL*	*TNR*
LK	**49.39%**	26.53%	16.33%	1.22%	6.53%
LCL	5.12%	**88.46%**	6.41%	0	0
LCR	8.93%	8.93%	**82.14%**	0	0
TNL	2.77%	0	0	**94.44%**	2.78%
TNR	0	0	0	3.33%	**96.67%**

TABLE 3. Driving event recognition results using processed smartphone data.

Predicted / Ground Truth	*LCL*	*LK*	*LCR*	*TNL*	*TNR*
LK	**49.40%**	32.93%	11.24%	3.61%	2.81%
LCL	10.46%	**81.40%**	6.98%	0	1.16%
LCR	1.79%	26.79%	**71.43%**	0	0
TNL	0	0	0	**97.30%**	2.70%
TNR	0	0	0	3.85%	**96.15%**

alignment step is performed by comparing the normalized acceleration values and GPS speed and bearing derivatives, with the assumption that vehicle vertical acceleration should be close to the gravity (i.e., $g \approx 9.8m/s$). Next, we map the interrelation of IMU and GPS data by a regression model. An adaptively filtering process in the last step is adopted to decouple the smartphone's relative movement while in the vehicle. For future applications, this framework could be employed as a pre-processing module and provides the starting point of succeeding data processing.

The pre-processed data are then used to divide and classify driving sequences into five groups: Lane-Keeping (LK), Left Lane-Chang (LCL), Right Lane-Change (LCR), Left-Turn (TNL), and Right-Turn (TNR). We treat our driving event recognition task as to assign labels to each entity within a sequence, which is similar to the sequence-tagging problem in Natural Language Processing (NLP) field. Therefore, we adopted a Bidirectional LSTM network [42], [43] illustrated in Fig. 5. A sequence of vehicle dynamic data from the mobile platform will replace the word embedding vector as the input, and their word contextual representation output will correspond with our driving event labels.

The long sequence driving data will first be segmented by sliding through a fixed time window (e.g., 0.1 seconds). Next, vehicle dynamic signals were concatenated to form the input vector. Output labels were assigned to each frame for the individual event. A total of 102 driving sequences were collected for this experiment. Comparable recognition results from Table 2 and Table 3 demonstrated the effectiveness of this framework. Real-time processing and communication capability are under development for better sensing and understanding of driver behavior.

D. DRIVING PERFORMANCE ANALYSIS AND ASSESSMENT

Driving often involves four key steps: surrounding monitoring, predicting, decision making and maneuver executing. Many factors could affect these steps and influencing a driver's driving performance and safety. Take the driving experience as an example, experienced drivers will have a better capability in situation assessment or hazard perception, so that they could earn them enough reaction time by predicting dangerous movements from surrounding environments in advance. Safety hazards could also be introduced by not familiar with the vehicle. Drivers may have trouble executing maneuvers in time-critical scenarios because the throttle response is not as sensitive as the driver's own vehicle. Or the driver may have difficulty in some instrumentation controls because the design and layout are different, which could cause distracted driving.

In this study [19], to analyze how the driver experience impacts their driving performance, a group of 20 subjects participated in our experiment by drove the UTDrive instrumented vehicle. We divided our participants into two groups: the novice driver group which has less than one year's driving experience, and the experienced driver group possesses an average of four years of driving experience.[1]

When trying to evaluate a driver's driving performance, it is very subjective because individuals have their own criteria. Therefore, one typical approach is to identify the driving event and then quantify if there is any deviation from the expected or "neutral" behavior. We employed a clustering-outlier detection grading method using the Support Vector Machine (SVM) model with the assumption that an experienced driver should express a stable vehicle dynamical data with low variance during normal driving sessions. In this method, driving events sharing the most common characteristics and classified in the innermost layer will be considered as normal/safe events and labeled with grade "A". Similarly, risky driving events will be outliers and labeled with a poor grade. We then calculate the "Grade Point Average (GPA)" score to provide a performance overview of the driving

1. IRB 06-19 approved by the Office of Research Integrity and Outreach, The University of Texas at Dallas.

TABLE 4. Quantitative analysis results of driving performance.

Driver Groups	*Forward (FWD)*	*Turning (TURN)*	*Acceleration (GAS)*	*Deceleration (BRK)*	*Overall*
			GPA		
Novice	3.432	3.413	3.033	2.943	3.247
Experienced	3.431	3.4256	3.115	3.001	**3.280**
			Average Discrepancy Score		
Novice	26.332	25.615	24.458	24.652	25.264
Experienced	22.921	22.122	21.147	21.737	**21.982**

session. Additionally, feature vectors extracted from two selected drivers' data were utilized as a "good driver" baseline to calculate the Euclidean distance between them and every other driver as discrepancy scores. A small discrepancy score implies that the subject behaves close to the experienced driver baseline, thus reflect a better driving performance.

Quantitative grading results are summarized in Table 4. A 10.16% GPA score gap between the two driver groups was observed. Both overall GPA score and discrepancy score suggested that the experienced driver group performs better when environmental conditions and vehicle familiarity are close. More experiments details and results were discussed in the original paper [19]. In this study, the ability to analyze and assess driving performance using vehicle dynamic data was demonstrated. For the next step, it would be worthwhile to refine baselines with a large amount of naturalistic driving data collected in different scenarios and test this in field operations to improve performance for future intelligent vehicle advancements.

E. ENHANCING MOBILE-UTDRIVE APP CAPACITY FOR ONBOARD DRIVER ASSESSMENT

One limitation of our previously developed Mobile-UTDrive App is the lack of real-time driving behavior analysis and visualization functions. Therefore, it can only serve as a mobile multi-modal data collection platform. Collected data have to be examined offline to provide driving behavior information.

The goal of this work is to expand the Mobile-UTDrive App's functionality to support real-time driver/driving behavior measurement and visualization. Parameters used to describe driving behavior in the current stage are speed, acceleration and deceleration obtained from smartphone sensors. We define the neutral state as drivers drive in a relatively conservative manner where no abnormal driving behavior is detected. Warning/moderate state is a transition state, which means drivers still drive under the limits, however, it should get some attention to monitor if the driving behavior will break through the limits. An aggressive state means aggressive or risky driving behavior is detected. The nearby traffic and pedestrians should be noticed. Based on our previous experiences, we set the evaluation interval as 1 second and acceleration/deceleration thresholds as 1 m/s^2, 2 m/s^2 and 4 m/s^2, which are corresponding to abovementioned

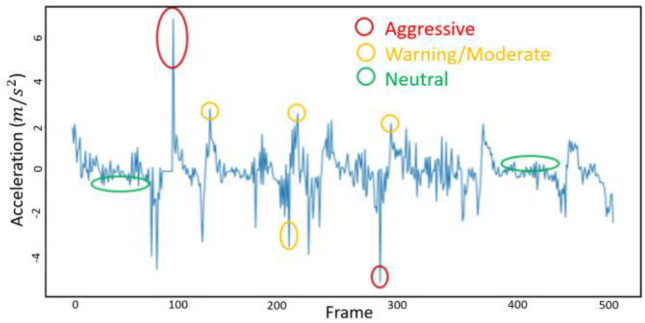

FIGURE 6. Driving sessions are categorized into neutral, warning/moderate, and aggressive states based on pre-defined acceleration thresholds.

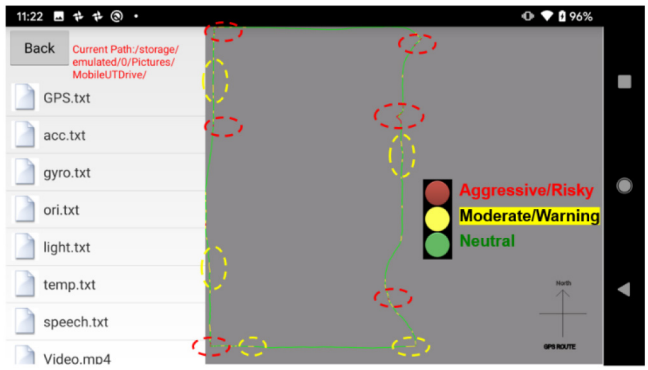

FIGURE 7. Screenshot of the driving behavior visualization from one field test. Green segments indicate neutral driving state, yellow and red correspond to warning/moderate and aggressive/risky driving states.

driving states. The driving performance score is also calculated using our previous algorithm. After each driving session, users could easily access the results in their devices. Fig. 6 and Fig. 7 illustrate the categorized driving states and visualization from one field test.

We believe this can provide users a straightforward understanding of their instantaneous state of driving behavior in detail. Typically, a neutral driving state can occur more than 80% of a driving session, while the remaining portions may result in crash or near-crash events. Therefore, the future direction of this study is to explore more accurate (e.g., deep learning-based, personalized) driver behavior prediction models, and implement them in our Mobile-UTDrive App to act as a virtual co-pilot to help reduce near-crash and crash events.

IV. ESTABLISHING DRIVER VISUAL GUIDANCE FOR IMPROVED SAFETY AND EFFICIENCY

After the intelligent vehicle or the cloud system predicted future risky maneuvers (e.g., aggressive/distracted driving) of neighboring vehicles or the ego-vehicle itself from learned driver behavior models, it is important to deliver warning or guidance information to drivers for safe and efficient driving. To achieve this goal, Augmented Reality (AR) with Head-Up Displays (HUDs) has attracted attention in the field of automotive research. Such AR-HUD systems have been

FIGURE 8. Multi-lane highway simulation environment built in Unity, where the furthest red vehicle is our ego-vehicle.

studied to supporting driver visual guidance or providing lane-level guidance services [44], [45], [46], [47]. One of our previous efforts trying to fusion vision and cloud data to display surrounding vehicles' lane-change probability will be briefly introduced in this section.

In this study [48], we assume that every vehicle in the driving scene has access to the Internet. The cloud system will run a lane-change prediction model based on data received from connected vehicles and the ego-vehicle is equipped with both RGB and depth camera. However, a key research issue is how to correctly overlay the cloud information received through the vehicle-to-cloud (V2C) communication onto the correct target vehicle. From a secondary viewpoint, one would also need to consider the ego vehicle's perspective, which motivates a second question on how to identify the target vehicle whose information has been shared [49].

Four key modules are included in our proposed architecture. The coordinate transformation module and the object detection module will output the target vehicle anchor point and bounding boxes information regarding target vehicles. The depth evaluation module will provide additional distance knowledge when necessary. This knowledge will assist the distance matching module to identify the correct target vehicle. Finally, predicted information from the cloud will be visualized to provide driver guidance. Details of each module are presented in [48].

We adopted the Unity game engine [50] to conduct modeling and evaluation of the proposed data fusion system, given its strengths in visualization, graphics design, as well as external joystick (i.e., driving simulator) integration. In the simulation, a multi-lane highway environment is built in Unity, illustrated in Fig. 8. An accident scene is designed in this environment, where two stopped trucks occupy the two right lanes (along their forward direction). When vehicles are approaching this accident zone, some of them might make left lane change actions to avoid a potential collision, and in turn create conflicts to the motion trajectory of the ego-vehicle (the furthest red vehicle). The lane change prediction model running on the cloud is therefore supposed to provide prediction information to the ego-vehicle's driver. An Augmented Reality (AR)-based information visualization approach is implemented in Unity to display the predicted lane change probability to the driver as a head-up display of the ego-vehicle.

We compared both scenarios (with and without the use of the proposed system) in terms of driving safety and comfort in the Human-in-the-Loop (HITL) simulation experiment. Three different measurement factors are compared in both scenarios: 1) average time-to-collision (TTC) between the ego-vehicle and the lane-change vehicle, 2) average absolute acceleration of the ego-vehicle, and 3) maximum jerk of the ego-vehicle during the trip. We observed a notable increase in average TTC value after implementing the proposed model. The increase of 30.31% means the driver of the ego-vehicle tended to keep further away from the lane-change vehicle when the visual guidance was provided. It is suggested that a larger TTC value can induce safer driving performance, and prevent potential rear-end collisions in such emergencies.

We also noticed that the average absolute acceleration decreases from 2.57 m/s^2 to 2.07 m/s^2 after the proposed vision-cloud data fusion model is implemented/deployed, as well as a decrease of 10.54% in the maximum jerk value.

Indicating that the proposed model can introduce a more comfortable and smooth driving/riding experience for the driver/passengers.

V. DISCUSSION AND CONCLUSION

The CRSS-UTDrive Lab has focused our researches on naturalistic driving studies with the motivation of contributing to improved intelligent driver-vehicle systems and the goal of driver behavior understanding from multiple modalities. In this article, we have reviewed several previous studies from the UTDrive project trying to address aforementioned research questions. Which are (i) how can we acquire sufficient data, (ii) how to evaluate and understand driving behavior, and (iii) how to deliver information effectively to drivers without introducing added distraction. We believe these studies, along with the fast-growing smartphone applications and cloud-connected services, would not only be able to help with the understanding of the driving scenario, driving performance variance, and drivers' mental conditions, but also help to establish a safe and efficient intelligent transportation system by utilizing information shared via vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications.

There are great progressives of vehicle technologies with a substantial amount and scope of research/development carried by researchers and manufacturers. Vehicle technologies and vehicle driving autonomy will keep developing as we

move forward into the next generation of intelligent vehicles and intelligent transportation systems. Next-generation intelligent vehicles will play a more significant role in a collaborative driver-vehicle engagement environment. Therefore, modeling and understanding of driver behavior, vehicle interaction, and the influence of environmental context from a large amount of naturalistic driving data reaming a key element to ensure safe driving.

REFERENCES

[1] S. Singh, *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey*, document DOT HS 812 115, Nat. Highway Traffic Safety Admin., Washington, DC, USA, 2015.

[2] K. Takeda, J. H. L. Hansen, P. Boyraz, L. Malta, C. Miyajima, and H. Abut, "International large-scale vehicle corpora for research on driver behavior on the road," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1609–1623, Dec. 2011.

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

[4] H. Caesar *et al.*, "nuScenes: A multimodal dataset for autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 11621–11631.

[5] M. Chang *et al.*, "Argoverse: 3D tracking and forecasting with rich maps," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 8748–8757.

[6] *Waymo Open Dataset: An Autonomous Driving Dataset*, Waymo, Mountain View, CA, USA, 2019.

[7] R. Kesten *et al.* (2019). *Lyft Level 5 AV Dataset 2019*. [Online]. Available: https://level5.lyft.com/dataset/

[8] F. Yu *et al.*, "BDD100K: A diverse driving dataset for heterogeneous multitask learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 2636–2645.

[9] *Strategic Highway Research: Saving Lives, Reducing Congestion, Improving Quality of Life*, vol. 260, Transp. Res. Board, Washington, DC, USA, 2001.

[10] V. L. Neale, S. G. Klauer, R. R. Knipling, T. A. Dingus, G. T. Holbrook, and A. Petersen, "The 100 car naturalistic driving study, phase I-experimental design," Nat. Highway Traffic Safety Admin., Washington, DC, USA, Rep. HS-809 536, 2002.

[11] S. C. Marshall *et al.*, "Protocol for Candrive II/Ozcandrive, a multicentre prospective older driver cohort study," *Accid. Anal. Prevent.*, vol. 61, pp. 245–252, Dec. 2013.

[12] 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," 2016. [Online]. Available: arXiv:1601.00740.

[13] H. Xu, Y. Gao, F. Yu, and T. Darrell, "End-to-end learning of driving models from large-scale video datasets," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 2174–2182.

[14] V. Ramanishka, Y. Chen, T. Misu, and K. Saenko, "Toward driving scene understanding: A dataset for learning driver behavior and causal reasoning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 7699–7707.

[15] M. Martin *et al.*, "Drive&Act: A multi-modal dataset for fine-grained driver behavior recognition in autonomous vehicles," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 2801–2810.

[16] Y. Zheng and J. H. L. Hansen, "Lane-change detection from steering signal using spectral segmentation and learning-based classification," *IEEE Trans. Intell. Veh.*, vol. 2, no. 1, pp. 14–24, Mar. 2017.

[17] E. Yurtsever *et al.*, "Risky action recognition in lane change video clips using deep spatiotemporal networks with segmentation mask transfer," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 3100–3107.

[18] Y. Liu, Y. Zheng, and J. H. L. Hansen, "1 MobileUTDrive: A portable device platform for in-vehicle driving data collection," in *Vehicles, Drivers, and Safety*. Berlin, Germany: De Gruyte, May 2020, pp. 3–24.

[19] Y. Liu and J. Hansen, "Analysis of driving performance based on driver experience and vehicle familiarity: A UTDrive/mobile-UTDrive app study," *SAE Int. J. Transp. Safety*, vol. 7, no. 2, pp. 175–190, Nov. 2019.

[20] J. C. McCall and M. M. Trivedi, "Video-based lane estimation and tracking for driver assistance: Survey, system, and evaluation," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 1, pp. 20–37, Mar. 2006.

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

[22] A. S. Huang, D. Moore, M. Antone, E. Olson, and S. Teller, "Finding multiple lanes in urban road networks with vision and lidar," *Auton. Robots*, vol. 26, nos. 2–3, pp. 103–122, Mar. 2009.

[23] X. Gu, A. Zang, X. Huang, A. Tokuta, and X. Chen, "Fusion of color images and LiDAR data for lane classification," in *Proc. 23rd SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, Nov. 2015, p. 69.

[24] G. Cao, F. Damerow, B. Flade, M. Helmling, and J. Eggert, "Camera to map alignment for accurate low-cost lane-level scene interpretation," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2016, pp. 498–504.

[25] D. Liang, Y.-C. Guo, S.-K. Zhang, T.-J. Mu, and X. Huang, "Lane detection: A survey with new results," *J. Comput. Sci. Technol.*, vol. 35, pp. 493–505, May 2020.

[26] A. Sathyanarayana, S. O. Sadjadi, and J. H. L. Hansen, "Automatic maneuver boundary detection system for naturalistic driving massive corpora," *SAE Int. J. Passenger Cars Electron. Electr. Syst.*, vol. 7, no. 1, pp. 149–156, Apr. 2014.

[27] Z. Ghahramani, "An introduction to hidden Markov models and Bayesian networks," in *Hidden Markov Models: Applications in Computer Vision*. Singapore: World Sci., 2001, pp. 19–41.

[28] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2017, pp. 2980–2988.

[29] M. Reininger, S. Miller, Y. Zhuang, and J. Cappos, "A first look at vehicle data collection via smartphone sensors," in *Proc. IEEE Sens. Appl. Symp. (SAS)*, 2015, pp. 1–6.

[30] C.-W. You *et al.*, "CarSafe demo: Supporting driver safety using dual-cameras on smartphones," in *Proc. ACM Conf. Ubiquitous Comput.*, 2012, p. 547.

[31] N. Kalra and D. Bansal, "Analyzing driver behavior using smartphone sensors: A survey," *Int. J. Electron. Electr. Eng.*, vol. 7, no. 7, pp. 697–702, 2014.

[32] F. Lindow and A. M. Kashevnik, "Driver behavior monitoring based on smartphone sensor data and machine learning methods," in *Proc. 25th Conf. Open Innovat. Assoc. (FRUCT)*, Nov. 2019, pp. 196–203.

[33] R. Chhabra, S. Verma, and C. R. Krishna, "Detecting aggressive driving behavior using mobile smartphone," in *Proc. 2nd Int. Conf. Commun. Comput. Netw.*, 2019, pp. 513–521.

[34] F. Li, H. Zhang, H. Che, and X. Qiu, "Dangerous driving behavior detection using smartphone sensors," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, 2016, pp. 1902–1907.

[35] R. Murugesh, U. Ramanadhan, N. Vasudevan, A. Devassy, D. Krishnaswamy, and A. Ramachandran, "Smartphone based driver assistance system for coordinated lane change," in *Proc. Int. Conf. Connected Veh. Expo (ICCVE)*, 2015, pp. 385–386.

[36] C. Campolo, A. Iera, A. Molinaro, S. Y. Paratore, and G. Ruggeri, "SMaRTCaR: An integrated smartphone-based platform to support traffic management applications," in *Proc. 1st Int. Workshop Veh. Traffic Manage. Smart Cities (VTM)*, Nov. 2012, pp. 1–6.

[37] G. F. Türker and A. Kutlu, "Survey of smartphone applications based on OBD-II for intelligent transportation systems," *Int. J. Eng. Res. Appl.*, vol. 6, no. 1, pp. 69–73, 2016.

[38] V. Kirthika and A. K. Vecraraghavatr, "Design and development of flexible on-board diagnostics and mobile communication for Internet of Vehicles," in *Proc. Int. Conf. Comput. Commun. Signal Process. (ICCCSP)*, 2018, pp. 1–6.

[39] K.-C. Su, H.-M. Wu, W.-L. Chang, and Y.-H. Chou, "Vehicle-to-vehicle communication system through Wi-Fi network using android smartphone," in *Proc. Int. Conf. Connected Veh. Expo (ICCVE)*, Dec. 2012, pp. 191–196.

[40] J. Wahlström, I. Skog, and P. Händel, "Smartphone-based vehicle telematics: A ten-year anniversary," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 10, pp. 2802–2825, Oct. 2017.

[41] Y. Zheng, N. Shokouhi, A. Sathyanarayana, and J. Hansen, "Free-positioned smartphone sensing for vehicle dynamics estimation," SAE Int., Warrendale, PA, USA, Technical Paper, Mar. 2017.
[42] G. Lample, B. Miguel, S. Sandeep, K. Kazuya, and D. Chris, "Neural architectures for named entity recognition," 2016. [Online]. Available: arXiv:1603.01360.
[43] X. Ma and E. H. Hovy, "End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguist.*, vol. 1, 2016, pp. 1064–1074.
[44] J.-W. Lee, C.-R. Yoon, J. Kang, B.-J. Park, and K.-H. Kim, "Development of lane-level guidance service in vehicle augmented reality system," in *Proc. 17th Int. Conf. Adv. Commun. Technol. (ICACT)*, 2015, pp. 263–266.
[45] L. Abdi and A. Meddeb, "Driver information system: A combination of augmented reality, deep learning and vehicular ad-hoc networks," *Multimedia Tools Appl.*, vol. 77, no. 12, pp. 14673–14703, 2018.
[46] J. Lin, C. Lin, C. Dow, and C. Wang, "Design and implement augmented reality for supporting driving visual guidance," in *Proc. 2nd Int. Conf. Innovat. Bio-Inspired Comput. Appl.*, 2011, pp. 316–319.
[47] V. Charissis *et al.*, "Employing emerging technologies to develop and evaluate in-vehicle intelligent systems for driver support: Infotainment AR HUD case study," *Appl. Sci.*, vol. 11, no. 4, p. 1397, 2021.
[48] Y. Liu, Z. Wang, K. Han, Z. Shou, P. Tiwari, and J. H. L. Hansen, "Sensor fusion of camera and cloud digital twin information for intelligent vehicles," in *Proc. IEEE Intell. Veh. Symp. (IV)*, 2020, pp. 182–187.
[49] H. Kobayashi, K. Han, and B. Kim, "Vehicle-to-vehicle message sender identification for co-operative driver assistance systems," in *Proc. IEEE 89th Veh. Technol. Conf. (VTC)*, 2019, pp. 1–5.
[50] Unity. (Feb. 29, 2020). *Unity for All*. [Online]. Available: https://www.unrealengine.com

YONGKANG LIU (Student Member, IEEE) received the B.S. degree in electronic information engineering from Shandong Normal University in 2015, and the M.S. and Ph.D. degrees in electrical engineering from the University of Texas at Dallas in 2017 and 2021, respectively. He has been a Ph.D. student of electrical engineering with the University of Texas at Dallas since 2017, working as a Research Assistant with the Center for Robust Speech Systems (CRSS)–UTDrive Lab. His research interests are focused on in-vehicle systems, advancements in smart vehicle technologies, and driver behavior modeling.

JOHN H. L. HANSEN (Fellow, IEEE) received the B.S.E.E. degree from Rutgers University, and the M.S. and Ph.D. degrees in electrical engineering from the Georgia Institute of Technology. He joined University of Texas-Dallas (UTDallas) in 2005, where he is currently an Associate Dean for Research, and previously served as the Department Head of Electrical Engineering. At UT-Dallas, he established the Center for Robust Speech Systems, and oversees the UTDrive Lab, which is focused on driver modeling and distraction. His research interests span the areas of digital speech processing, machine learning for speech and speaker traits, feature estimation in noise, cognitive modeling, and driver distraction analysis for effective human-machine interactions. He is a past elected SLTC Chair and a member of IEEE Signal Processing Society Speech-Language Technical Committee and Educational Technical Committee.