VTD: Visual and Tactile Database for Driver State and Behavior Perception

Jie Wang[1],Mobing Cai[2],Zhongpan Zhu*,*Member,IEEE,*Hongjun Ding,Jiwei Yi and Aimin Du

Abstract— **In the domain of autonomous vehicles, the human-vehicle co-pilot system has garnered significant research attention. To address the subjective uncertainties in driver state and interaction behaviors, which are pivotal to the safety of Human-in-the-loop co-driving systems, we introduce a novel visual-tactile perception method. Utilizing a driving simulation platform, a comprehensive dataset has been developed that encompasses multi-modal data under fatigue and distraction conditions. The experimental setup integrates driving simulation with signal acquisition, yielding 600 minutes of fatigue detection data from 15 subjects and 102 takeover experiments with 17 drivers. The dataset, synchronized across modalities, serves as a robust resource for advancing cross-modal driver behavior perception algorithms.**

Keywords: Visual and tactile data, driver state, driver behavior, intelligent cockpit, autonomous vehicles

I. INTRODUCTION

In recent years, data-driven autonomous vehicles have encountered SOTIF (Safety of the Intended Functionality) issues and long-tail challenges in their AI algorithms. These challenges arise due to the complexity and unpredictability of real-world scenarios where autonomous vehicles must navigate, requiring robust algorithms for safe operation under various conditions. Furthermore, the uncertainty in driver non-driving behavior[1] and abnormal state[2] within the context of human-vehicle co-driving further exacerbates the challenges faced by autonomous vehicles. Perceiving and predicting the driver's uncertain behaviors accurately is crucial for developing effective AI algorithms that can adapt and respond appropriately.

As an individual with fully independent behavior possessing subjective initiative uncertainty and individual discrepancy, drivers may become increasingly reliant on autonomous systems as driving intelligence advances, leading to drowsiness, slower reaction times, and reduced risk perception[1, 2]. Furthermore, the addition of in-vehicle entertainment features introduces a potential risk of distracted driving. Research indicates that driver behavior is a key factor in most road traffic accidents, with fatigue and distraction being the primary causes[3, 4]. Changes in a driver's state increase the uncertainty of driving behavior and raise the risks associated with human-machine interaction, impeding the development of intelligent human-machine hybrid driving modes. Therefore, perceiving and understanding driver behavior and state has become a critical area for research breakthroughs.

To address these challenges, it is essential to collect high-quality datasets on fatigue and distraction, enhancing the ability to detect driver risks in intelligent co-driving systems and reducing potential dangers. These datasets should cover a wide range of driving conditions, human behavioral characteristics, and the specific interaction patterns found in human-machine co-driving scenarios.

The availability of human-vehicle co-driving datasets is crucial for the development and evaluation of artificial intelligence algorithms in autonomous vehicles. However, existing datasets for human-vehicle joint driving applications are still quite limited, especially in monitoring key states such as driver fatigue and distraction, and existing datasets are often limited in scope and depth, lacking the nuanced information required to address these challenges effectively. This paper proposes a multimodal cross-sensing method combining visual and haptic channels under controlled environmental conditions and constructs the VTD (Visual and Tactile Database for Driver State and Behavior Perception), a comprehensive, well-structured, large-scale dataset. The VTD dataset not only covers diverse driving conditions and human behavior characteristics but also places particular emphasis on monitoring and recording driver fatigue and distraction states. These data will help enhance the perception and understanding capabilities of autonomous driving systems regarding driver states in intelligent co-driving scenarios, thereby improving the safety and reliability of human-machine interaction and effectively reducing potential risks.

The contributions of VTD dataset are as follows:

1) This paper presents the VTD, a long-sequence multi-modal natural driving dataset based on the fusion of visual and haptic data, which includes over 10 hours of fatigue driving data and 102 takeover scenarios. It effectively captures the multi-path driving conditions, as well as the mental and physical states and behavioral characteristics of drivers in

*This work was co-supported by the National Key Research and Development Pro-gram of China (No. 2021YFE0193900 and 2020AAA0108901), the National Natural Science Foundation of China (No. 52002286, 51975415 and 62088101), the Fundamental Research Funds for the Central Universities (22120220642). (Corresponding author: Zhongpan Zhu).

[1]Jie Wang, Zhongpan Zhu are with the College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China, and with Frontiers Science Center for Intelligent Autonomous Systems, Shanghai 201210,China(e-mail:2054310@tongji.edu.cn;521bergsteiger@tongji.edu.cn;)

[2]Mobing Cai is with the Department of Engineering and Trinity College, University of Oxford, Oxford, OX1 3BH, United Kingdom. (email: mobing.cai@trinity.ox.ac.uk)

Hongjun Ding, Jiwei Yi, Aimin Du are with the school of Automotive Studies, Tongji University, Shanghai 201804, China (e-mail: 2031656@tongji.edu.cn; 2033522@tongji.edu.cn; duaimin@tongji.edu.cn;)

[1]The uncertainty in driver non-driving behavior refers to situations where the driver is not fully engaged in the primary driving task due to factors like fatigue or distraction.

[2]Abnormal state refers to the unpredictability in driving behavior caused by changes in the driver's state, which elevates the risk in human-machine interaction.

multimodal environments.

2)The VTD is developed using a human-in-the-loop algorithm to ensure that the collected data accurately reflects real-world driving behavior. Additionally, the boundaries of driving environment are clearly defined, facilitating the quantification of the influencing mechanisms behind the driving behaviors of different groups.

3) Serving as a standardized platform for benchmarking in the field of human-vehicle co-driving, VTD offers valuable data support for cross-modal perception algorithms and scenarios related to driver fatigue and distraction. This capability significantly advances research in driver behavior perception.

II. RELATED WORK

In designing human-vehicle collaboration systems, recognizing the driver's intent, modeling behavior, and monitoring their state are critical. Despite advancements in autonomous driving technologies, the driver remains the core of system coordination[5]. Therefore, accurately monitoring the driver's fatigue and distraction is essential to ensure the safe operation of human-machine collaborative systems.

Driver state monitoring primarily focuses on physiological and psychological factors. Driver behavior monitoring involves analyzing specific behaviors during driving to infer their state and decision-making process. Behavior is an external manifestation of the state: the former reflects specific actions, while the latter represents the mental and physical condition. By observing driving behavior and measuring physiological and psychological indicators, a more comprehensive inference of the driver's fatigue or distraction can be made. This paper will introduce datasets in the field of human-vehicle collaboration from the perspectives of driver state and behavior, along with their applications and limitations in driver monitoring and human-vehicle co-driving system optimization.

A. Datasets for Driver State

As mentioned earlier, fatigue and distraction are the two main factors affecting driving safety. Therefore, this section will primarily focus on the current methods and datasets for monitoring driver distraction and fatigue, and analyze their applicability and limitations in real-world scenarios.

Research on driver distraction monitoring[3] has varied focuses. In recent years, the rise of AI algorithms for computer vision has led to a growing interest in analyzing driver behavior through visual perception methods. This includes studying facial expressions [6] and head gestures [7]. Key features commonly extracted in these studies are eye fixation duration, scan paths, eye-opening and eye-closing patterns, and head rotation angles. In addition, driving distraction monitoring and behavior analysis can also be achieved by using vehicle sensors to monitor vehicle conditions, such as steering wheel angle and accelerator pedal position, or by physiological indices such as driver Electrocardiogram (ECG) [8, 9], Electroencephalogram (EEG) [10–12], Electromyography (EMG), and Galvanic Skin Response (GSR) [11].

Visual-based driver distraction monitoring accuracy is still limited in an actual driving environment due to problems including low resolution, motion blur, dynamic background, and occlusions [13]. Hand movements, head gestures [7], gaze direction [14], and pedal control are the keys to addressing the problems. Moreover, existing datasets still cannot characterize diverse, ambiguous, and personalized distraction behaviors influenced by the driver's physiological and psychological state[15–20]. The investigations of the above datasets demonstrate a need for fine-grained distraction datasets with controllable and quantifiable conditions, multi-modal synchronized data, and data about driver distraction feature diversity.

Aside from driving distraction detection, driver fatigue is likewise an important research direction of driver behavior and state perception. Fatigue is displayed in various forms. Based on eye feature extractions, PERCLOS (percent eye closure), eye-white reflex, eye states, and yawning conditionsare included[21]. It can also be displayed through detection based on eye-mouth combinations [22], eyelid closure and eye closure percentage combinations [23], FatigueTree [24], and other combinations.

We gathered and analyzed existing DMS datasets[16, 24–29] and discovered that in many datasets, fatigue is monitored by recording drivers' facial features under natural driving conditions using RGB cameras. These natural-driving datasets face challenges in extracting driver fatigue features under nighttime dimmed lighting conditions or facial occlusions, and different types of physiological fatigue signals cannot be captured with a single vision. Differences in road environments may also lead to differences in driving loads, making it impossible to analyze the cause of fatigue and extract the differentiated impact on drivers. Additionally, most of the existing single-mode visual datasets about driver fatigue only concentrate on relatively monotonous visual features like blinking and yawning [25, 26], which lack time series and contextual features. Algorithms carried out on these single-mode datasets are too restrictive to be applied in reality and cannot contain all challenges[24]. It is increasingly essential to determine how to provide greater flexibility and diversity in driver monitoring using multi-modal data to characterize fatigue signals and more complicated state combinations. However, the lack of a complete and comprehensive dataset in this field has bottlenecked the progress in algorithm development of driver fatigue detection [16].

B. Datasets for Driver Behaviors

Driver behavior is distinct from the driver's state. It refers to the specific movements a driver makes during the driving process, such as turning, accelerating, decelerating, and braking. Driver behavior can be captured and analyzed using vehicle sensors, cameras, and other monitoring devices. It is often associated with specific driving skills and traffic regulations, such as speeding, frequent lane changes, and

[3] NHTSA describes the distraction process as "any activity that diverts a driver's attention away from the task of driving" and classifies it into visual, auditory, bio-mechanical, and cognitive distractions

running red lights. In essence, driver behavior encompasses the specific actions and maneuvers made by the driver, whereas driver state refers to their physical and mental condition. The driver's state can be indirectly inferred by observing and evaluating their behaviors and by measuring relevant physiological indicators. Modeling and understanding a driver's state through their behavior is essential for ensuring safety and facilitating assisted driving [30]. To enhance driving safety, future intelligent vehicles should be capable of autonomously assessing the driver's behavior and competence using onboard sensors and operational data.

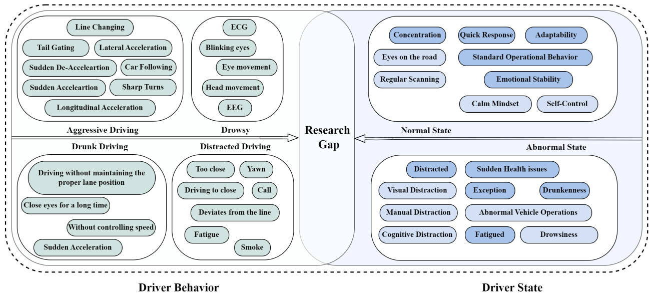


Fig. 1: Research Gap in Driver Behavior and Driver State

Natural driving data is a vital resource for learning and understanding driving behaviors [30]. Vehicle operations and driver behavior data are captured and collected through cameras and sensor arrays, providing essential support for research in this field. However, natural driving datasets often lack a unified boundary [31, 32], and there is limited coupling of data on driver behaviors and states, making it challenging to perform quantitative analysis on contributing factors. Additionally, most datasets are outdated and suffer from low data quality due to limited device accuracy, rendering them inadequate for current needs. Publicly available natural driving datasets are also scarce because of the high cost of data collection[4]. The VTD dataset aims to provide new options for advancing research in this field.

III. METHODOLOGY

This section proposes a multi-view and multi-modal database named VTD for studying driver distraction and fatigue behaviors. It consists of multi-modal signals, including frontal images, ECG signals, and vehicle signals. To emphasize the practicality and authenticity of VTD, a platform was developed that integrates driving simulation and multi-modal signal acquisition functions to conduct experiments. Finally, the dataset was constructed through data processing and analysis to extract features related to driver state and behavior perception.

A. Overall Framework

Driving data from 15 subjects in a fatigued condition and takeover experiment data from 17 distracted participants are included in the VTD dataset. All participants were fully informed about the research background and procedures, and they consented to participate by signing a written informed consent form. Detailed information about the subjects is provided in Table I. This paper will separately describe the specifics of the two data collection experiments and the data processing procedures.

Experiment	Gender		Age		Driving Experience	
	male	female	mean	SD	mean	SD
Fatigue	12	3	32.2	6.14	3.73	3.16
Distraction	14	3	33.7	6.45	4.61	3.65

TABLE I: Basic Information of Participants

Figure II summarizes the VTD experimental process and data collection infrastructure, as illustrated. (a) Physical diagram of the data acquisition platform, including the monitor, G29 steering wheel and pedal set, RGB camera, and ECG equipment. (b) Framework for fatigue driving experiments. (c) Flowchart for generating fatigue data labels. Data is labeled based on subjective and objective evaluations to assess driver behavior. (d) Examples of facial videos of drivers in distracted and fatigued states during driving simulation. (e) Experimental setup for fatigue and distraction driving models[5]. (f) Framework for distracted driving experiments. (g) Process for distracted driving takeover experiments.

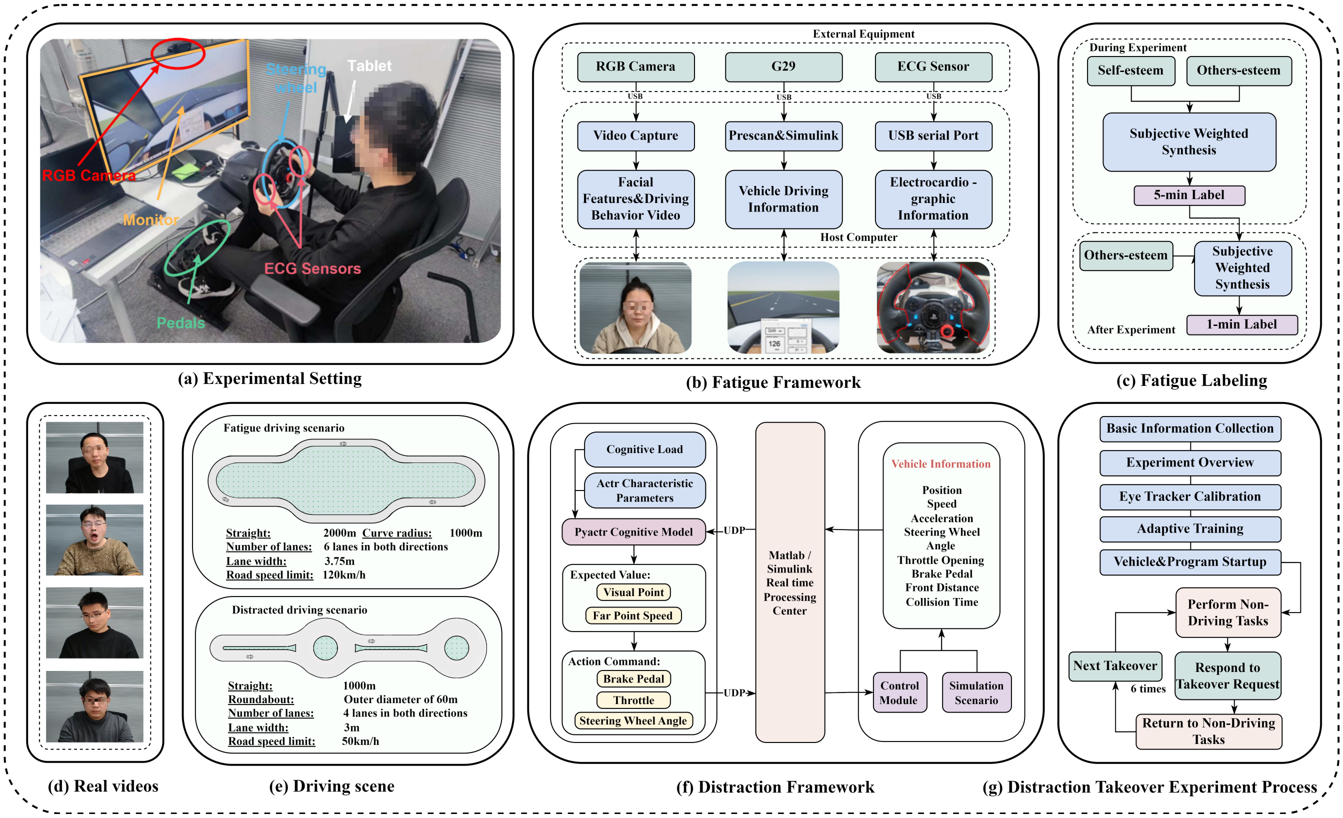
B. Fatigue Driving Experiment

The Fatigue dataset includes a processed time-series dataset and a raw simulation scenario video dataset. The time series dataset comprises eleven dimensions and is divided into a training set, a validation set, and a testing set in a ratio of 4:1:1.The total number of valid samples is 480, with a time slice of 60 seconds and a frame number of 1800. The dataset labels were modified based on KSS and SSS fatigue scales, and subjective evaluations were completed.

To construct a driver fatigue detection dataset, we collected information from participants and assigned pre-fatigue status according to age, sleep duration, and napping habits. Table II shows the pre-fatigue status of the participants. Participants were required to complete the pre-fatigue accumulation according to the assigned status. During the testing stage, the functionality of the platform connection and the data collection program was verified. Participants were instructed to complete the experiment preparations and driving adaptation under the operation instructions. They then performed the experiment wearing eye movement equipment and holding the electrode area of the steering wheel. They entered the appointed conditions and drove for 40 minutes continuously, during which they should keep their hands on the electrodes on the two sides. When the staff issued the prompt ”report the current status” every five minutes, the participants should complete their self-evaluations while the staff completed their assessments on the participants, combining the observation results and reported results.

[4] UTDrive, a large-scale natural dataset, includes driving data from 500 drivers across three countries, covering vehicles, sensors, and routes. In addition to UTDrive, the SHRP2 and MIT-AVT projects also focus on gathering natural driving data.

[5] To trigger distraction and fatigue, the scene is sparsely populated with vehicles, providing a wide driving view.



(a) Experimental Setting (b) Fatigue Framework (c) Fatigue Labeling

(d) Real videos (e) Driving scene (f) Distraction Framework (g) Distraction Takeover Experiment Process

Fig. 2: VTD Data Collection Infrastructure

State	Requirements
A1	Regular sleep: Ensure normal adequate sleep the night before the experiment. The experiment can be conducted in both the morning and afternoon. If in the afternoon, the nap time should be more than 0.5 hours.
A2	Nap deprivation: Ensure normal adequate sleep the night before the experiment. Conduct nap deprivation and the experiment only in the afternoon.
A3	Partial night sleep deprivation: Deprived of 40%-60% sleep the night before the experiment. No additional sleep on the day of the experiment. The experiment could be conducted in the morning and afternoon.
A4	Total night sleep deprivation: Deprived of over 80% sleep the night before the experiment. No additional sleep on the day of the experiment. The experiment could be conducted in the morning and afternoon.

TABLE II: The Pre-Fatigue States of the Participants

We generalized the different types of data collected as video data, vehicle data, and ECG data. The data were processed in different ways according to their characteristics.

For the driver frontal behavior information, the Mediapipe Facemesh model[33] was adopted for face landmark estimation,extracting 478 key coordinates of the driver's face[6] and characterized their mouths and eyesby Mouth Aspect Ratio (MAR) and Eye Aspect Ratio (EAR). The MAR and EAR values of each image frame are concatenated to form a time series. As indicated by the key points in the figure, the EAR and MAR are calculated as follows:

$$MAR = \frac{\| P_{82} - P_{87} \|_2 + \| P_{312} - P_{317} \|_2}{4 \| P_{78} - P_{308} \|_2} \tag{1}$$

$$EAR = \frac{\| P_{387} - P_{373} \|_2 + \| P_{385} - P_{380} \|_2}{4 \| P_{263} - P_{362} \|_2} + \frac{\| P_{158} - P_{153} \|_2 + \| P_{160} - P_{144} \|_2}{4 \| P_{133} - P_{33} \|_2} \tag{2}$$

For driver head pose capture, we used Euler angles to characterize the rotational pose of the head. The Euler angle is derived by utilizing the facial key point information obtained from Facemesh and solving the rotation matrix by the PnP (Perspective-n-Point) algorithm. Note that the camera needs to be calibrated because its parameters affect the mapping relationship of the spatial points in the plane. This study employs a planar tessellated grid, calibrated using the Zhang Zhengyou calibration method to obtain the camera's internal parameters and distortion coefficients.

During driving, the steering wheel installed with flexible electrodes can be connected to the host computer via USB, and the driver's ECG signals are acquired in real time by holding the steering wheel, with a sampling frequency of 250 Hz. This paper focuses on processing and feature extraction of ECG signals,primarily concentrating on the R-wave and utilizing the R-R intervals to compute the characteristics of

[6] Facial coordinate information can refer to: https://github.com/google-ai-edge/mediapipe/blob/master/mediapipe/modules/face_geometry/data/canonical_face_model_uv_visualization.png

heart rate and heart rate variability. For heart rate variability, the standard deviation of the R-R interval (SDNN) and the root mean square of the difference of the R0R interval (RMSSD) are used as metrics. To address baseline drift, IF interference, and EMG noise that occurr during the acquisition process, a Butterworth bandpass filter is applied.

For the vehicle data, we obtained them during real-time driving through the control signal of G29 with a sampling frequency of 100Hz, including the steering wheel angle, gas pedal signal, brake pedal signal, vehicle speed, and vehicle traverse angle speed.

C. Distracted Driving and Takeover Experiment

Regarding the driving takeover dataset, each participant is asked to perform three takeovers during visual and auditory subtasks in the investigation. Initially, the vehicle is in the autonomous driving phase while participants perform non-driving tasks on a tablet computer. When the system prompts a message to take over, drivers operate the vehicle while staff record relevant data and address unexpected situations.

This paper establishes 34 groups of visual and auditory subtasks for autonomous vehicles under three conditions (straight path, roundabout cut-in, and roundabout obstacle avoidance) and conducts 102 takeover experiments. After data screening and criteria extraction through nodes, takeover segments are extracted and divided according to the time nodes marking the start and end of the entire process, as well as those of the takeover.

Regarding takeover time, it is divided into takeover reaction time and takeover execution time. Takeover reaction time refers to the duration between the system's takeover request and the driver's return to the driving task (both hands back on the steering wheel), while takeover execution time is the sum of the duration during which the steering wheel angle$\geq 2^\circ$ and the pedal was pressed $\geq 10\%$.

VTD's takeover and distraction data can also be used to calculate a driver's load rate in human-vehicle co-pilot tasks through cognitive architecture models like QN-ACTR. Based on the load rate, driver's fatigue level can be assessed, and by combining the vehicle's displacement information, a safer and more reasonable human-vehicle driving right switching strategy can be designed.VTD also includes unscreened raw time series, raw videos, and tactile data so that users can filter and combine data based on research needs and goals.

D. VTD Experiment Setup Innovation

1) Multi-channel and Multi-angle Videos: Owing to the lack of public driver behavior datasets, most datasets are single-mode (RGB). For safety reasons, only simple visual signals can be collected in actual driving processes. These visual features usually depend on cameras and sensors directed toward the driver to obtain input data. The large-scale multi-view multi-modal database we constructed, VTD, can fill the gap for single visual signals. The accuracy of features extracted from facial detection, head pose estimation, and eye status analysis can be enhanced using multi-view information like driving view, eye movement, and facial view [34]. Moreover, eye trackers' high sampling rate, high precision, and low noise are advantageous compared to visual feature detection. Other modal features can be tuned to enhance the overall recognition rate using multi-view feature extraction and fusion.

Eye detection and eye status analysis are crucial to driver distraction and fatigue detection. Head rotation and eye closure rate can be calculated by applying PERCLOS to measure a driver's fatigue level and PERLOOK [35] to measure ametropia duration. Due to limitations in resolution, camera-eye distance, and lighting conditions, it is not easy to calculate and distinguish the accuracy of data results from current mainstream datasets. However, VTD adopted Tobii Glasses3 to obtain omnidirectional eye movement tracking data from various angles, thus achieving the capture of high-precision eye movement data in an extensive range.

In current mainstream research methods, behavior analysis and fatigue detection have also been conducted by fully using the drivers' diverse characteristics. These include detecting physiological signals, such as using EOG(Electrooculography) and ECG[35, 36] or combining driving measurements (Steering wheel angle, steering speed, accelerator pedal angle, etc.) [37, 38]. In all of the above scenarios, VTD is adaptable.

2) Tactile sensing device for driver's ECG: To minimize the impact of the ECG devices on the driver, these devices are fixed on the G29 steering wheel. Signals are acquired through two flexible electrodes and transmitted to the collection program via USB, where they are saved as real-time texts. In contrast to signal acquisition from the participants' left and right earlobes, participants only need to hold the electrode area of the steering wheel to perform real-time heart rate detection. This approach significantly reduces the chance of distraction and mitigates the devices' impact on the experiment.

E. Data Processing Method Innovations

To better align the dataset with the training model, normalization preprocessing is performed on the time-series data, which is then categorized according to research directions. Regarding driving fatigue detection, the VTD dataset contains 11-dimensional time series information and includes data series of the drivers' frontal image, ECG signals, and the vehicle's motion state. Fatigue levels are then graded by subjective evaluations combined with self-assessment and other's assessment, thus realizing data calibration of Human-in-the-loop. Subsequently, dimension reduction and screening are performed on the above data to ensure a strong correlation between the data and the driver behaviors.

Table III[7] presents the 10 dimensional sequence information and analysis results of VTD fatigue data. The time series in some dimensions are chosen and investigated using One-way ANOVA (Analysis of Variance) to determine whether time series features are salient under different fatigue levels.

[7] "++++" represents very significant differences ($\alpha < 0.01$); "+++" represents significant differences ($0.01 \leq \alpha < 0.05$).

Time series	Signal	Clue	F	P	S
EAR	Driver Frontal Image Signal	PERCLOS	10.2095	4.3141×10^{-6}	++++
		Blinking Rate	4.2819	5.5569×10^{-3}	++++
MAR		MAR(SD)	4.0222	8.1897×10^{-3}	++++
Head Tilt		Head Tilt(SD)	14.1214	2.0341×10^{-8}	++++
Head Yaw		-	-	-	-
Head Roll		-	-	-	-
R-R	ECG	SDNN	12.3479	1.7280×10^{-7}	++++
		RMSSD	14.1791	7.8400×10^{-9}	++++
Steering Wheel Angle	Vehicle Signals	Steering Angle(SD)	6.4363	5.5569×10^{-2}	+++
Pedal		Pedal(SD)	3.5743	1.4349×10^{-2}	++++
Vehicle Speed		Speed(SD)	7.0730	1.2934×10^{-4}	++++
Transverse Angular Velocity		Transverse Angular Velocity(SD)	5.7549	1.6380×10^{-4}	++++

TABLE III: Analysis of 10-Dimensional Time Series Information and F, P, and Significance Levels in Fatigue Data

Dataset	Features	People	Quantity	Environment
PUBLIC DISTRACTION DATASETS				
3MDAD[15]	Comprehensive	50	507 min videos, 20-34 sec each	Act + Real
DMD[16]	Comprehensive	37	41h RGBD+IR videos	Real + Lab
VIVA[17]	Hands	8	2000+ images	Act
DriveAHead[18]	Heads	20	100 million Depth & IR images	Real
DAD[19]	Behavior	31	783 min videos	Act
MDAD[20]	Driver action	50	2x2x800 video sequences	Real + Lab
Ours VTD	**Comprehensive**	**17**	**6 types of scenarios, 102 takeover experiments,630 min videos**	**Real + Lab**
DMS PUBLIC FATIGUE DATASETS				
YawDD[25]	Yawning	107	342 videos, 15-40 sec each	Act + Real
ZJU[26]	Eye Blinking	20	80 videos	Act + Lab
NTHU[27]	Drowsiness	36	360 videos, 1 min each	Act + Simulated
RLDD[28]	Drowsiness	12	180 videos, 10 min each	Real + Lab
NTHU-DDD[27]	Comprehensive	36	RGB + TXT	Act
DMD[16]	Comprehensive	37	41h RGBD+IR videos	Real + Lab
CMU-PIE[29]	Head Pose	72	1503 images	Real + Lab
Ours VTD	**Comprehensive**	**15**	**600 min videos,10-Dimensional Time-Series Signals**	**Real + Lab**

TABLE IV: Comprehensive Summary and Comparison of Public Fatigue and Distraction Datasets with VTD

From the ten features analyzed, VTD's fatigue data and driver fatigue are strongly correlated[8]. All are valid inputs for the fatigue classification model.

This paper examines the differences in various takeover scenarios and subtasks under conditions of distraction and takeover. The findings indicate significant differences in collision avoidance conditions between straight roads and roundabouts ($P = 5.536 \times 10^{-10} < 0.05$,$P = 2.879 \times 10^{-2} < 0.05$); there are also significant differences under visual and auditory subtasks ($P = 9.120 \times 10^{-4} < 0.05$,$P = 6.060 \times 10^{-3} < 0.05$). These results suggest that different subtasks and takeover scenarios impact takeover reaction time.

IV. PROPERTIES

A. Video Data Characteristics and Applications

VTD contains various complex combinations of visual indications and different fatigue levels and distractions in takeover tasks. The experiment includes 10-hour fatigue driving data from 15 participants and 102 takeover multi-dimensional experiment data from 17 participants (including recorded video data).Table IV lists a comparison of attributes between VTD and existing datasets. Compared to

[8] The F-value in Table III needs to be averaged over the serial information of each dimension and the total data, after which the specific value is obtained by calculating the ratio of the between-groups variance (MSA) to the within-groups variance (MSE). We assume that it satisfies the distribution $F(k-1, n-k)$, and obtain the probability P based on the F distribution, setting the significance level $\alpha = 0.05$ as the benchmark. When $p < \alpha$, it is considered that there is a significant difference between different serial data.

other publicly available data, VTD offers multimodal, multi-view, diverse, and fine-grained data that is controllable and quantifiable. This provides rich data support for research on human-machine collaborative driving systems and driver safety monitoring systems.

Another key highlight of VTD is the construction of a long-sequence multimodal natural data based on visual-tactile data fusion. We designed extended time-series segments integrating multiple modalities, including RGB facial video, vehicle motion data, tactile ECG data and images captured by the eye trackers in driving scenarios. This comprehensive approach enhances data diversity and granularity, providing critical insights into driver fatigue and distraction.These video data can be utilized in studies including driver fatigue detection, distraction monitoring, and human-vehicle driving control transitions.

B. Properties and Functions of Tactile Data

VTD provides piezoelectric tactile ECG data and steering wheel data. In the 40-minute experiment, we gathered drivers' ECG and PPG data using their tactile feedback to the flexible electrode during real-time driving. At the same time, 7-dimensional data of the steering wheel, the vehicle direction, the brake, the accelerator, the gear position, and the turning angle were collected. The driver's feedback and steering wheel data formed cross-validation.

Traditional visual driving behavior detection methods are limited by lighting conditions and the vehicle's location. Consequently, they are unable to satisfy continuous, high-quality visual signal collection. Additionally, visual identity systems are also faced with problems including but not limited to computing power issues and communication delays. We can partly solve the above issues with the wearable piezoelectric tactile device without creating constraints or disturbance caused by traditional wearable devices. However, difficulties still exist, for example, too many environmental interference sources and insufficient robustness for dynamic changes in signals and environments[39, 40]. A breakthrough that future driver behavior perception technology should anticipate is a combination of vision and tactile sensations that can balance the driver's state, the accuracy of behavior recognition, and application adaptability.

C. Properties and Functions of Visual-tactile Combinations

Combining vision and tactile sensation can compensate for the robustness of visual perception by incorporating the visual modality's sensitivity to position and movement and the tactile modality's rapidity. It can reduce the system delay under the risk of data overload and form mutual complementary effects under vehicle tracking and collision avoidance control [41]. Visual-tactile fusion requires temporal embedding when combined with multi-dimensional time-series data. Positional embeddings are also added to non-linear transformed time series to leverage the sequential correlations based on time steps. While using Transformer to classify time series and make predictions, positional embeddings can be employed to solve the scene adaptation issue of position data and time series in Transformer. These positional embedding vectors, along with multi-dimensional time series, can be injected into the model as additional input.

V. CONCLUSIONS

This paper presents a method for constructing a long-sequence multimodal natural dataset based on visual-tactile data fusion. The aim is to provide data support for quantifying and validating drivers' fatigue and distraction detection across identical driving scenarios, as well as for cross-modal perception algorithms related to driver behaviors, such as driving takeover monitoring. To meet various research demands, the VTD dataset includes data on fatigue driving and the drivers' visual and tactile behaviors during human-vehicle driving control transitions. This work aims to establish a standardized platform for benchmark testing, thereby advancing the development of driver behavior perception and enhancing research on driving safety.

ACKNOWLEDGMENT

The authors would like to thank all the people who participated in the data acquisitions and give their great support in this project.

REFERENCES

[1] Y. Li, Y. Su, X. Zhang, Q. Cai, H. Lu, and Y. Liu, "A simulation system for human-in-the-loop driving," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 2022, pp. 4183–4188.

[2] J. Wu, Z. Huang, Z. Hu, and C. Lv, "Toward human-in-the-loop ai: Enhancing deep reinforcement learning via real-time human guidance for autonomous driving," *Engineering*, vol. 21, pp. 75–91, 2023.

[3] T. Arakawa, "Trial verification of human reliance on autonomous vehicles from the viewpoint of human factors," *Int. J. Innov. Comput. Inf. Control*, vol. 14, no. January 2017, pp. 491–501, 2018.

[4] H. Singh and A. Kathuria, "Analyzing driver behavior under naturalistic driving conditions: A review," *Accident Analysis & Prevention*, vol. 150, p. 105 908, 2021.

[5] S. Gnatzig, F. Schuller, and M. Lienkamp, "Human-machine interaction as key technology for driverless driving-a trajectory-based shared autonomy control approach," in *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, IEEE, 2012, pp. 913–918.

[6] M.-H. Sigari, M.-R. Pourshahabi, M. Soryani, and M. Fathy, "A review on driver face monitoring systems for fatigue and distraction detection," *International Journal of Advanced Science and Technology*, vol. 64, pp. 73–100, 2014.

[7] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation in computer vision: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 4, pp. 607–626, 2008.

[8] S. V. Deshmukh and O. Dehzangi, "Ecg-based driver distraction identification using wavelet packet transform and discriminative kernel-based features," in *2017 IEEE International Conference on Smart Computing (SMARTCOMP)*, IEEE, 2017, pp. 1–7.

[9] M. Taherisadr, P. Asnani, S. Galster, and O. Dehzangi, "Ecg-based driver inattention identification during naturalistic driving using mel-frequency cepstrum 2-d transform and convolutional neural networks," *Smart health*, vol. 9, pp. 50–61, 2018.

[10] G. Li, W. Yan, S. Li, X. Qu, W. Chu, and D. Cao, "A temporal–spatial deep learning approach for driver distraction detection based on eeg signals," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 4, pp. 2665–2677, 2021.

[11] P. Manikandan, M. R. S. Reddy, S. Mehatab, and P. M. Sai, "Automobile drivers distraction avoiding system using galvanic skin responses," in *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, IEEE, 2021, pp. 1818–1821.

[12] X. Zuo, C. Zhang, F. Cong, J. Zhao, and T. Hämäläinen, "Driver distraction detection using bidirectional long shortterm network based on multiscale entropy of eeg," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 19 309–19 322, 2022.

[13] I. Jegham, A. B. Khalifa, I. Alouani, and M. A. Mahjoub, "Vision-based human action recognition: An overview and real world challenges," *Forensic Science International: Digital Investigation*, vol. 32, p. 200 901, 2020.

[14] I. Jegham, A. B. Khalifa, I. Alouani, and M. A. Mahjoub, "Safe driving: Driver action recognition using surf keypoints," in *2018 30th International Conference on Microelectronics (ICM)*, IEEE, 2018, pp. 60–63.

[15] I. Jegham, A. B. Khalifa, I. Alouani, and M. A. Mahjoub, "A novel public dataset for multimodal multiview and multispectral driver distraction analysis: 3mdad," *Signal Processing: Image Communication*, vol. 88, p. 115 960, 2020.

[16] J. D. Ortega *et al.*, "Dmd: A large-scale multi-modal driver monitoring dataset for attention and alertness analysis," in *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16*, Springer, 2020, pp. 387–405.

[17] N. Das, E. Ohn-Bar, and M. M. Trivedi, "On performance evaluation of driver hand detection algorithms: Challenges, dataset, and metrics," in *2015 IEEE 18th international conference on intelligent transportation systems*, IEEE, 2015, pp. 2953–2958.

[18] A. Schwarz, M. Haurilet, M. Martinez, and R. Stiefelhagen, "Driveahead-a large-scale driver head pose dataset," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 1–10.

[19] O. Kopuklu, J. Zheng, H. Xu, and G. Rigoll, "Driver anomaly detection: A dataset and contrastive learning approach," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 91–100.

[20] I. Jegham, A. Ben Khalifa, I. Alouani, and M. A. Mahjoub, "Mdad: A multimodal and multiview in-vehicle driver action dataset," in *Computer Analysis of Images and Patterns: 18th International Conference, CAIP 2019, Salerno, Italy, September 3–5, 2019, Proceedings, Part I 18*, Springer, 2019, pp. 518–529.

[21] M. Omidyeganeh *et al.*, "Yawning detection using embedded smart cameras," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 3, pp. 570–582, 2016.

[22] Y. Ying, S. Jing, and Z. Wei, "The monitoring method of driver's fatigue based on neural network," in *2007 International Conference on Mechatronics and Automation*, IEEE, 2007, pp. 3555–3559.

[23] L. M. Bergasa, J. Nuevo, M. A. Sotelo, R. Barea, and M. E. Lopez, "Real-time system for monitoring driver vigilance," *IEEE Transactions on intelligent transportation systems*, vol. 7, no. 1, pp. 63–77, 2006.

[24] C. Yang, Z. Yang, W. Li, and J. See, "Fatigueview: A multicamera video dataset for vision-based drowsiness detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 1, pp. 233–246, 2022.

[25] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, and B. Hariri, "Yawdd: A yawning detection dataset," in *Proceedings of the 5th ACM multimedia systems conference*, 2014, pp. 24–28.

[26] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based antispoofing in face recognition from a generic webcamera," in *2007 IEEE 11th international conference on computer vision*, IEEE, 2007, pp. 1–8.

[27] C.-H. Weng, Y.-H. Lai, and S.-H. Lai, "Driver drowsiness detection via a hierarchical temporal deep belief network," in *Computer Vision–ACCV 2016 Workshops: ACCV 2016 International Workshops, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part III 13*, Springer, 2017, pp. 117–133.

[28] R. Ghoddoosian, M. Galib, and V. Athitsos, "A realistic dataset and baseline temporal model for early drowsiness detection," in *Proceedings of the ieee/cvf conference on computer vision and pattern recognition workshops*, 2019, pp. 0–0.

[29] K. Diaz-Chito, A. Hernández-Sabaté, and A. M. López, "A reduced feature set for driver head pose estimation," *Applied Soft Computing*, vol. 45, pp. 98–107, 2016.

[30] Y. Liu and J. H. Hansen, "A review of utdrive studies: Learning driver behavior from naturalistic driving data," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 2, pp. 338–346, 2021.

[31] M. H. Alkinani, W. Z. Khan, and Q. Arshad, "Detecting human driver inattentive and aggressive driving behavior using deep learning: Recent advances, requirements and open challenges," *Ieee Access*, vol. 8, pp. 105 008–105 030, 2020.

[32] X. Hu, R. Eberhart, and B. Foresman, "Modeling drowsy driving behaviors," in *Proceedings of 2010 IEEE International Conference on Vehicular Electronics and Safety*, IEEE, 2010, pp. 13–17.

[33] Y. Kartynnik, A. Ablavatski, I. Grishchenko, and M. Grundmann, "Real-time facial surface geometry from monocular video on mobile gpus," *arXiv preprint arXiv:1907.06724*, 2019.

[34] K. Yuen, S. Martin, and M. M. Trivedi, "Looking at faces in a vehicle: A deep cnn based approach and evaluation," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 2016, pp. 649–654.

[35] J. Jo, S. J. Lee, H. G. Jung, K. R. Park, and J. Kim, "Vision-based method for detecting driver drowsiness and distraction in driver monitoring system," *Optical Engineering*, vol. 50, no. 12, pp. 127 202–127 202, 2011.

[36] L. Wang, J. Li, and Y. Wang, "Modeling and recognition of driving fatigue state based on rr intervals of ecg data," *Ieee Access*, vol. 7, pp. 175 584–175 593, 2019.

[37] X. Li, L. Hong, J.-c. Wang, and X. Liu, "Fatigue driving detection model based on multi-feature fusion and semi-supervised active learning," *IET Intelligent Transport Systems*, vol. 13, no. 9, pp. 1401–1409, 2019.

[38] S. Lim and J. H. Yang, "Driver state estimation by convolutional neural network using multimodal sensor data," *Electronics Letters*, vol. 52, no. 17, pp. 1495–1497, 2016.

[39] L. Boon-Leng, L. Dae-Seok, and L. Boon-Giin, "Mobile-based wearable-type of driver fatigue detection by gsr and emg," in *TENCON 2015-2015 IEEE Region 10 Conference*, IEEE, 2015, pp. 1–4.

[40] B.-G. Lee and W.-Y. Chung, "Wearable glove-type driver stress detection using a motion sensor," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 7, pp. 1835–1844, 2016.

[41] J. Yi, A. Du, Z. Zhu, and H. Ding, "A survey of driver behavior perception methods for human-computer hybrid enhancement of intelligent driving," in *Proceedings of China SAE Congress 2021: Selected Papers*, Springer, 2022, pp. 754–766.