Project 1928 | July 2020

SJSU SAN JOSÉ STATE UNIVERSITY

MTI MINETA TRANSPORTATION INSTITUTE

A Multimodal Approach for Monitoring Driving Behavior and Emotions

Arash Tavakoli, PhD
Vahid Balali, PhD
Arsalan Heydarian

CSUTC California State University Transportation Consortium

CALIFORNIA STATE UNIVERSITY LONG BEACH

REPORT 20-27

A MULTIMODAL APPROACH FOR MONITORING DRIVING BEHAVIOR AND EMOTIONS

Arash Tavakoli, PhD
Vahid Balali, PhD
Arsalan Heydarian

July 2020

A publication of
Mineta Transportation Institute
Created by Congress in 1991

College of Business
San José State University
San José, CA 95192-0219

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. 20-27	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle A Multimodal Approach for Monitoring Driving Behavior and Emotions		**5. Report Date** July 2020
		6. Performing Organization Code
7. Authors Vahid Balali, PhD, https://orcid.org/0000-0002-5553-7599 Arash Tavakoli, PhD; Arsalan Heydarian		**8. Performing Organization Report** CA-MTI-1928
9. Performing Organization Name and Address Mineta Transportation Institute College of Business San José State University San José, CA 95192-0219		**10. Work Unit No.**
		11. Contract or Grant No. ZSB12017-SJAUX
12. Sponsoring Agency Name and Address State of California SB1 2017/2018 Trustees of the California State University Sponsored Programs Administration 401 Golden Shore, 5th Floor Long Beach, CA 90802		**13. Type of Report and Period Covered** Final Report
		14. Sponsoring Agency Code
15. Supplemental Notes DOI: 10.31979/mti.2020.1928		

16. Abstract

Studies have indicated that emotions can significantly be influenced by environmental factors; these factors can also significantly influence drivers' emotional state and, accordingly, their driving behavior. Furthermore, as the demand for autonomous vehicles is expected to significantly increase within the next decade, a proper understanding of drivers'/passengers' emotions, behavior, and preferences will be needed in order to create an acceptable level of trust with humans. This paper proposes a novel semi-automated approach for understanding the effect of environmental factors on drivers' emotions and behavioral changes through a naturalistic driving study. This setup includes a frontal road and facial camera, a smart watch for tracking physiological measurements, and a Controller Area Network (CAN) serial data logger. The results suggest that the driver's affect is highly influenced by the type of road and the weather conditions, which have the potential to change driving behaviors. For instance, when the research defines emotional metrics as valence and engagement, results reveal there exist significant differences between human emotion in different weather conditions and road types. Participants' engagement was higher in rainy and clear weather compared to cloudy weather. More-over, engagement was higher on city streets and highways compared to one-lane roads and two-lane highways.

17. Key Words Naturalistic Driving Study, Emotions and Behaviors, Human Sensing, Computer Vision, Human-in-the-loop systems	**18. Distribution Statement** No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161		
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 20	**22. Price**

Form DOT F 1700.7 (8-72)

DOI: 10.31979/mti.2020.1928

Mineta Transportation Institute
College of Business
San José State University
San José, CA 95192-0219

Tel: (408) 924-7560
Fax: (408) 924-7565
Email: mineta-institute@sjsu.edu

transweb.sjsu.edu

110520

ACKNOWLEDGMENTS

The author would like to thank the National Center for Transportation, Green Technology, and Education (TransGET) at California State University Long Beach. Any opinions, findings, and conclusions or recommendations expressed in this report are those of the author and do not necessarily reflect the views of the CSUTC or other sponsors.

The author thanks Editing Press, for editorial services, as well as MTI staff, including Executive Director Karen Philbrick, PhD; Deputy Executive Director Hilary Nixon, PhD; Graphic Designer Alverina Eka Weinardy; and Executive Administrative Assistant Jill Carter.

TABLE OF CONTENTS

LIST OF FIGURES

EXECUTIVE SUMMARY

Studies have indicated that emotions can be significantly influenced by environmental factors, contextual settings, and social interactions. These factors can also expressively influence a driver's emotional state, and when it comes to driving, they can significantly affect a person's driving behavior. In addition, studies suggest that different emotions result in specific variations in humans' facial expressions, physiological measures, and specific behaviors. This research proposes to develop a novel semi-automated approach for understanding the effect of environmental factors on drivers' emotions and behavioral changes: specifically, computer vision and data processing techniques are used to classify drivers' emotions based on data collected in environmental settings through a naturalistic driving study. This setup includes a frontal road and facial camera, noise sensors, a smart watch for tracking physiological measurements, and a Controller Area Network (CAN) serial data logger. A framework is proposed for conducting a long-term naturalistic study with participants, where elements such as environmental information and drivers' emotional and behavioral information are automatically monitored and collected. To achieve this, a semantic segmentation using GIST feature is implemented on frontal road videos to automatically understand different environmental factors, such as traffic density, traffic signs, road type, and the amount of vegetation. Moreover, the noise level variation inside and outside of the vehicle is measured by noise sensors. In addition to environmental changes, the research team analyzes the video of the front-seat passengers using the state-of-the-art facial emotional analysis software (e.g., Affectiva SDK) to identify passengers' positive and negative emotions. Through the use of wearables (i.e., smart watches), participants' physiological information (i.e., heart rate and skin temperature) is collected in order to identify sudden physiological changes while driving. This research will provide an effective way for Caltrans to identify how different factors may positively and negatively influence certain driving behaviors by analyzing the changes in drivers' emotions based on different environmental and contextual settings.

If Caltrans wants to implement the methods developed herein at a larger scale, the research team will seek opportunities with Caltrans' current vendors that provide driver behavior analytics, doing so in order to transfer the developed technology. Furthermore, as the demand for autonomous vehicles is expected to significantly increase within the next decade, a proper understanding of drivers'/passengers' emotions, behavior, and preferences will be needed in order to create an acceptable level of trust with humans.

I. INTRODUCTION

According to a recent study by the American Automobile Association (AAA), nearly 80 percent of drivers in the United States experience significant anger, aggression, or road rage annually (Johnson, 2016). Anger, which is referred to here as “road rage,” is not the only emotion that affects driving behaviors. Studies suggest that emotions such as happiness can degrade driving performance and affect the driver’s risk perception (Jeon, Walker, and Yim, 2014; Hu, Xie, and Li, 2013).

Research in psychology suggests that emotions can significantly be influenced by environmental factors; for instance, human emotions are affected by the type of color and picture within the field of view (Nijdam, 2005; AL-Ayash et al., 2016), the noise level in the environment, type of music an individual is listening to (Koelsch, 2014), and so on. As a result, driving behaviors can also be affected by individuals’ emotion and mood (Hu, Xie, and Li, 2013). For instance, the surrounding landscape can have effects on driving behavior.

In addition, autonomous cars are just entering the market and are expected to grow and account for 15 percent of total cars sold by 2030 (“Automotive Revolution & Perspective Towards 2030”, 2016). Prior to the large-scale introduction and adoption of autonomous vehicles, a number of underlying factors related to safety and trust in these systems must be evaluated and applied (Cunningham and Regan, 2015). For instance, with the recent reported crashes of semi-autonomous vehicles, serious concerns have been raised about how these vehicles can be fully adopted and trusted by the consumer (Higgins, Spector, and Colias, 2018). These accidents have flagged a significant concern in terms of safety, trust, and how to keep drivers alert in events of failure, where human decision-making and control need to overtake the autonomous system.

In order to achieve healthy human–vehicle collaboration and trust, in addition to environmental sensing, autonomous vehicles need to be coded with a proper understanding of driver and passengers’ behavioral changes, which is currently non-existent in today’s semi-autonomous vehicles. By coupling environmental sensing and human sensing, one can understand how specific human behaviors and emotions are affected by spatial and temporal environmental factors (e.g., highways vs. country roads, traffic, scenery, and so on). Furthermore, a naturalistic driving study can provide detailed insight about how different drivers respond in various contextual settings; with access to such information, driving profiles can be developed, allowing autonomous vehicles to move towards user-centered autonomy, where they can better respond to the users’ needs, habits, or preferences in real time.

This paper proposes a novel automated approach for understanding the effect of environmental factors on drivers’ emotions and behavioral changes; specifically, computer vision and data processing techniques are used to classify drivers’ emotion and based on environmental settings through a naturalistic driving study. First is a review of the background literature on the effects of environmental factors on driving behaviors; then, a framework is proposed for understanding and identifying the influence of environmental factors on driving behaviors and emotions. Furthermore, a discussion is presented based on the preliminary results and a roadmap set out for future research in this area.

II. BACKGROUND STUDY

Defining behaviors and emotions has a long history in the psychological literature. The definition of emotion is still not entirely clear (Reisenzein, 2007; Burton, 2016). One of the popular definitions identifies an emotion as a complex mental state phenomenon that occurs spontaneously, has neurobiological activity, and is accompanied by psychological and physiological signs (e.g., change in blood pressure, voice tone, and so on) (Izard, 2009; Solomon, 2003). Researchers mostly agree that there exist a fixed number of emotions, called basic emotions, that underlie the other human emotions; however, there exists a debate about the number and type of basic emotions. For instance, Ekman (1992) defines six basic emotions (i.e., anger, disgust, fear, happiness, sadness, and surprise) and Izard (2009) defines ten (i.e., anger, contempt, disgust, distress, fear, guilt, interest, joy, shame, and surprise). One of the recently pursued areas of emotion and behavior research is automating emotion and behavior recognition.

During the past few years, automated emotion recognition, through facial, audio, and physiological cues, has gained more attention, as it has found applications in multiple areas such as marketing, medicine, entertainment, law, etc. The most important cues for understanding emotions in humans are visual. Humans use visual cues (i.e., facial expressions) for understanding each other when expressing emotions such as sadness, anger, happiness, etc. (Ratliff and Patterson, 2008). In addition to visual cues, physiological signs have been used in different areas of emotion recognition, as these measures are highly affected by the type of emotion being experienced; for instance, researchers have used electroencephalogram (EEG) results (Wu et al., 2017; Lan et al., 2016), temperature (T) (Gouizi, Reguig, and Maaoui, 2011; Lisetti and Nasoz, 2004), heart rate (HR) (Valenza et al., 2014; Guo et al., 2016), galvanic skin response (GSR) (Wen et al., 2014), and respiration (RSP) (Wong et al., 2010) to classify different emotional states participants were feeling. Although both facial emotional analysis and physiological cues have been used to understand driver's emotions, using facial analysis has gained more attention as it is less costly and more practical compared to physiological data collection, such as EEGs, which are more intrusive (Vural et al., 2007).

The general procedure for understanding emotional cues using automated facial analysis in a spatial manner (i.e., with a single image) is as follows. (1) The image of a person showing an emotion goes through a pre-processing stage which removes noise from the picture. (2) The face is detected, and different facial features (i.e., eyes, eyebrows, lips, etc.) are tracked and extracted. There are a number of feature extraction algorithms that are currently widely used: e.g., geometric-based versus appearance-based (Kumari, Rajesh, and Pooja, 2015). (3) These features are fed into a classifier, such as support vector machine (SVM) or nearest neighbor (NN), to classify each expression (Kumari, Rajesh, and Pooja, 2015). Using these methods, many studies have attempted to better understand the causes of unsafe behaviors such as fatigue (Zhang and Hua, 2015), aggressive driving (Moriyama, Abdelaziz, and Shimomura, 2012), frustration (Abdíc et al., 2016), cognitive load (Fridman et al., 2018), and so on. Some of the studies, such as the one conducted by Fridman and colleagues (2019), have only used gaze and eye tracking, while others have looked at complete facial analysis (Fridman et al., 2019). Most of the studies were conducted either in a driving simulator or in an experimental setup (Jabon et al., 2011), while very few studies have looked at naturalistic driving behavior (Fridman et al., 2019).

A naturalistic driving study referred as a research on drivers by recording non-intrusive data from a participant in driving, with the goal of keeping the situation as close as possible to a natural driving or so-called driving “in the wild” (Fridman et al., 2019). There have been a multiple naturalistic studies in the past that have aimed at gathering long-term data (Klauer et al., 2006; Neale et al., 2005; Dingus et al., 2006; Regan et al., 2013). However, most of these studies were conducted with the aim of understanding safety and crash-related events, not individual emotion analysis in a human-centered manner.

One of the important environmental factors that can affect drivers’ emotions and behavior is vegetation and the cross-section of the road. Vegetation has a paradoxical two-fold impact on the driving process: (1) the effect of vegetation on the severity of crashes involving drivers in run-off crashes, and (2) the effect of vegetation on driving behavior and emotions (Fitzpatrick, Samuel, and Knodler, 2016).

The effect of trees on crash rate and severity is still not clear. Some studies suggest that having trees in the surrounding area may decrease the crash rate but increase the severity; as a result, these studies propose to remove roadside trees as a safe solution (Zeigler, 1987; Sullivan and Daly, 2005). However, that solution is neither sustainable nor environmentally friendly. Crashes involving fixed objects such as trees represent 46 percent of all fatal crashes (Dixon and Wolf, 2007). The most important reason behind this high fatality rate is that fixed objects are near the roadside; from both an economic and environmental point of view, it is costly to remove trees in large amounts (Fitzpatrick, Samuel, and Knodler, 2016). In addition, it should be noted that trees can make it difficult for the drivers to observe or predict a hazard (Fitzpatrick, Samuel, and Knodler, 2016). On the other hand, some studies (Mok, Landphair, and Naderi, 2006; Zeigler, 1987) suggest that in some cases trees do not increase the crash severity and actually significantly decrease the crash rate. All this together (Dixon and Wolf, 2007) suggests that more data collection is required for understanding the true effect of vegetation and trees in urban areas on the crashes.

Moreover, as autonomous vehicles enter the transportation system, where they can communicate with each other through Internet of Things- (IoT-) enabled infrastructure, crashes due to human errors with roadside fixed elements (e.g., trees, poles, etc.) would be significantly decreased. In this situation, the vegetation and other environmental factors can have a second function: to enhance passengers’ emotions and mood.

This study uses environmental sensors, facial emotion analysis, and heart rate data collected from smart watches to better understand the effect of environmental conditions (e.g., different road types) on driving behavior metrics (i.e., the number of hard accelerations, hard brakes, and so on). Specifically, the following research questions are investigated:

- How do environmental and contextual factors influence drivers’ emotion and behavior?
- Do specific behaviors change with respect to road type and condition?
- Does the amount of vegetation affect drivers’ emotions?

III. METHODOLOGY

A naturalistic driving study platform has been developed for this study and is described below. The pilot study presented in this paper aims to test the system and understand potential improvements.

DATA COLLECTION

The in-cabin data collection setup for this study is depicted in Figure 1. Different streams of data are being collected, including: (1) video of the driver, (2) video of the outside area, (3) Controller Area Network (CAN) data for the driver's trip details, and (4) heart rate of the driver. The videos of the driver and outside road are collected using a Z-Edge S3 Dual Dash Cam and stored on a micro SD card. The videos are then manually transferred to a local computer for processing. The travel data are collected using a CAN data collector. (A CAN bus is a serial broadcast bus which enables communication between all parts of the car system, similar to a microcontroller acting instead of a computer in a car.) For this study, the Automatic Pro device is used as the CAN data retriever. This device connects to the CAN bus serial of the car; collects all data related to travel such as start and end point, hard brakes, hard accelerations, mean speed, fuel usage, etc.; and transfers data to the online account of the user using 3G cellular coverage. The data then can be retrieved from the user's online account as a CSV file. The heart rate of the participant is gathered using a Samsung Gear smart watch and exported as a CSV file.

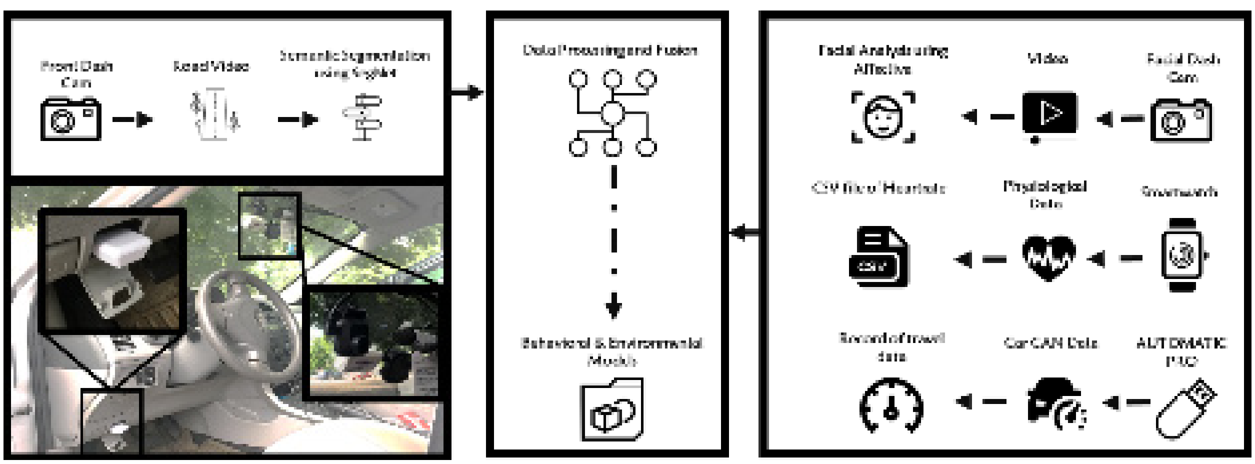

Figure 1. The Proposed Framework for Fusing Data Collected through Human Sensing and Environmental Sensing

DATA ANALYSIS

First, by inspecting the frontal road videos, different types of road are classified. In this paper, four different types of roads are classified, which include: (1) city streets, (2) one-lane roads, (3) two-lane roads, and (4) highways of three or more lanes. The video of the frontal camera is then fed into SegNet (Badrinarayanan, Kendall, and Cipolla, 2017; Badrinarayanan, Handa, and Cipolla, 2015), a semantic segmentation algorithm that recognizes pixels representing a specific object such as a tree or car. Using a preliminary SegNet implementation in MATLAB, the average fraction of the tree in each video is calculated. One frame per second of each frontal video is analyzed to calculate the average. Figure 1, A shows the tree fraction calculated for each road type sample. In addition, Figure 1, C3 shows an example of how the semantic segmentation works, in this instance for defining the tree percentage. In this example, the green color shows the detected trees in a given frame. This implementation of semantic segmentation runs with 85 percent accuracy on defining trees on the training dataset, which is in an acceptable level for this study.

Second, each epoch of road video is categorized based on the weather conditions. The classification categories are 1 (clear weather), 2 (cloudy weather), 3 (very cloudy weather), and 4 (rainy weather). For instance, Figure 1, A3 represents cloudy weather, and A4 represents clear weather.

Videos of the driver and passengers are fed to Affdex SDK (McDuff et al., 2016), a software program for analyzing emotions in videos. As a summary, it first uses the Viola–Jones method for detecting the user's face (Viola and Jones, 2001). Through detecting 34 facial landmarks, the software defines the regions of interest in the picture. Then, a histogram of gradients (HOG) is extracted from the image's regions of interest and is fed to a SVM for classifying the emotions expressed by the driver. For each frame, the percentage of each one of the specific emotions (i.e., contempt, surprise, anger, sadness, disgust, fear, and joy), valence (i.e., a number between -100 and 100 which shows how much the person displays a positive or negative emotion, respectively), and engagement (i.e., a number between 0 to 100 which shows how much the person is expressive of his/her emotion) is reported. For this pilot study, 105 video epochs of one driver have been analyzed using Affdex SDK. Each video has been analyzed using 10 frames per second. An example of the output for one specific frame is depicted in Figure 2, B1 and B2. In this study, the values of engagement and valence were chosen to be the metric for analyzing the emotions. Hence, for each video epoch, the average engagement and valence of the driver are calculated.

Figure 2. (A): Different Road Types; (B): Participants' Emotion and Heart Rate; (C): Car CAN Data and Semantic Segmentation

IV. RESULTS AND DISCUSSION

In order to understand the correlation between heart rate data correlation and driving behavior, heart rate data for each trip should be compared with facial analysis results and CAN data. In this study, heart rate data were collected using Samsung Gear devices, with data retrieved once every two hours. This frequency of data retrieval is not sufficient for a proper understanding of the correlation between heart rate data and other sources of data because, generally, the duration of a trip is not sufficiently long to produce enough heart rate data with this frequency; however, the variations in heart rate give some insight with respect to the proposed method.

As an example, the heart rate data for two specific trips of the participant are discussed. As shown, the two trips happened on July 13 and July 14 (Figure 2, C2) between Northern and Central Virginia (Figure 2, C1). In addition, the heart rate data variation for these two different trips are depicted in Figure 2, C3. Heart rate data from the hours before and after the trips were included as well, as the frequency of heart rate data retrieval was not enough at this point. Comparing the heart rate data and the number of hard accelerations for these two trips reveals that the participant had less hard acceleration events on the day when he experienced significantly lower heart rate. The number of hard accelerations increases from 5 to 13 as the heart rate on average changes from 60 to 80 beats per minute.

For each different road type and weather condition, the average valence and engagement of the video epoch is calculated. Using semantic segmentation, the roads are classified based on the amount of trees they include. Figure 3 shows that there is a significant difference between the tree pixel fraction for different road types, which is aligned with the author's manual classification of the road types. As shown, a mean of tree fraction in the field of view increases from 0.33 (ranging from 0.21 to 0.39) in city streets to 0.37 (ranging from 0.30 to 0.45) in one-lane roads, and it decreases to 0.32 (ranging from 0.24 to 0.43) in two-lane roads and 0.27 (ranging from 0.20 to 0.35) in three-lane highways.

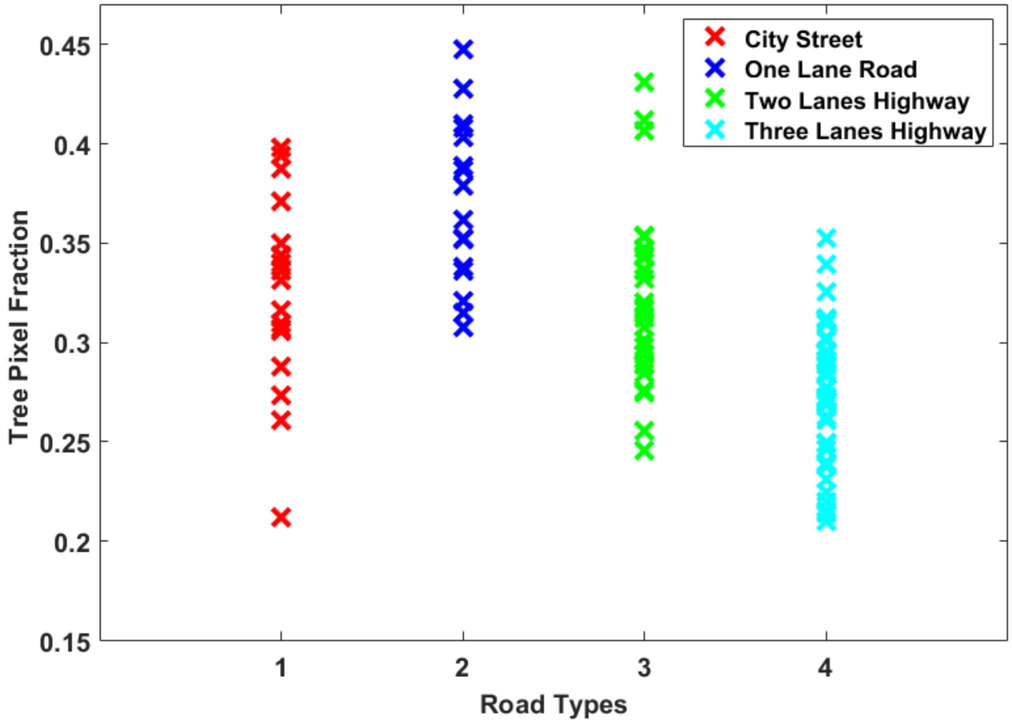

Figure 3. Tree Pixel Fraction vs. Different Road Types

Using facial analysis and the results of semantic segmentation, the engagement and valence values are plotted for each road type (Figure 4). As depicted, there is a significant difference in the value of engagement and valence on different road types. On average, the value of engagement is 8.9 (ranging from 2 to 30) for city streets, 6.5 (ranging from 1.5 to 12) for one-lane roads, 4.5 (ranging from 0 to 12) for two-lane highways, and 9.86 (ranging from 2 to 23) for three-or-more-lane highways. On average, the value of valence is -5.089 (ranging from 0 to -15) for city streets, -3.90 (ranging from 0 to 12) for one-lane roads, -2.06 (ranging from 2 to -7) for two-lane highways, and -3.20 (ranging from 6 to -12) for three-or-more-lane highways. It should be noted that the values of engagement and valence are generally very close to zero for a five-minute-long video (more than 3000 frames on a 10 frame per second rate), meaning that a person generally has a neutral face for a long duration of time, which makes both valence and engagement to be close to zero. Thus, a small change in the average valence or engagement can be counted as significant in a few-minute-long video. Thus, this difference indicates there exist enough frames (a few seconds) with very high values of engagement/valence that could change the average to be above zero.

The variation in emotions (engagement and valence) across different road types (Figure 4) suggests that engagement is a function of road type and scenery within the driver’s field of view. The average value of engagement is highest for both city streets and highways. City streets are surrounded by events that can distract the driver and increase his engagement: the driver might move her head to look at buildings, people, etc. Meanwhile, highways have a higher traffic density, which keeps the driver more focused on cars, further increasing the driver’s engagement expressed through his face. Further inspection of the range of values for engagement depicts that the range is greater in city and highway contexts, which again confirms the above argument. In additions, analyzing the range of valence values reveals that on highways, the valence values can reach positive values as high as 6 for a specific video. Valence values greater than zero are positive emotions, which can give insight into driver's more positive situations while driving in highways. Additionally, the higher traffic density on the highways confirms that the drivers' emotion is also a function of traffic density. As the driver's emotions may affect the driving behavior, it may indicate that certain behaviors are more likely to happen in higher-density traffic; however, more data are required to examine such an effect.

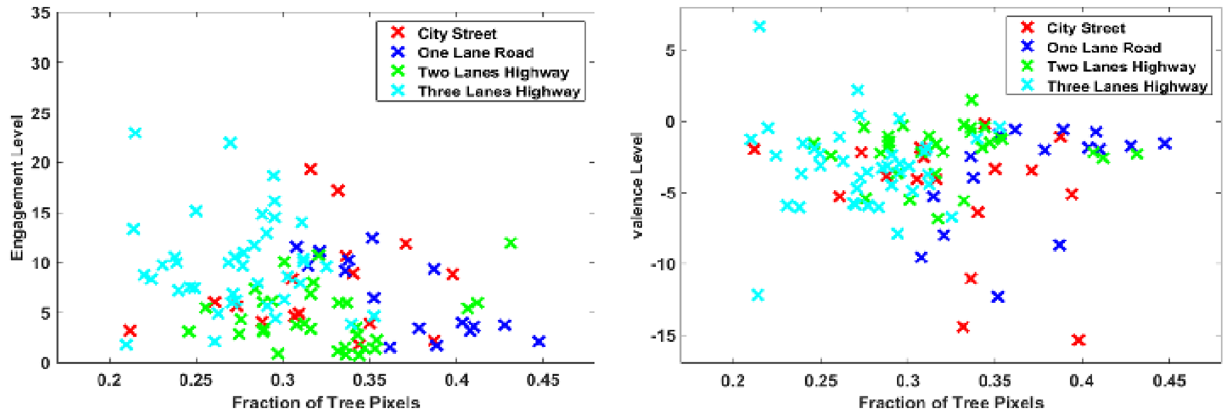

Figure 4. Variations in Engagement and Valence With Respect to Different Road Types

Variations in engagement and valence values for each weather type are plotted in Figure 5. The engagement value is higher in the clear sky and rainy situations; on average, the values for weather conditions 1, 2, 3, and 4 are 8.4 (ranging from 3 to 12), 5.64 (ranging from 0 to 30) , 6.33 (ranging from 0 to 18), and 10.95 (ranging from 3 to 22), respectively. The valence level is more negative in clear sky and rainy conditions; on average, the values for weather conditions 1, 2, 3, and 4 are -4.3 (ranging from -5 to -12), -2.9 (ranging from 1 to -15), -3.03 (ranging from 0 to -7), and -3.31 (ranging from 6 to -12), respectively. As Figure 5 reveals, the clear and rainy cases have the highest engagement by the driver. The clear weather increases the sunshine on the driver, which by itself might affect the driver's vision and naturally increase his engagement. In addition, rainy weather increases the driver's attention, which may affect his/her engagement.

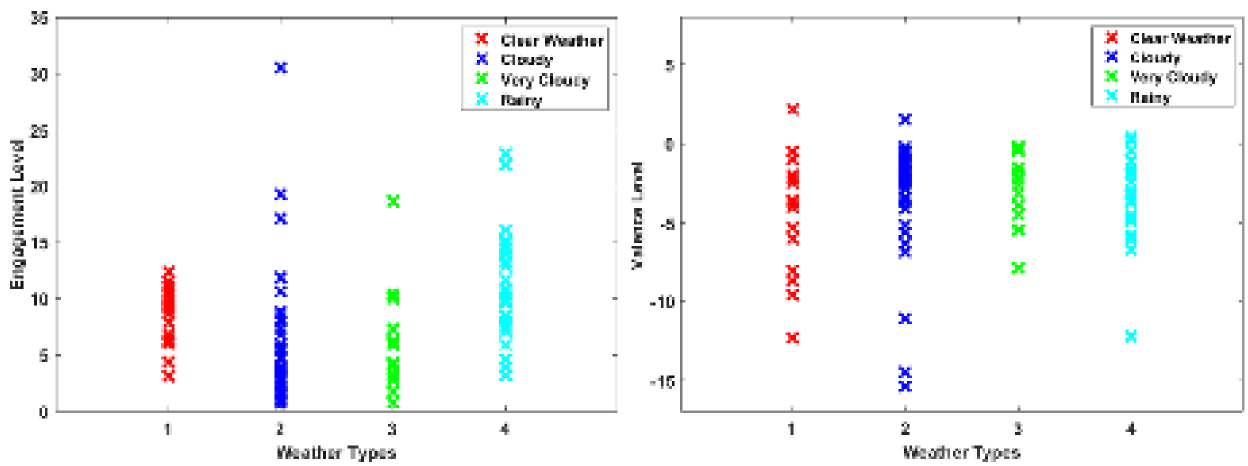

Figure 5. Variations in Engagement and Valence Level in Different Weather Conditions

V. CONCLUSION, LIMITATIONS, AND FUTURE WORK

Environmental factors significantly affect driving behaviors. In order to better understand these effects, this paper has proposed a novel approach for considering different modalities of human sensing data, together with environmental sensing. The study includes a naturalistic driving setup consisting of videos recording both the driver and the road, heart rate data, and data from the car's CAN serial bus. One hundred and five (105) epochs of frontal and facial videos have been analyzed using Affdex SDK, semantic segmentation, and manually annotated videos. Results suggest that weather conditions and road type may significantly change driver emotions and driving behavior. However, more naturalistic driving data from different drivers are needed to better identity the influence of emotions on driving behavior.

Through a pilot study and preliminary data analysis, it was possible to classify the road types by quantifying the number of tree pixels in each frame. It should be noted that only the tree pixel percentage was considered, whereas if the research had included pixel percentage of other environmental factors affecting drivers' behavior, such as traffic density, it would be possible to more accurately classify the road type condition for the behavior model. As part of future work, researchers should increase the accuracy of semantic segmentation by improving/increasing the training data as well as using semantic segmentation to automatically examine more properties of the road, weather, and traffic.

The heart rate data show that, in certain driving situations, there can be significant differences in the heart rate of the driver. Although this difference does not have statistical significance due to the fact that this result only reflects one participant, it reveals that using heart rate can be very helpful for understanding driving-related emotions and behaviors; specifically, in certain driving situations, heart rate can be a significant indicator of certain emotions or behaviors. This claim is aligned with previous studies showing that emotions have significant effects on humans' physiological signs such as heart rate (Nardelli et al., 2015; AL-Ayash et al., 2016). In addition, considering the heart rate data from the hours before and after the driving shows that the driving behavior may be a function of the driver's state before starting the driving process. This illustrates that a proper model of driving behavior might need to also consider the history of the driver's time spent beginning some point before the start of a journey.

The results of emotion detection in the analyzed videos reveal valuable insight with respect to overall emotion recognition; however, inspecting the videos frame-by-frame reveals situations for which facial recognition is not sufficient for understanding the drivers'/passengers' emotions. It should be noted that while this incorrect detection might not be of a high value for today's manually driven cars, the data gain more value when looking at the driver in a semi-autonomous or fully autonomous vehicle, where the driver herself will be a passenger. For instance, in the case of a driver talking to a passenger or eating food, as the cars move towards being autonomous systems, these situations will take place more often and in longer duration compared to a manually driven car where the driver is mostly focused on driving. Some of the identified false positive emotions are listed below:

- There are specific human facial expressions that are similar to specific emotions but actually represent a neutral status of that specific face (Figure 6, A1). For instance, a person might generally seem to show the emotion disgust, the software recognizes it as disgust, but that is his/her general facial style. In addition, there are situations wherein the participant does not show his actual emotions on their face. For instance, the person might seem neutral, but she is actually sad or angry. In this case, using other sources of data such as physiological measures is necessary to more accurately predict the correct emotion.

- There are specific situations wherein the drivers' or passengers' faces cannot be detected or will be detected with a false emotion due to the angle of the face. The Affdex software, like many other facial analysis software programs, works at its best accuracy while the head is not angled more than 25 degrees relative to the camera (Affectiva, n.d.). These situations of facial tilt include talking with the passenger, looking at one's cell phone, looking in the side mirror, eating food, etc. For instance, Figure 6, A2 shows certain situations wherein the software was not able to detect the face or detected a wrong emotion.

- There are certain situations when software can classify the engagement but is not able to classify the emotion. Figure 6, B1 shows an example of this situation.

- There are certain situations when the driver is clearly showing an emotion but the software does not detect any emotion (Figure 6, B2)

Figure 6. Variations in False Positive Detections

In addition, it should be noted that this study considered engagement and valence as emotional traits. Future work should look more into (1) the measurement of specific emotions (e.g., happiness, sadness, etc.), and (2) the interaction between emotion and cognition in the framework of a driving situation, where the participant's real-time emotions might interact with the cognitive load, perhaps a result of the workload of the driver.

Aforementioned situations affirm the value of multimodal analysis. As seen in the scenarios discussed above where the software fails to correctly classify the emotion, there needs to be a source of data other than facial analysis for verification, to confirm or deny the result of the facial analysis. For instance, in the cases for which facial analysis is not capable of detecting the face, other sources of data could include (1) human sensing data such as physiological signs and the driver’s audio; (2) environmental sensing data such as the type of road and weather condition. These should be fused with facial analysis to correctly classify drivers’ emotions. Thus, there needs to be a systematic approach to detecting from all sources of non-intrusive data collection; understanding the weaknesses and accuracy of each source of data based on the contextual setting (e.g. situations when facial analysis failure can be covered by heart rate analysis); and proposing a generalized human sensing model for in-cabin experience.

ABBREVIATIONS AND ACRONYMS

AAA	American Automobile Association
CalTrans	California Department of Transportation
CAN	Controller Area Network
GSR	Galvanic Skin Response
HOG	Histogram of Oriented Gradients
HR	Heart Rate
IoT	Internet of Things
NN	Nearest Neighbor
RDD	Random Digit Dialing
SVM	Support Vector Machine

BIBLIOGRAPHY

Abdíc, I., Fridman, L., McDuff, D., Marchi, E., Reimer, B., and Schuller, B. (2016). Driver Frustration Detection from Audio and Video in the Wild. In S. Kambhampati (Ed.), *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (pp. 1354-1360). AAAI Press.

Affectiva. (n.d.). *Affectiva Automotive AI for Driver Monitoring Systems*. https://www.affectiva.com/product/affectiva-automotive-ai-for-driver-monitoring-solutions/

AL-Ayash, A., Kane, R. T.,Smith, D., and Green-Armytage, P. (2016). The Influence of Color on Student Emotion, Heart Rate, and Performance in Learning Environments. Color Research & Application, 41(2), 196–205. https://doi.org/10.1002/col.21949.

Automotive Revolution & Perspective Towards 2030. (2016). *Auto Tech Review, 5*(4), 20–25. https://doi.org/10.1365/s40112-016-1117-8.

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(12), 2481-2495. https://doi.org/10.1109/TPAMI.2016.2644615

Badrinarayanan, V., Handa, A., and Cipolla, R. (2015). Segnet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling. https://arxiv.org/pdf/1505.04366.pdf

Burton, N. (2016, January 3). *What Is an Emotion? The Answer Is Not Entirely Clear*. Psychology Today. https://www.psychologytoday.com/us/blog/hide-and-seek/201601/what-is-emotion.

Cunningham, M., and Regan, M. A. (2015). Autonomous Vehicles: Human Factors Issues and Future Research. *Proceedings of the 2015 Australasian Road Safety Conference* (pp. 14-16). https://pdfs.semanticscholar.org/f62e/ed1494efcc1247933b93c46d51ca8656142b.pdf?_ga=2.212166684.928594510.1592323535-736533661.1591813491

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J. D., Perez, M. A., Hankey, J., Ramsey, D. J., Gupta, S., Bucher, C., Doerzaph, Z. R., Jermeland, J., and Knipling, R. R. (2006). *The 100-Car Naturalistic Driving Study, Phase II: Results of the 100-Car Field Experiment* (DOT HS 810 593). https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/100carmain.pdf

Dixon, K. K., and Wolf, K. L. (2007). Benefits and Risks of Urban Roadside Landscape: Finding a Livable, Balanced Response. *Proceedings of the 3rd Urban Street Symposium*. Transportation Research Board of the National Academies of Science. https://nacto.org/docs/usdg/benefits_and_risks_of_an_urban_roadside_ landscape_dixon.pdf

Ekman, P. (1992). An Argument for Basic Emotions. *Cognition & Emotion, 6*(3/4), 169–200. http://www.paulekman.com/wp-content/uploads/2013/07/An-Argument-For-Basic-Emotions.pdf

Fitzpatrick, C. D., Samuel, S, and Knodler, M. A. (2016). Evaluating the Effect of Vegetation and Clear Zone Width on Driver Behavior Using a Driving Simulator. *Transportation Research Part F: Traffic Psychology and Behaviour*,*42*, 80–89. https://doi.org/10.1016/j.trf.2016.07.002

Fridman, L., Langhans, P., Lee, J., and Reimer, B. (2016). Driver Gaze Region Estimation without Use of Eye Movement. *IEEE Intelligent Systems, 31*(3), 49–56. https://doi.org/10.1109/MIS.2016.47

Fridman, L., Brown, D. E., Glazer, M., Angell, W., Dodd, S., Jenik, B., Terwilliger, J., Aleksandr, P., Kindelsberger, J., Ding, L., Seaman, S., Mehler, A., Sipperley, A., Pettinato, A., Seppelt, B., Angell, L., Mehler, B., and Reimer, B. (2019). MIT Autonomous Vehicle Technology Study: Naturalistic Driving Study of Driver Behavior and Interaction with Automation. *IEEE Access, 7*, 102021 - 102038. https://doi.org/10.1109/ACCESS.2019.2926040

Fridman, L., Reimer, B., Mehler, B., and Freeman, W. T. (2018). Cognitive Load Estimation in the Wild. *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* Association for Computing Machinery, New York, United States. https://doi.org/10.1145/3173574.3174226

Gouizi, K., Reguig, F. B., and Maaoui, C. (2011). Analysis Physiological Signals for Emotion Recognition. *International Workshop on Systems, Signal Processing and their Applications, WOSSPA* (pp. 147-150). https://doi.org/10.1109/WOSSPA.2011.5931436

Guo, H., Huang, Y., Lin, C., Chien, J., Haraikawa, K., and Shieh, J. (2016). Heart Rate Variability Signal Features for Emotion Recognition by Using Principal Component Analysis and Support Vectors Machine. *2016 IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE)* (pp. 274-277). https://doi.org/10.1109/BIBE.2016.40

Higgins, T., Spector, M., and Colias, M. (2018, April 2). *Tesla, Uber Deaths Raise Questions About the Perils of Partly Autonomous Driving. T*he Wall Street Journal. https://www.wsj.com/articles/tesla-uber-deaths-raise-questions-about-the-perils-of-partly-autonomous-driving-1522661400

Hu, T. Y., Xie, X., and Li, J. (2013). Negative or Positive? The Effect of Emotion and Mood on Risky Driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 16*, 29–40. https://doi.org/10.1016/j.trf.2012.08.009.

Izard, C. E. (2009). Emotion Theory and Research: Highlights, Unanswered Questions, and Emerging Issues. *Annual Review of Psychology*, *60*, 1–25. https://doi.org/10.1146/annurev.psych.60.110707.163539.Emotion.

Jabon, M., Bailenson, J., Pontikakis, E., Takayama, L., and Nass, C. (2011). Facial Expression Analysis for Predicting Unsafe Driving Behavior. *IEEE Pervasive Computing, 10*(4), 84–95. https://doi.org/10.1109/MPRV.2010.46

Jeon, M., Walker, B. N., and Yim, J. B. (2014). Effects of Specific Emotions on Subjective Judgment, Driving Performance, and Perceived Workload. *Transportation Research Part F: Traffic Psychology and Behaviour, 24*, 197–209. https://doi.org/10.1016/j.trf.2014.04.003

Johnson, T. (2016, July 14). *Nearly 80 Percent of Drivers Express Significant Anger, Aggression or Road Rage*. AAA Newsroom. https://newsroom.aaa.com/2016/07/nearly-80-percent-of-drivers-express-significant-anger-aggression-or-road-rage/.

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., and Ramsey, D. J. (2006). T*he Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data* (DOT HS 810 594). https://pdfs.semanticscholar.org/3ba7/6db90b340d57a9aed9160f76ff7496e011fe.pdf?_ga=2.217917214.928594510.1592323535-736533661.1591813491

Koelsch S. (2014). Brain correlates of music-evoked emotions. *Nature reviews Neuroscience, 15*(3), 170–180. https://doi.org/10.1038/nrn3666

Kumari, J., Rajesh, R., and Pooja, K. M. (2015). Facial Expression Recognition: A Survey. *Procedia Computer Science, 58,* 486–491. https://doi.org/10.1016/j.procs.2015.08.011

Lan, Z., Sourina, O., Wang, L., and Liu, Y. (2016). Real-Time EEG-Based Emotion Monitoring Using Stable Features. *The Visual Computer*, *32*(3),347–358. https://doi.org/10.1007/s00371-015-1183-y

Lisetti, C. L., and Nasoz, F. (2004). Using Noninvasive Wearable Computers to Recognize Human Emotions from Physiological Signals. *EURASIP Journal on Advances in Signal Processing*, *11*, 1672–1687. https://doi.org/10.1155/S1110865704406192

McDuff, D., Mahmoud, A. N., Mavadati, M., Amr, M., Turcot, J., and Kaliouby, R. E. (2016). AFFDEX SDK: A Cross-Platform Real-Time Multi-Face Expression Recognition Toolkit. *CHI EA ‘16: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 3723-3726). https://doi.org/10.1145/2851581.2890247

Mok, J., Landphair, H. C., and Naderi, J. R. (2006). Landscape Improvement Impacts on Roadside Safety in Texas. *Landscape and Urban Planning,* 78(3), 263–274.

https://doi.org/10.1016/j.landurbplan.2005.09.002

Moriyama, T., Abdelaziz, K., and Shimomura, N. (2012). Face Analysis of Aggressive Moods in Automobile Driving Using Mutual Subspace Method. *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)* (pp. 2898-2901). IEEE. https://projet.liris.cnrs.fr/imagine/pub/proceedings/ICPR-2012/media/files/1801.pdf

Nardelli, M., Valenza, G., Greco, A., Lanata, A., and Scilingo, E. P. (2015). Recognizing Emotions Induced by Affective Sounds through Heart Rate Variability. *IEEE Transactions on Affective Computing, 6*(4), 385–394. https://doi.org/10.1109/TAFFC.2015.2432810.

Neale, V. L., Dingus, T. A., Klauer, S. G., Sudweeks, J., and Goodman, M. (2005). An Overview of the 100-Car Naturalistic Study and Findings. *Proceedings - 19th International Technical Conference on the Enhanced Safety of Vehicles (ESV)*. National Highway Traffic Safety Administration. https://pdfs.semanticscholar.org/0024/aa33b7db2e0054ce106a7b9a2676d463a886.pdf?_ga=2.208012058.928594510.1592323535-736533661.1591813491

Nijdam, N. A. (2005). Mapping Emotion to Color. https://pdfs.semanticscholar.org/5f0d/e6e7bc1d5443243f9f42f2379db9639a933d.pdf?_ga=2.117321486.928594510.1592323535-736533661.1591813491

Ratliff, M. S., and Patterson, E. (2008). Emotion Recognition Using Facial Expressions with Active Appearance Models. *HCI '08: Proceedings of the Third IASTED International Conference on Human Computer Interaction* (pp. 138-143). http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.503.944&rep=rep1&type=pdf

Regan, M. A., Williamson, A. M., Grzebieta, R., Charlton, J. L., Lenne, M. G., Watson, B. C., Haworth, N. L., Rakotonirainy, A., Woolley, J. E., Anderson, R. W. G., Senserrick, T., and Young, K. L. (2013). In N. Leal (Ed.), The Australian 400-Car Naturalistic Driving Study: Innovation in Road Safety Research and Policy. *Proceedings of the 2013 Australasian Road Safety Research, Policing and Education Conference* (pp. 1 - 13). Australasian College of Road Safety

Reisenzein, R. (2007). What is a definition of emotion? And are emotions mental-behavioral processes? *Social Science Information, 46*(3), 424–428. https://doi.org/10.1177/05390184070460030110

Solomon, R. C. (2003). *What Is an Emotion? Classic and Contemporary Readings*. Oxford University Press

Sullivan, E., and Daly, J. (2005). Investigation of Median Trees and Collisions on Urban and Suburban Conventional Highways in California. *Transportation Research Record: Journal of the Transportation Research Board, 1908*(1), 114–120. https://doi.org/10.1177/0361198105190800114

Valenza, G., Citi, L., Lanatá, A., Scilingo, E. P., and Barbieri, R. (2014). Revealing Real-Time Emotional Responses: A Personalized Assessment Based on Heartbeat Dynamics. *Scientific Reports, 4*. https://doi.org/10.1038/srep04998

Viola, P., and Jones, M. (2001). Rapid Object Detection Using a Boosted Cascade of Simple Features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. I–I). https://doi.org/10.1109/CVPR.2001.990517

Vural, E., Cetin, M., Ercil, A., Littlewort, G., Bartlett, M., and Movellan, J. (2007). Drowsy Driver Detection through Facial Movement Analysis. In Lew, M., Sebe, N., Huang, Th.S., Bakker, E.M. (Eds.), *Human-Computer Interaction: International Workshop, HCI 2007 Rio de Janeiro, Brazil, October 20, 2007 Proceedings, vol 4796.* Springer, Berline, Heidelberg. https://doi.org/10.1007/978-3-540-75773-3_2

Wen, W., Liu, G., Cheng, N., Wei, J., Shangguan, P., and Huang, W. (2014). Emotion Recognition Based on Multi-Variant Correlation of Physiological Signals. *IEEE Transactions on Affective Computing, 5*(2), 126-140. https://doi.org/10.1109/TAFFC.2014.2327617

Wong, W. M., Tan, A. WC., Loo, C. K., and Liew, W. S. (2010). PSO Optimization of Synergetic Neural Classifier for Multichannel Emotion Recognition. *Proceedings of 2010 Second World Congress on Nature and Biologically Inspired Computing (NaBIC)* (pp.316-321). IEEE. https://doi.org/10.1109/NABIC.2010.5716292

Wu, S., Xu, X., Shu, L., and Hu, B. (2017). Estimation of Valence of Emotion Using Two Frontal EEG Channels. *Proceedings of 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp.1127-1130). IEEE. https://doi.org/10.1109/BIBM.2017.8217815

Yang, L., Li, X., Guan, W., Zhang, H. M., and Fan, L. (2018). Effect of Traffic Density on Drivers' Lane Change and Overtaking Maneuvers in Freeway Situation-A Driving Simulator Based Study. *Traffic Injury Prevention, 19*(6), 594-600. https://doi.org/10.1080/15389588.2018.1471470

Zeigler, A. J. (1987). Risk of Vehicle-Tree Accidents and Management of Roadside Trees. *Transportation Research Record* 1127, 37-43. http://onlinepubs.trb.org/Onlinepubs/trr/1987/1127/1127-006.pdf

Zhang, Y., and Hua, C. (2015). Driver Fatigue Recognition Based on Facial Expression Analysis Using Local Binary Patterns. *Optik: International Journal for Light and Electron Optics,126*(23), 4501–4505. https://doi.org/10.1016/j.ijleo.2015.08.185

ABOUT THE AUTHOR

ARASH TAVAKOLI

Arash Tavakoli is a PhD candidate, Engineering Systems and Environment (ESE) at the Link Lab of the University of Virginia. He has earned his BSc. and MSc. in Civil Engineering from the Sharif University of Technology and Virginia Tech respectively. Arash's research interest lies on the intersection of psychology, transportation engineering, and computer science. Specifically, Arash implements psychological theories and experimental design together with affective computing and computer vision techniques to enhance human-autonomous vehicle interaction.

VAHID BALALI

The Principal Investigator, Vahid Balali, PhD, is an Assistant Professor in the Department of Civil Engineering and Construction Engineering Management at California State University Long Beach. Dr. Balali's research focuses on visual data sensing and analytics, virtual design and construction for civil infrastructure and interoperable system integration, and smart cities in transportation for sustainable decision-making. He also has experience as a visual data analyst and has developed a video-based construction resource tracking and action recognition for activity analysis of operators at Caterpillar. He has the knowledge, technical skills, and experience that are crucial to the successful completion of the proposed work.

Dr. Vahid Balali has been named a recipient of the 2020 Early Academic Career Excellence Award by California State University Long Beach. He was also selected as one of the Top 40 under 40 by the Consulting-Specifying-Engineer for the year 2017 and as the top young professional in California by the Engineering News Record (ENR) for the year 2016. He received the 2014 second-best poster award from the Construction Research Congress, as well as the 2013 CMAA national capital chapter scholarship award. He is currently an associate member of ASCE and CMAA, a committee member of the ASCE Data Sensing and Analysis and ASCE Visual Information Modeling and Simulation committees, and a friend member of relevant TRB committees. He is also serving as a reviewer of several top-tier journals. He is actively collaborating with industrial partners and is involved in professional and outreach activities.

ARSALAN HEYDARIAN

Arsalan Heydarian is an Assistant Professor in the Department of Engineering Systems and Environment at the University of Virginia and he's part of the the UVA Cyber-physical systems lab, the Link Lab. His research broadly focuses on user-centered design, construction, and operation of intelligent infrastructure with the objective of enhancing their sustainability, adaptability, and resilience. Dr. Heydarian received his Ph.D. in Civil Engineering from the University of Southern California (USC), M.Sc in System Engineering from USC, and B.Sc. and M.Sc in Civil Engineering from Virginia Tech.

PEER REVIEW

San José State University, of the California State University system, and the Mineta Transportation Institute (MTI) Board of Trustees have agreed upon a peer review process required for all research published by MTI. The purpose of the review process is to ensure that the results presented are based upon a professionally acceptable research protocol.