WIP: Body Posture Analysis as an Objective Measurement for Human Trust Dynamics in AVs

Cherin Lim
University of Washington
cherinl@uw.edu

Tianhao Xu
University of Washington
tx29@uw.edu

Prashanth Rajivan
University of Washington
prajivan@uw.edu

***Abstract*—Human trust is critical for the adoption and continued use of autonomous vehicles (AVs). Experiencing vehicle failures that stem from security threats to underlying technologies that enable autonomous driving, can significantly degrade drivers' trust in AVs. It is crucial to understand and measure how security threats to AVs impact human trust. To this end, we conducted a driving simulator study with forty participants who underwent three drives including one that had simulated cybersecurity attacks. We hypothesize drivers' trust in the vehicle is reflected through drivers' body posture, foot movement, and engagement with vehicle controls during the drive. To test this hypothesis, we extracted body posture features from each frame in the video recordings, computed skeletal angles, and performed k-means clustering on these values to classify drivers' foot positions. In this paper, we present an algorithmic pipeline for automatic analysis of body posture and objective measurement of trust that could be used for building AVs capable of trust calibration after security attack events.**

I. Introduction

Advancements in transportation systems have transformed contemporary vehicles into smart modes of transportation capable of driving autonomously using sensors, artificial intelligence technologies, and communication network protocols [1]. While such development aims to ensure safer and more efficient operation of vehicles, advancements in the autonomy of vehicles also expose them to cybersecurity threats. Despite the relatively rare occurrence of security breaches, the impact of such incidents on the trust between drivers and autonomous vehicles (AVs) can be profound. Trust is a crucial element in the successful adoption and acceptance of autonomous technologies, influencing the overall user experience and perception of safety.

Like any other Internet-enabled technology, AVs are also vulnerable to a variety of threats that can vary based on type of attacker (source), attack vector (method), target, motive (objective/reason) and potential consequences (outcome) (Thing and Wu [17] for a taxonomy of potential security threats that an AV can experience). For example, the attack vectors categorized as 'physical access' and 'remote access' could include attacks such as code injection, in-vehicle spoofing, and packet fuzzing, and the latter could include GPS spoofing and jamming.

Despite the risks posed by such security threats to human interactions, there has been limited research examining the human aspects of autonomous vehicle (AV) security. Specifically, there is a gap in understanding about the methods for assessing and calibrating drivers' degraded trust in the system following a cyberattack. As highlighted by [11], trust plays a pivotal role in the acceptance of AVs, and once compromised, rebuilding it becomes a formidable challenge. Experiencing failure can significantly erode drivers' trust levels. However, the specific impacts and the extent to which experiencing failures from security threats erode drivers' trust remains to be measured and understood. Developing methods capable of robustly assessing and calibrating drivers' trust in autonomous vehicles is also imperative.

A. Trust Measurements

The development of higher levels of automation in AVs has created a dynamic between the human driver and the system in which the driving task can be shared. One safety implication of this dynamic relationship is the driver's behavior after relinquishing control to the autonomous driving system. For example, a driver may choose to sleep or use their phone instead of supervising the system, resulting in potentially dangerous road situations. This decision to appropriately supervise the AV or not can be heavily influenced by the driver's trust in the system [2].

Past studies have widely relied on administering trust questionnaires that result in self-reported responses from individuals to measure human trust levels in such autonomous driving scenarios. For instance, Jian et al. [10] developed a 12-item trust scale for automated systems based on different types of trust, i.e., human–human trust, human–machine trust, and trust in general. In another study conducted by Holthausen et al. [9], a situational trust scale for autonomous driving was constructed to rate various components such as trust, performance, non-driving related tasks (NDRT), risk, judgement, and reaction. Other studies have used self-defined scales to evaluate individuals' trust levels and overall driving experience of AVs [5], [7], [12].

However, due to the inherent limitations of self-reporting data based on the susceptibility to individual and situational biases, subjective trust assessments provided by drivers may

Symposium on Vehicles Security and Privacy (VehicleSec) 2024
26 February 2024, San Diego, CA, USA
ISBN 979-8-9894372-7-6
https://dx.doi.org/10.14722/vehiclesec.2024.23061
www.ndss-symposium.org

Fig. 1: Experiment Setup

not align with the level of trust that could be objectively measured through observed behaviors. This discrepancy is attributed to the non-objective and potentially inaccurate nature of self-reporting, as highlighted by various studies [8], [14], and the presence of unaccounted confounding effects [21]. Furthermore, methods that involve self-reported measures of trust are not amenable to the development of trust-aware autonomous vehicles.

To overcome such drawbacks, few studies have proposed the use of objective measures of driver trust centered around drivers' behavioral responses including hand position frequencies and transition probabilities [21], drivers' gas pedal control during car following [6], and eye glance patterns [18]. Also, on-road studies have been conducted to collect more realistic driving behavior data, and observe the interaction between drivers and AVs under dynamic real-world environments [15].

II. Methodology and Data

A driving simulator study was conducted at a large public university in the USA with 40 participants (22 males, 18 females) to investigate their responses to autonomous driving situations and their trust levels toward self-driving cars during cybersecurity attacks. The recruitment criteria were (1) 18 years old and above, (2) at least one year of US driving experience, and (3) English proficiency. The participants had an average age of 24.5 years and an average driving experience of 5.95 years.

The driving scenarios were created using the National Advanced Driving Simulator miniSim and ISAT. The simulator featured a realistic setup with a steering wheel, pedals, displays, and other in-vehicle controls. To manipulate an "autonomous" driving environment, the scenarios were driven by one of the researchers, prerecorded, and shown to the participants during the experiment. To heighten participant attentiveness and engagement during each of the drives, participants were told to feel free to step on the brakes, the accelerator, or use other vehicle controls if they felt the need to do so, and were not explicitly informed that the experiment stimuli were prerecorded videos.

At the beginning of the study, participants were informed about the research's focus on understanding drivers' reactions to autonomous driving and their trust in the face of cybersecurity threats, and signed an informed consent form. However, in order to prevent participants from creating bias in how they interact with the simulator, no specific information regarding when the threats would occur and what types of threats they would encounter was not provided. Then participants underwent three drives: a Baseline Drive with no cybersecurity threats, an Attack Drive with simulated cybersecurity attacks at specific time intervals, and a Post-Attack Drive resembling the Baseline Drive. Thus, the Baseline Drive was considered as the control drive with the absence of security perturbations, to which drivers' behaviors in the other drives would be compared. All drives involved city driving with various events such as traffic lights, pedestrian crossings, and stop signs. The Attack Drive featured three cybersecurity attacks, including failure to recognize a stop sign, running a red light, and not identifying a pedestrian crossing. In other words, the cybersecurity attacks simulated in the experiment were modeled on the premise that the perception module of the AV was compromised. After the Attack Drive, participants received a notification about the cybersecurity attack, simulating compromised vehicle controls.

The experiment lasted approximately 60 minutes, and participants received a $25 gift card as compensation. All study protocols were approved by the university's Institutional Review Board. The data collected from this experiment is used for developing a novel approach to measuring trust from drivers' foot positions.

III. Results

Drawing from prior research that employed alterations in body positions as a measure for evaluating driver confidence levels [20], [21], the study involved the extraction of foot position data from video recordings of the experiment. First, 0.5 second frames were extracted from the experiment recordings of each of the participants. With each drive being around 10 minutes resulting in a 30-minute recording for a single participant, a total of 144,522 frames were extracted as JPG files from all recordings.

Then, using the OpenPose library, which is a real-time system to jointly detect human body, hand, facial, and foot keypoints on single images [3], [4], [16], [19], key joints in the lower body were captured from the image files and saved as JSON files. Next, using python code, we defined utility functions to extract and manipulate pose information, as well as to draw skeleton lines and overlay the detected skeleton on RGB images. Thus, the first section of the code showcases the process of extracting, analyzing, and visually presenting body pose data in the context of video analysis.

Based on the identified skeletal lines, we calculated the angle between two lines defined by four input points (the starting and ending points of the two lines), employing vector operations and trigonometry. This is done by computing the Euclidean distance between two points in a 2D space,

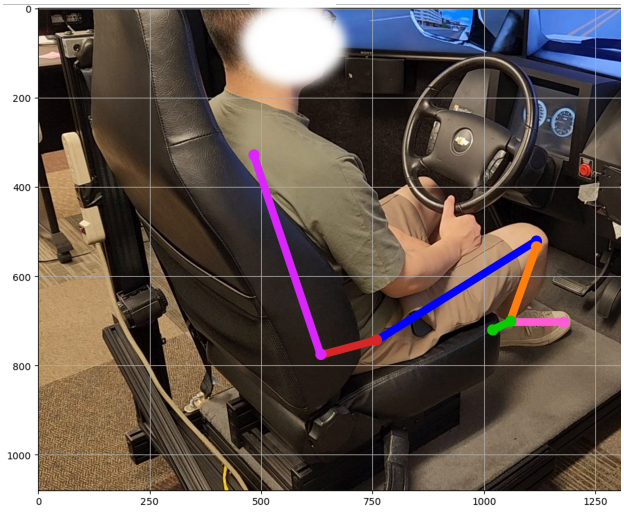

Fig. 2: Example of Feature Extraction Results

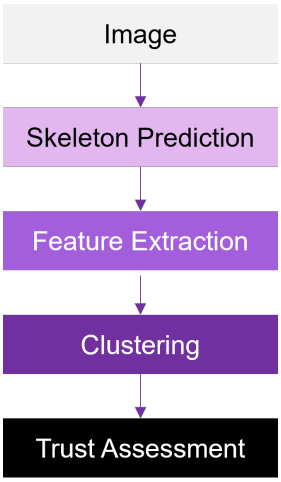

Fig. 3: Trust Assessment Pipeline

extracting the pixel coordinates of specified points from the OpenPose body pose estimation data, and calculating the angles between pairs of lines specified by point indices.

From the resulting angle values, we performed k-means clustering to classify foot positions using the scikit-learn library. The results from an initial analysis showed that hand positions and overall body posture were maintained in a relatively similar position throughout the drives, while foot positions changed according to each driving scenario. Thus, we specifically focused on foot posture for trust measurement for the scope of this research. We hypothesized that there would be mainly four different foot positions being: foot on brakes, foot on accelerator, hovering between pedals, and foot completely off both pedals. Thus, we used k=4 and were able to observe 5912, 10618, 23426, and 17503 occurrences respectively for each of the four clusters. Currently, this is a work in progress. Our focus is twofold. Firstly. we are evaluating the accuracy and reliability of such methods for accurately detecting foot positions in driving conditions. Secondly, we are developing a novel approach to automatically generate trust scores which is predicated on analyzing drivers' foot positions as an indicative measure of their trust levels.

IV. Discussion

A. Trust Assessment Pipeline

In this study, we propose a pipeline for trust assessment based on the analysis of drivers' body postures, specificially foot positions. The initial step involves the segmentation of the recording into discrete 0.5-second frames. Subsequently, joint coordinates such as knees and ankles are extracted and the skeletal structure in each frame is detected. Based on the resulting feature extraction, further computations can be performed to obtain distance and angle between skeletal features, in order to classify the foot position in the given frame. Next, unsupervised clustering techniques, (K-means clustering in this study) can be applied to group frames exhibiting similar postures. These clusters can subsequently be labeled to denote distinct foot positions, such as 'Foot on Brakes,' 'Foot on Accelerator,' 'Hovering Between Pedals,' and 'Completely Off Both Pedals.' This classification sets the basis for subsequent application of supervised learning algorithms, designed to search for a correlation between these categorized postures and the corresponding levels of trust, as reported by participants. This correlation process is underpinned by the dataset being labelled with trust levels derived from participants' self-reported trust levels. Moving forward, our model selection and training phase would incorporate classification models, using algorithms such as Support Vector Machines (SVM) and Random Forests.

By monitoring the driver's body postures, particularly the foot positions, the system can infer the driver's dependence on the autonomous system. For example, the driver's foot positioned more frequently on the brake pedal may indicate a higher level of caution and lower levels of trust towards AV. On the other hand, if the foot is consistently off both pedals, it may suggest a higher trust level in the autonomous system. The clustering and labeling of different postures can thus be used to identify patterns of trust, and the size of such clusters or frequency in the appearance of certain labels could be used as indicators of driver trust.

B. Objective Trust Measurement

While post-experiment questionnaires have been widely used for assessing trust in AVs, the incorporation of machine-learning-based algorithms that analyze body postures introduces a more objective dimension to trust measurement. Body posture can be a significant indicator of human attitudes toward the system in use, providing a non-intrusive and real-time source of information. Unlike questionnaires, which are retrospective and may be influenced by recall bias, machine learning algorithms analyzing body postures hold the capacity to provide a real-time assessment of driver trust levels towards AVs. This enables continuous monitoring and dynamic adap-

tation of the vehicle's responses to enhance trust during the driving experience.

Also, machine learning models can generate objective metrics related to posture and movement, such as skeletal angles as computed in this study. This allows for quantifiable measurements of trust, and such metrics can reduce the subjectivity associated with self-reported questionnaire results.

Moreover, information derived from machine learning algorithms analyzing body postures can be used to develop driver-adaptive systems. For instance, higher engagement with vehicle controls including frequent usage of brakes/accelerator may be consistently associated with decreased levels of trust. In such a situation, the vehicle's interface or behavior could be adjusted in real-time to address concerns and enhance drivers' overall perception of safety.

C. Lessening Manual Labor

The proposed method for posture analysis and trust quantification can significantly help future experiments on drivers' trust towards AV. Analyzing video recordings of drivers often necessitates extensive manual effort [13]. Extracting specific features at certain time frames within a video recording, a common task in posture analysis, can be particularly time-consuming for human observers. This process involves meticulously identifying and tracking key body posture indicators, which can be challenging and prone to subjectivity.

By employing machine learning algorithms for posture analysis, this labor-intensive aspect of the analysis can be significantly reduced. Machine learning models, once trained on a sufficiently diverse dataset, can autonomously recognize and extract relevant features from video frames. Moreover, as this particular method utilizes unsupervised learning algorithms that operate based on intrinsic patterns and relationships within the data itself, data labeling is not required. This automation not only accelerates the analysis process but also mitigates the potential for human error and variability in qualitative assessments of data collected from driving studies. As a result, the adoption of machine learning in posture analysis contributes to increased efficiency, consistency, and objectivity in extracting valuable insights from driver behavior, ultimately facilitating a more streamlined and effective approach to understanding and interpreting posture dynamics.

D. Limitations and Future Work

One notable limitation of the method proposed for posture analysis is the occasional identification of a single individual as two separate individuals, resulting in inaccuracies in foot position categorization. Particularly, it was shown that there was a difference in estimations when the algorithm limited the number of people it needed to identify as one or two. Also, there were errors where the full skeleton could not be identified and a loss of the lower body skeleton occurred. Additionally, the quality of the data could be constrained by inconsistencies in video recording conditions, including variations in recording angles and zoom levels. These inconsistencies can introduce noise and impact the reliability of the pose classification. Moreover, for a more comprehensive analysis in future studies, there is a potential avenue for improvement by incorporating additional features. For instance, including measurements such as distances to the brake and gas pedals or absolute positions of joints using x-y coordinates could provide a more accurate result for estimating body postures.

V. Conclusion

When faced with vehicle failures, drivers' confidence in AVs can rapidly diminish, necessitating the imperative task of rebuilding this trust for the widespread adoption and acceptance of autonomous technologies. To address this, we conducted a driving simulator study involving forty participants, subjecting them to three driving scenarios, one of which simulated cybersecurity attacks. Recognizing that drivers' trust is manifest in their body posture, movements, and interaction with vehicle controls, we extracted body posture features from each video frame, computed skeletal angles, and applied k-means clustering to classify drivers' foot positions. Consequently, we present an algorithmic pipeline for the automated analysis of body posture, providing an objective means to measure trust, which can be further utilized for trust recalibration and building driver adaptive systems.

Acknowledgment

This work was supported, in part, by a grant from the National Science Foundation (NSF grant #2142888). The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

References

[1] U. Ahmad, H. Song, A. Bilal, M. Alazab, and A. Jolfaei, "Securing smart vehicles from relay attacks using machine learning," *The Journal of Supercomputing*, vol. 76, pp. 2665–2682, 2020.

[2] M. Blanco, J. Atwood, H. Vasquez, T. Trimble, V. Fitchett, J. Radlbeck, G. Fitch, S. Russell, C. Green, B. Cullinane, and J. Morgan, "Human factors evaluation of level 2 and level 3 automated driving concepts," NHTSA, Tech. Rep., 2015.

[3] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh, "Openpose: Realtime multi-person 2d pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[4] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *CVPR*, 2017.

[5] M. Dikmen and C. Burns, "Trust in autonomous vehicles: The case of tesla autopilot and summon," in *2017 IEEE International conference on systems, man, and cybernetics (SMC)*. IEEE, 2017, pp. 1093–1098.

[6] F. Feng, S. Bao, J. Sayer, and D. LeBlanc, "Spectral power analysis of drivers' gas pedal control during steady-state car-following on freeways," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 60, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2016, pp. 729–733.

[7] N. Gang, S. Sibi, R. Michon, B. Mok, C. Chafe, and W. Ju, "Don't be alarmed: Sonifying autonomous vehicle perception to increase situation awareness," in *Proceedings of the 10th international conference on automotive user interfaces and interactive vehicular applications*, 2018, pp. 237–246.

[8] S. Hergeth, L. Lorenz, R. Vilimek, and J. F. Krems, "Keep your scanners peeled: Gaze behavior as a measure of automation trust during highly automated driving," *Human factors*, vol. 58, no. 3, pp. 509–519, 2016.

[9] B. E. Holthausen, P. Wintersberger, B. N. Walker, and A. Riener, "Situational trust scale for automated driving (sts-ad): Development and initial validation," in *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2020, pp. 40–47.

[10] J.-Y. Jian, A. M. Bisantz, and C. G. Drury, "Foundations for an empirically determined scale of trust in automated systems," *International journal of cognitive ergonomics*, vol. 4, no. 1, pp. 53–71, 2000.

[11] S. Khalid Khan, N. Shiwakoti, and P. Stasinopoulos, "A conceptual system dynamics model for cybersecurity assessment of connected and autonomous vehicles," *Accident Analysis & Prevention*, vol. 165, p. 106515, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001457521005467

[12] A. Kunze, S. J. Summerskill, R. Marshall, and A. J. Filtness, "Conveying uncertainties using peripheral awareness displays in the context of automated driving," in *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2019, pp. 329–341.

[13] D. Mortelmans, "Analyzing qualitative data using nvivo," *The Palgrave handbook of methods for media policy research*, pp. 435–450, 2019.

[14] R. Rosenman, V. Tennekoon, and L. G. Hill, "Measuring bias in self-reported data," *International Journal of Behavioural and Healthcare Research*, vol. 2, no. 4, pp. 320–332, 2011.

[15] S. M. Simmons, A. Hicks, and J. K. Caird, "Safety-critical event risk associated with cell phone tasks as measured in naturalistic driving studies: A systematic review and meta-analysis," *Accident Analysis & Prevention*, vol. 87, pp. 161–169, 2016.

[16] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, "Hand keypoint detection in single images using multiview bootstrapping," in *CVPR*, 2017.

[17] V. L. Thing and J. Wu, "Autonomous vehicle security: A taxonomy of attacks and defences," in *2016 ieee international conference on internet of things (ithings) and ieee green computing and communications (greencom) and ieee cyber, physical and social computing (cpscom) and ieee smart data (smartdata)*. IEEE, 2016, pp. 164–170.

[18] Y. Wang, S. Bao, W. Du, Z. Ye, and J. R. Sayer, "Examining drivers' eye glance patterns during distracted driving: Insights from scanning randomness and glance transition matrix," *Journal of safety research*, vol. 63, pp. 149–155, 2017.

[19] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," in *CVPR*, 2016.

[20] T. J. Wright, W. J. Horrey, M. F. Lesch, and M. M. Rahman, "Drivers' trust in an autonomous system: Exploring a covert video-based measure of trust," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 60, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2016, pp. 1334–1338.

[21] B. Yu, S. Bao, Y. Zhang, J. Sullivan, and M. Flannagan, "Measurement and prediction of driver trust in automated vehicle technologies: An application of hand position transition probability matrix," *Transportation research part C: emerging technologies*, vol. 124, p. 102957, 2021.