2022 26th International Conference on Pattern Recognition (ICPR)
August 21-25, 2022, Montréal, Québec, Canada

Contact Versus Noncontact Detection of Driver's Drowsiness

Salem Sharak
University of Michigan
Dearborn, Michigan
sharak@umich.edu

Kapotaksha Das
University of Michigan
Dearborn, Michigan
takposha@umich.edu

Kais Riani
University of Michigan
Dearborn, Michigan
kriani@umich.edu

Mohamed Abouelenien
University of Michigan
Dearborn, Michigan
zmohamed@umich.edu

Mihai Burzo
University of Michigan
Flint, Michigan
mburzo@umich.edu

Rada Mihalcea
University of Michigan
Ann Arbor, Michigan
mihalcea@umich.edu

2022 26th International Conference on Pattern Recognition (ICPR) | 978-1-6654-9062-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICPR56361.2022.9956336

***Abstract*—With an estimated number of injuries in the millions, accidents caused due to drowsy driving remain a significant source of financial costs and loss of life. Accurate detection of driver's drowsiness could provide a clear avenue towards eliminating a great majority of the associated accidents and losses. Existing research on the subject could be defined as either contact-based or noncontact-based alertness detection. This paper utilizes a novel multimodal driver's alertness dataset consisting of 45 subjects via seven recorded channels, including four contact-based and three noncontact-based channels, to investigate the performance of said modalities in detecting driver's drowsiness as well as provide a novel comparison between the results of multiple contact and noncontact methods. Our results highlight the viability of noncontact methods to detect driver's drowsiness as an implementable technology in automobiles.**

I. Introduction

Driving accidents pose one of the great dangers globally, with the Center for Disease Control and Prevention (CDC) in 2020 estimating about three million non-fatal injuries yearly in the United States alone [1]. Additionally, the World Health Organization (WHO) reported in 2020 that annual road traffic deaths are estimated to have reached 1.35 million globally [2]. In addition to the significant number of injuries and loss of life, there are heavy costs impacted by motor vehicle accidents. The CDC estimated a loss of $75 billion dollars in 2017 from medical care costs and productivity losses associated with injuries and deaths from crashes [3]. Additionally, it is estimated by Chen et al. [4] that the global macroeconomic burden of road injuries between 2015 and 2030 will reach $1.8 trillion, a figure that does not include direct losses to capital.

Of the causes most associated with traffic accidents is driver's drowsiness, with an estimated seven percent of traffic accidents in the United States in 2017, including 16 percent of fatal accidents [5]. Additionally, in a study by AAA, it was found that about 9% of all crashes and over 10% of all crashes involving property damage, airbag deployment, or injury, involved driver's drowsiness [6]. Other studies indicate that drowsiness might be significantly under-reported, with data showing drowsiness as a contributing factor in 20% of all crashes [7]. Drowsiness is significant as it results in impaired reaction times, attention, mental processing, judgement, and decision making according to the National Highway Traffic Safety Administration (NHTSA) [8]. Additionally, previous work has shown that prolonged drowsiness results in driving behavior similar to driving under the influence of alcohol [9], [10], [11].

Given the severity of the problem, there is a great deal of interest in addressing this issue, especially when discussed in the context of motor vehicle crashes. Driver's alertness has been researched extensively, including earlier work by Brookhuis and Waard [12], which used contact-based sensors, such as an electrocardiogram (ECG) to model driver's states, as well as the work by Reimer and Mehler [13], which used an ECG and skin conductance electrodes to model driver's states. Other recent work focused on modeling driver's alertness built on data from sensors, such as electroencephalogram (EEG), respiration rate and skin temperature sensors, among others [14], [15], [16]. Finally, works such as those by Naqvi et al. [17], Raorane et al. [18], Lopez et al. [19], and Kiashari et al. [20], among others utilized noncontact methods, including visual, and more recently, near infra-red (NIR) and thermal imaging to model driver's drowsiness.

In this paper, we present two main contributions. Firstly, we model driver's alertness and provide a novel comparison between the performance of the different recorded channels. Secondly, we introduce a novel multimodal dataset of 45 subjects, with each subject recorded across two recording sessions in a simulated driver's environment, and each recording session consisting of two segments, as detailed in section III. Importantly, this contribution includes eleven recorded channels, as detailed in section III. For comparison, a review of existing datasets found in literature can be found in Table I. In review, the strengths of our dataset when compared to previous work would include the relatively large subject body as well as the large number of recorded channels and their varied source modalities. Of the eleven available channels we use seven in this study, namely: All four physiological channels,

one visual RGB, one NIR, and one thermal channel. Using our two main contributions, we aim to answer the following research questions:

- How well do noncontact-based approaches compare to contact-based methods?
- Does the integration of multiple noncontact-based modalities as well as the combination of multiple contact-based signals improve drowsiness detection?

II. Related Work

Several studies have been conducted in the past to detect driver alertness using different modalities, such as visual, physiological, behavioral, and environmental data. The vast majority of them used only one modality, while more recent work combined multiple modalities to detect driver's alertness [38], [39], [40], [41], [31], [42].

As modern cars are being fitted with more advanced sensors, drowsiness modeling utilizing contact-based methods is still attracting the attention of related research [43]. The review provided by Chowdhury et al. [44] showed that the type of physiological signals did not see a drastic change as the community keeps addressing the topic based on signals related to respiration, heart rate, skin conductance, ECG, EEG, and skin temperature. However, research has gradually moved from statistical correlation to driver-centered behavioral models focusing on the aforementioned signals. In the study of 2017 conducted by Awais et al. [14], results showed that combining both electrocardiogram (ECG) and electroencephalogram (EEG) information could lead to better drowsiness detection. Whereas, more recently in 2019, Persson et al. [15] utilized ECG signals and were able to categorize different levels of alertness by identifying specific features of interest. Finally, the work proposed by Papakostas et al. in 2021 [16] used a deep learning method on four different physiological signals, including Blood Volume Pulse (BVP), respiration, skin conductance, and skin temperature to detect both distracted and drowsy driving.

Despite physiological features showing great potential in detecting driver's alertness, the necessity of contact-based sensor intrusiveness is a major setback to the practicality of the approach. For that reason, researchers explored noncontact methods, such as visual and thermal. The work proposed by D'Orazio et al. [45] in 2007 was an early work published on this topic, which implemented a neural classifier to detect driver's drowsiness by detecting the closure of eyes in image sequences. Several articles have been published since, addressing the same problem using novel and sophisticated computational methods. A set of facial and head movement features were used by Mbouna et al. [46] to detect driver's alertness. In more recent works, deep-learning methods have been used on visual features to tackle this problem. Naqvi et al. [17] suggested a deep learning-based gaze detection system based on a NIR camera sensor that takes the driver's head and eye expression into account. In order to obtain the gaze features, the authors used a pre-trained CNN model to fine-tune the extracted cropped images of the face, left eye, and right eye using the VGG-face network.

As an alternative noncontact method, researchers began to explore thermal imaging as a new noncontact-based modality. Lopez et al. [19] used a Therm-App mobile thermal camera to classify fatigued individuals by detecting, segmenting, and aligning thermal facial images based on coordinates of the eye and nose. They then utilized convolutional neural networks to extract facial feature vectors, followed by using a Support Vector Machine (SVM) to determine a subject's condition as in a state of fatigue or a state of rest. A study by Kiashari et al. [24] showed that a person's respiratory state can be tracked with thermal imaging without interference to monitor the variation in the driver's respiration rate from wakefulness to drowsiness. More recently, Kiashari et al. [20] extended their previous work by implementing a SVM and K-Nearest Neighbor (KNN) classifiers on 30 subjects to detect drowsiness.

III. Dataset

We collected a multimodal dataset from 45 subjects for our experiments, which included thermal, visual, acoustic, and physiological data. Of these 45 subjects, 30 were male and 15 were female. Out of the 30 male participants six were White/Caucasian, and 24 participants were Asian/Middle Eastern. The group of female subjects consisted of five White/Caucasians and 10 Asian/Middle Eastern subjects. In total 11 participants were considered White/Caucasians and 34 were Asians/Middle Eastern. All subjects were between 20 and 33 years old. For the purposes of this research we focus entirely on the thermal, visual, and physiological modalities . The data was collected during two recording sessions. The recordings were taken while the subjects were driving in a driver's simulator and included a recording taken usually before 11 a.m. followed by an afternoon recording, usually between 4 p.m. and 8 p.m. This setting was chosen to resemble the typical day-to-day routine of 'nine-to-five' employees as closely as possible, where they are alert as they head to their place of employment and drowsy after a long day of work. According to [47], the 4:00 to 8:00 time slot was the slot with the highest crashes throughout the day for both fatal and nonfatal crashes in 2020, which coincided with our approach. This is also similar to the approach taken by [48] and [49], in which recordings taken earlier in the day are categorized as alert, while later recordings are set as drowsy. We asked all participants to schedule the earlier recording as their first activity of the day in order to make sure the subjects were alert while driving. On the other hand, the second recording was captured later in the day to simulate drowsiness, where the subjects were explicitly warned not to nap the whole day prior to the recording time. Every session consisted of two separate sub-recordings, 'baseline' and 'free-driving'.

For the first half of the baseline recording, participants were asked to remain still without moving, breathe naturally, and look at the center of the central camera. In the second half, they were asked to follow a target shown on the screens

Paper Reference	No. of Channels	Modalities Used	No. of Subjects
[21]	2	Thermal, Visual RGB	6
[22]	3	EOG (Electrooculogram), ECG (Electrocardiogram), EEG (Electroencephalogram)	6
[23]	7	ECG, EOG, EEG, EMG (Electromyogram), Visual RGB, NIR (Near Infrared), Depth Map	14
[24]	2	NIR, Thermal	12
[25]	2	Cardiopulmonary, Thermal	12
[26]	1	ECG	12
[27]	4	Vehicle telemetry, Self-reported drowsiness ratings, ECG, EMG	15
[28]	1	Thermal	1
[29]	10	Eight physiological channels, Visual RGB, NIR	60 (All Male)
[30]	6	Visual RGB, NIR Depth map, Audio, Pulse, Vehicle telemetry	8
[31]	3	Visual RGB, Audio, Vehicle telemetry	30
[32]	4	Vehicle telemetry, Eye closure, Head position, Visual RGB	13
[14]	5	Four visual RGB cameras, ECG	22
[33]	4	EOG, Vehicle telemetry, Eye tracking, Survey	16
[34]	2	Visual RGB, NIR Depth Map, Audio, Pulse, Vehicle telemetry	36
[35]	4	Visual RGB, ECG, Pulse, Physioloigcal	30
[36]	3	EEG, EOG, EMG	10
[37]	3	EEG, Two NIR cameras	38

TABLE I: A review of existing datasets found in literature.

with their gaze while sitting normally. Each half lasted about 2.5 minutes on average. This was followed by the free-driving recording, where the participants were directed to drive using the virtual environment for approximately 15 minutes. The subjects started driving on a low-traffic highway in the simulated environment and then were allowed to freely stay on the highway or divert to city-like driveways. No pedestrians were included and clear weather conditions under daytime were selected for our simulations.

A system of cameras and sensors was prepared and attached to the driver's simulator. In particular, the visual, NIR, thermal, and physiological data has been captured during each recording. By utilizing a suite of recording modalities, we intended to account for the strengths and weaknesses of each recording method in their respective environments. This would allow, for example, the system to overcome the visual modality's weakness in low light with the NIR modality's superior low-light performance, or to overcome the thermal modality's weakness in an environment with fluctuating temperatures by utilizing the visual modality's respective robustness. By providing varied input modalities, we intend to maximize the system's potential ability to overcome a variety of challenging environments, including those found in real driving scenarios.

Our modalities were recorded using the following equipment:

1) Logitech HD web camera recording at 30 fps with audio and providing a top-down oblique view of the subject
2) Raspberry-Pi camera recording at 25 fps and providing a face closeup view
3) RGB camera from Imaging Development Systems (IDS) recording at 20 fps and providing a face closeup view
4) NIR camera from IDS providing a face closeup view, recording at 20fps.
5) FLIR ONE thermal camera, recording the subject's face at a slight angle at 7 fps
6) FLIR SC6700 thermal camera, capturing the subject's face at 100 fps
7) Thought Technology Ltd.'s four physiological sensors; three of them were attached to the subject's non-dominant hand and one to the torso, capturing at 2048 Hz: a) *B*VP, b) *Skin Temperature*, c) *Skin Conductance* and d) *Respiration*. To reduce noise in the physiological data, we asked subjects to drive mostly with their dominant hand and only use their other hand when necessary.

Fig. 1 illustrates the experimental setup environment with each channel labeled accordingly.

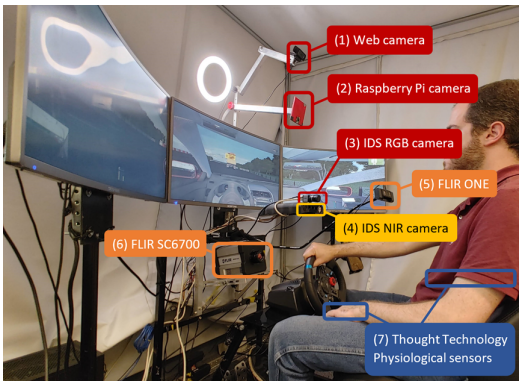


Fig. 1: Experimental setup showing sensor layout

After processing the data from each modality, we obtained a total of 367 recordings per modality, with 182 of the recordings belonging to the alert class and 185 of the recordings belonging to the drowsy class. The imbalance in the recordings between the classes occurred due to missing recordings for three subjects in the Morning session in the thermal modality.

Finally, of note is that the institutional review board of the University of Michigan has reviewed and approved all research procedures using the identification code HUM00132603.

IV. Modality Processing

A. Thermal Modality

For the thermal modality, we formed a thermal map by extracting the following features from the thermal frames: mean, maximum, minimum, range, and a histogram over the values of the pixels that represented the temperature distribution in the Regions of Interest (ROI).

To form the thermal map, three main processing steps have been followed. The first step consists of segmentation of the frame into five different regions, including the entire face, forehead, eyes, cheeks, and nose. Afterwards, those regions were tracked throughout the recording using the thermal-based tracking algorithm we proposed in [50]. A variation of the Shi-Tomasi corner detection algorithm [51] was used to detect points of interest in the Regions of Interest (ROIs) by computing the weighted square difference between two successive frames. The algorithm predicts the relocation of points of interest from one frame to another, with a small displacement between the pixels, which is very suitable for our tracking needs. A geometric transformation has been applied to map the points of interest between frames following the tracking process and the displacement estimation. This above geometric transformation estimated the transformation of interesting points on the basis of similarity.

Finally, we generated a thermal map that illustrates the thermal features in the ROIs, in order to extract potentially indicative thermal features of drowsy behavior in our five areas of interest. To that end, the following steps were taken in this order: a) ROI segmentation, b) segment binarization, c) image masking and d) thermal map cropping for each ROI. This method can be seen in Fig. 2.

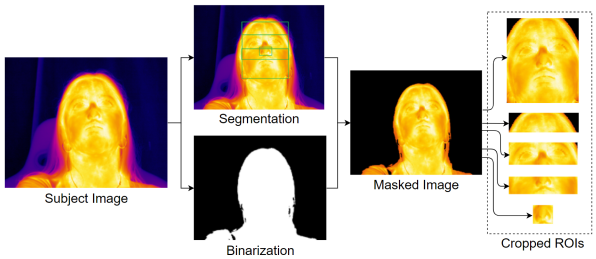


Fig. 2: The process of segmenting, binarizing, masking and cropping the thermal faces

Ultimately, we created the final thermal feature vectors from the generated thermal maps for all ROIs. The feature vectors were created by extracting the mean pixel values within the ROI, the minimum temperature, the maximum pixel value representing the highest temperature, the difference between the maximum and the minimum temperatures, and the histogram above the pixel values in the ROI that represents the temperature distribution in that region to form a total of 24 features per feature vector. After all features were extracted, they were then averaged across the whole recording to create a single feature vector per subject and then normalized using the afternoon baseline recording. The goal here is to explore and compare the behavior of the contact versus noncontact modalities in classification. Accordingly, each recording is represented as a single instance of being alert or drowsy in order to observe the performance of the different modalities and decide on the feasibility of using a fully noncontact approach, in addition to allowing us to exclusively explore the behavior of the modalities in the classification comparison. It should be noted also that the metrics varied no more than 1-2% depending on whether the morning or afternoon baseline was used for normalization.

B. Visual Modality

In processing the data from our visual modality, we utilized the OpenFace library to extract features from our RGB and NIR cameras. These features include facial landmarks, eye gaze, head pose, and a series of detected Action Units (AUs), which represent facial deformations due to facial muscle movement as defined by the Facial Action Coding System (FACS) [52]. To extract these features, we deployed a Convolutional Experts - Constrained Local Model (CE-CLM) within OpenFace, as described by Zadeh et al. [53].

After extracting landmarks in the region associated with the eyes, the system estimated the direction of the gaze by comparing the location of the center of the pupil and the eyeball sphere, then defining a ray intersecting them. This produced a feature vector describing the position of each pupil. Additionally, utilizing the facial landmarks of the face at every frame of the video, the system estimated the 3D positioning of the head, producing a feature vector describing the pose of the head. Finally, the system extracted facial appearance features by finding the difference between the expression found in the current frame and a neutral expression. These features were then summarized through the use of AUs, which represented a higher-level abstraction of individual features into common facial expressions; in addition, these AUs were recorded with the detected intensity. Examples of such AUs are shown in Fig. 3.

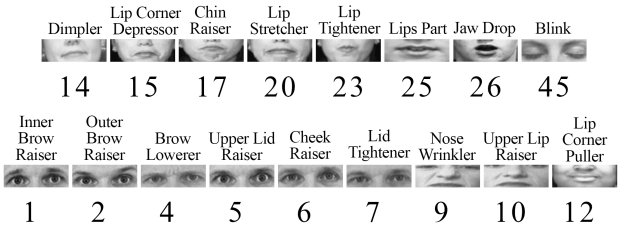


Fig. 3: AU definitions and their numbering as defined by FACS [54], [52]

After extraction of all 709 features across every frame of a given recording with OpenFace, the features were averaged across all frames, collapsing the time dimension per recording. These results were then normalized by dividing against the corresponding subjects' baseline recording.

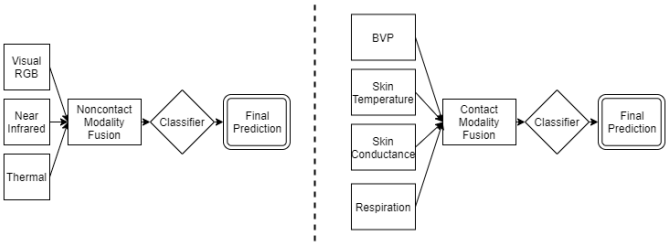


Fig. 4: Representation of Fusion for Noncontact (left) and Contact (right) based Modalities

C. Physiological Modality

Four primary sensors were used to collect physiological data; the BVP sensor at a 2048 Hz sample rate, and the Skin Conductance, Skin Temperature and Respiration sensors at a 256 Hz sample rate. The latter three sensor data feeds were then upscaled to 2048Hz to match the sample rate of the BVP sensor. Statistical features were then computed across the four modalities for the time and frequency domains. For BVP, mean, minimum, maximum, and standard deviation were extracted to describe overall behavior, as well as obtain information on the consecutive inter-beat interval. Further normal to normal heart beat related features provided information on the interval between two normal heartbeats. Finally, in order to describe spectral power statistics for the BVP channel, power related statistics for three frequency bands were also calculated. This resulted in a total of 49 features from the BVP channel. Standard time domain statistics were computed from the remaining three modalities to provide 24 more features, including mean, minimum, maximum, and standard deviation, among others.

The readings were then normalized using the afternoon baseline recording. This normalization resulted in a vector representing the variations of the subjects' features compared to their baseline for the alert and drowsy recordings. Finally, an average feature vector per recording was obtained by taking the mean value per feature across a given recording. Using each sensor feed separately provided four different types of contact-based signals. The extracted features of all four sensors were also combined to form a fused contact-based physiological modality.

V. Experimental Discussion

A. Experimental Setup

Prior to experimentation, all the data was normalized using the afternoon baselines, as described for each modality in the previous section. The afternoon baseline was selected experimentally over the morning baseline due to its improved performance. For our experiments, we looked at the performance of contact-based modalities versus noncontact-based ones. The target label was the alertness level of the driver, being modeled as a binary label of being either Alert or Drowsy. Two classifiers were used for training, namely, Random Forest Classifier (RFC), and Extreme Gradient Boosted Classifier (XGB) due to their improved performance in literature [55], [56]. Model performance was evaluated based on the metrics of accuracy, F1 score, specificity (alertness recall) and sensitivity (drowsiness recall), using Leave-One-Subject-Out cross validation. This form of Cross Validation means that each fold used during testing contained all recordings belonging to a single subject at a time, so that at no point during training, a classifier would have prior knowledge about any of the test subject data.

Following the evaluation of each separate channel, we explored the performance of noncontact versus contact-based modalities. In particular, we fused the features extracted from the visual RGB, NIR, and thermal data to form the noncontact integrated modality, and the features from all four physiological sensors to form the integrated contact-based modality as represented in Fig. 4. This fusion can be described as early fusion, where we concatenated features across all modalities before using the merged feature set for training a classifier. This fusion is beneficial as each unique recording is represented by a larger feature vector that encompasses information from multiple modalities, which provides richer information for the classifiers during the training phase. This process resulted in a total of 1538 features for the fused noncontact modalities, and 73 features for the fused contact-based modalities.

B. Contact-based Modality Results

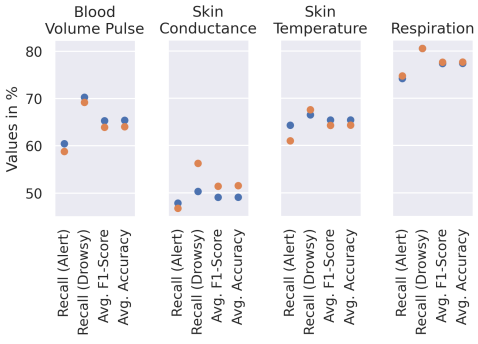
First, we present the performance of contact-based modalities, as shown in Fig. 5a. It can be observed that the respiration rate features provide the best signal for alertness detection, with an accuracy of 77.5% and a drowsy class recall of 80% using the Gradient Boosted classifier. Skin temperature and BVP have performance metrics in the mid-60s, with skin conductance being the worst indicator for this target label. Overall, the performance of the respiration signal indicates a superior and more reliable performance in modeling driver's alertness amongst the contact-based modalities.

C. Noncontact-based Modality Results

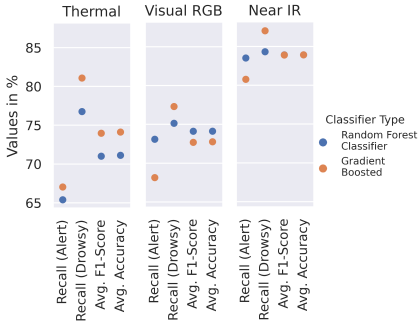
Next, we evaluated the performance of noncontact-based modalities, as shown in Fig. 5b. We observed improved performance compared to the contact-based modalities, with average accuracy for all three modalities above 71%. The best performing modality is the NIR modality, scoring 82% in average accuracy and 87% in recall for the drowsy class when using the Gradient Boosted classifier. However, the Random Forest classifier is preferred for this modality as it has the same accuracy as that of Gradient Boosting, but does not suffer from the recall tradeoff, having a more uniform recall of 83% for both classes instead.

D. Contact versus Noncontact-based Results

As shown in Fig. 6a, it is clear that the noncontact-based modalities outperform contact-based modalities in all metrics measured. One possible theory as to why noncontact-based modalities outperform contact-based ones is that visual signals

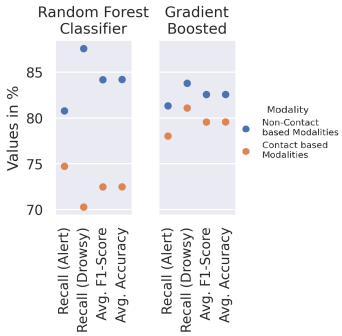


(a) Performance of Contact-based Modalities

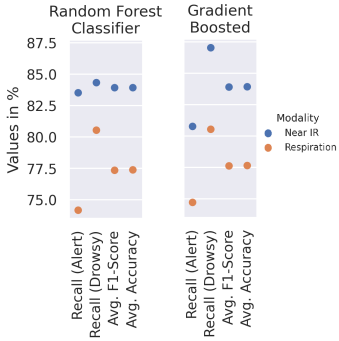


(b) Performance of Noncontact-based Modalities

Fig. 5: Performance of Contact-based versus Noncontact-based methods



(a) Comparison of Contact versus Noncontact-based Modality Performance using Early Fusion



(b) Comparison of the best Contact (Respiration) versus Noncontact (NIR) based Modality Performance

Fig. 6: Performance comparison of Contact-based versus Noncontact-based methods

are less susceptible to noise when factoring driver's movements and their effect on contact-based signals. If this theory proves to be true, it would further reinforce the practicality of noncontact-based methods within an automotive environment.

However, of further interest is the effect of fusion on the modalities' performance. Fig. 6b shows the comparative results of the best performing modality from the contact-based and noncontact-based modalities, being respiration and NIR respectively. When using the Random Forest classifier, we observe that the fusion of contact-based modalities performs worse than the single modality of respiration alone, implying that fusion here is detrimental to classification. But on the contrary, for noncontact-based modalities, we see that the fusion outperforms NIR slightly by about 1% when using the Random Forest classifier. This implies that fusion is beneficial to the noncontact-based modalities by allowing the possibility for multiple modalities to work well in tandem as a combined feature set without degrading performance. This would imply that different non-contact modalities could address each other's weaknesses when needed while ensuring the best possible performance at any time.

VI. Conclusion

In this paper we presented a multimodal dataset that evaluated a variety of contact and noncontact-based modalities in order to model driver's drowsiness. We see that noncontact-based modalities performed better than contact-based ones, highlighting the viability of using NIR and visual cameras as a noninvasive way of monitoring a person's alertness state. On the other hand, the respiration rate signal provided a significantly improved performance compared to other contact-based signals. However, its performance was not superior to that of the combined noncontact-based modalities or the the NIR modality by itself. The benefits of using such noncontact-based modalities include cheaper and easier to install hardware in vehicles, as well increased driver's convenience and acceptance of noninvasive monitoring methods. Furthermore, we showed the potential of fusion when using noncontact-based modalities, using more than one modality to monitor the subject's drowsiness state, which provides a more reliable and supported decision potentially resulting in developing better warning systems in vehicles if driver's drowsiness is detected. In future work, we are planning to explore the temporal dimension of the dataset using segmentation as well.

Acknowledgment

This paper is based on research funded by the Toyota Research Institute ("TRI"). Any observations, assumptions, or recommendations are the authors' and do not necessarily represent the views of TRI or any other Toyota entity.

References

[1] C. for Disease Control and Prevention. (2020) Cost of injury data. [Online]. Available: https://www.cdc.gov/injury/wisqars/cost/index.html

[2] W. H. Organization. (2020) Road traffic injuries. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

[3] C. for Disease Control and Prevention. (2020) Cost data and prevention policies. [Online]. Available: https://www.cdc.gov/transportationsafety/costs/index.html

[4] S. Chen, M. Kuhn, K. Prettner, and D. E. Bloom, "The global macroeconomic burden of road injuries: estimates and projections for 166 countries," *The Lancet Planetary Health*, vol. 3, no. 9, pp. e390–e398, 2019.

[5] O. U. P. USA. (2018) Sleep deprived people more likely to have car crashes. [Online]. Available: https://www.sciencedaily.com/releases/2018/09/180918082041.htm

[6] J. M. Owens, T. A. Dingus, F. Guo, Y. Fang, M. Perez, J. McClafferty, and B. Tefft. (2018) Prevalence of drowsy driving crashes: Estimates from a large-scale naturalistic driving study. [Online]. Available: https://aaafoundation.org/prevalence-drowsy-driving-crashes-estimates-large-scale-naturalistic-driving-study/

[7] "100-car naturalistic study fact sheet," https://www.csg.org/sslfiles/dockets/2011cycle/31B/31Bbills/100-Car_Fact-Sheet.pdf, accessed: 2021-05-12.

[8] "Nhtsa drowsy driving research and program plan," https://www.nhtsa.gov/sites/nhtsa.gov/files/drowsydriving_strategicplan_030316.pdf, accessed: 2021-04-28.

[9] A. M. Williamson and A.-M. Feyer, "Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication," *Occupational and Environmental Medicine*, vol. 57, no. 10, pp. 649–655, 2000. [Online]. Available: https://oem.bmj.com/content/57/10/649

[10] D. Dawson and K. Reid, "Fatigue, alcohol and performance impairment," *Nature*, vol. 388, no. 6639, pp. 235–235, 1997. [Online]. Available: https://app.dimensions.ai/details/publication/pub.1019789243andhttps://www.nature.com/articles/40775.pdf

[11] J. T. Arnedt, G. Wilde, P. Munt, and A. Maclean, "How do prolonged wakefulness and alcohol compare in the decrements they produce on a simulated driving task?" *Accident; analysis and prevention*, vol. 33, pp. 337–44, 06 2001.

[12] K. A. Brookhuis and D. De Waard, "Monitoring drivers' mental workload in driving simulators using physiological measures," *Accident Analysis & Prevention*, vol. 42, no. 3, pp. 898–903, 2010.

[13] B. Reimer and B. Mehler, "The impact of cognitive workload on physiological arousal in young adult drivers: a field study and simulation validation," *Ergonomics*, vol. 54, no. 10, pp. 932–942, 2011.

[14] M. Awais, N. Badruddin, and M. Drieberg, "A hybrid approach to detect driver drowsiness utilizing physiological signals to improve system performance and wearability," *Sensors*, vol. 17, no. 9, p. 1991, 2017.

[15] A. Persson, H. Jonasson, I. Fredriksson, U. Wiklund, and C. Ahlström, "Heart rate variability for driver sleepiness classification in real road driving conditions," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 6537–6540.

[16] M. Papakostas, K. Das, M. Abouelenien, R. Mihalcea, and M. Burzo, "Distracted and drowsy driving modeling using deep physiological representations and multitask learning," *Applied Sciences*, vol. 11, no. 1, p. 88, 2021.

[17] R. A. Naqvi, M. Arsalan, G. Batchuluun, H. S. Yoon, and K. R. Park, "Deep learning-based gaze detection system for automobile drivers using a nir camera sensor," *Sensors*, vol. 18, no. 2, p. 456, 2018.

[18] A. Raorane, H. Rami, and P. Kanani, "Driver alertness system using deep learning, mq3 and computer vision," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE, 2020, pp. 406–411.

[19] M. B. Lopez, C. R. del Blanco, and N. Garcia, "Detecting exercise-induced fatigue using thermal imaging and deep learning," in *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*. IEEE, 2017, pp. 1–6.

[20] S. E. H. Kiashari, A. Nahvi, H. Bakhoda, A. Homayounfard, and M. Tashakori, "Evaluation of driver drowsiness using respiration analysis by thermal imaging on a driving simulator," *Multimedia Tools and Applications*, pp. 1–23, 2020.

[21] I. Pavlidis, J. Levine, and P. Baukol, "Thermal imaging for anxiety detection," in *Proceedings IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications (Cat. No. PR00640)*. IEEE, 2000, pp. 104–109.

[22] F.-C. Lin, L.-W. Ko, C.-H. Chuang, T.-P. Su, and C.-T. Lin, "Generalized eeg-based drowsiness prediction system by using a self-organizing neural fuzzy system," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, no. 9, pp. 2044–2055, 2012.

[23] Q. Massoz, T. Langohr, C. François, and J. G. Verly, "The ulg multimodality drowsiness database (called drozy) and examples of use," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2016, pp. 1–7.

[24] S. E. H. Kiashari, A. Nahvi, A. Homayounfard, and H. Bakhoda, "Monitoring the variation in driver respiration rate from wakefulness to drowsiness: A non-intrusive method for drowsiness detection using thermal imaging," *Journal of Sleep Sciences*, vol. 3, no. 1-2, pp. 1–9, 2018.

[25] C. Puri, L. Olson, I. Pavlidis, J. Levine, and J. Starren, "Stresscam: non-contact measurement of users' emotional states through thermal imaging," in *CHI'05 extended abstracts on Human factors in computing systems*, 2005, pp. 1725–1728.

[26] M. Patel, S. K. Lal, D. Kavanagh, and P. Rossiter, "Applying neural network analysis on heart rate variability data to assess driver fatigue," *Expert systems with Applications*, vol. 38, no. 6, pp. 7235–7242, 2011.

[27] A. Sahayadhas, K. Sundaraj, and M. Murugappan, "Drowsiness detection during different times of day using multiple features," *Australasian physical & engineering sciences in medicine*, vol. 36, no. 2, pp. 243–250, 2013.

[28] A. Kolli, A. Fasih, F. Al Machot, and K. Kyamakya, "Non-intrusive car driver's emotion recognition using thermal camera," in *Proceedings of the Joint INDS'11 & ISTET'11*. IEEE, 2011, pp. 1–5.

[29] M. Rimini-Doering, D. Manstetten, T. Altmueller, U. Ladstaetter, and M. Mahler, "Monitoring driver drowsiness and stress in a driving simulator," in *First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*. Citeseer, 2001, pp. 58–63.

[30] C. Craye, A. Rashwan, M. S. Kamel, and F. Karray, "A multi-modal driver fatigue and distraction assessment system," *International Journal of Intelligent Transportation Systems Research*, vol. 14, no. 3, pp. 173–194, 2016.

[31] Y. Du, C. Raman, A. W. Black, L.-P. Morency, and M. Eskenazi, "Multimodal polynomial fusion for detecting driver distraction," *arXiv preprint arXiv:1810.10565*, 2018.

[32] P. W. Kithil, R. D. Jones, and J. McCuish, *Driver alertness detection research using capacitive sensor array*. Society of Automotive Engineers, 2001.

[33] C. Fors, C. Ahlstrom, and A. Anund, "A comparison of driver sleepiness in the simulator and on the real road," *Journal of Transportation Safety & Security*, vol. 10, no. 1-2, pp. 72–87, 2018.

[34] C.-H. Weng, Y.-H. Lai, and S.-H. Lai, "Driver drowsiness detection via a hierarchical temporal deep belief network," in *Asian Conference on Computer Vision*. Springer, 2016, pp. 117–133.

[35] T. Kundinger, N. Sofra, and A. Riener, "Assessment of the potential of wrist-worn wearable sensors for driver drowsiness detection," *Sensors*, vol. 20, no. 4, p. 1029, 2020.

[36] M. B. Kurt, N. Sezgin, M. Akin, G. Kirbas, and M. Bayram, "The ann-based computing of drowsy level," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2534–2542, 2009.

[37] A. Quddus, A. S. Zandi, L. Prest, and F. J. Comeau, "Using long short term memory and convolutional neural networks for driver drowsiness detection," *Accident Analysis & Prevention*, vol. 156, p. 106107, 2021.

[38] V. Saini and R. Saini, "Driver drowsiness detection system and techniques: a review," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 3, pp. 4245–4249, 2014.

[39] C. Anitha, "Detection and analysis of drowsiness in human beings using multimodal signals," in *Digital Business*. Springer, 2019, pp. 157–174.

[40] Q. Abbas and A. Alsheddy, "A methodological review on prediction of multi-stage hypovigilance detection systems using multimodal features," *IEEE Access*, vol. 9, pp. 47 530–47 564, 2021.

[41] A. Sengupta, A. Dasgupta, A. Chaudhuri, A. George, A. Routray, and R. Guha, "A multimodal system for assessing alertness levels due to cognitive loading," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 7, pp. 1037–1046, 2017.

[42] K. Riani, M. Papakostas, H. Kokash, M. Abouelenien, M. Burzo, and R. Mihalcea, "Towards detecting levels of alertness in drivers using multiple modalities," in *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, 2020, pp. 1–9.

[43] H.-S. Shin, S.-J. Jung, J.-J. Kim, and W.-Y. Chung, "Real time car driver's condition monitoring system," in *SENSORS, 2010 IEEE*. IEEE, 2010, pp. 951–954.
[44] A. Chowdhury, R. Shankaran, M. Kavakli, and M. M. Haque, "Sensor applications and physiological features in drivers' drowsiness detection: A review," *IEEE Sensors Journal*, vol. 18, no. 8, pp. 3055–3067, 2018.
[45] T. D'Orazio, M. Leo, C. Guaragnella, and A. Distante, "A visual approach for driver inattention detection," *Pattern recognition*, vol. 40, no. 8, pp. 2341–2355, 2007.
[46] R. O. Mbouna, S. G. Kong, and M.-G. Chun, "Visual analysis of eye state and head pose for driver alertness monitoring," *IEEE transactions on intelligent transportation systems*, vol. 14, no. 3, pp. 1462–1469, 2013.
[47] "Car crashes by time of day and day of week," Mar 2022. [Online]. Available: https://injuryfacts.nsc.org/motor-vehicle/overview/crashes-by-time-of-day-and-day-of-week/
[48] L. Oliveira, J. S. Cardoso, A. Lourenço, and C. Ahlström, "Driver drowsiness detection: a comparison between intrusive and non-intrusive signal acquisition methods," in *2018 7th European Workshop on Visual Information Processing (EUVIP)*. IEEE, 2018, pp. 1–6.
[49] H. Lee, J. Lee, and M. Shin, "Using wearable ecg/ppg sensors for driver drowsiness detection based on distinguishable pattern of recurrence plots," *Electronics*, vol. 8, no. 2, p. 192, 2019.
[50] M. Abouelenien, V. Pérez-Rosas, R. Mihalcea, and M. Burzo, "Detecting deceptive behavior via integration of discriminative features from multiple modalities," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 5, pp. 1042–1055, 2017.
[51] J. Shi *et al.*, "Good features to track," in *1994 Proceedings of IEEE conference on computer vision and pattern recognition*. IEEE, 1994, pp. 593–600.
[52] P. Ekman and W. V. Friesen, *Manual for the facial action coding system*. Consulting Psychologists Press, 1978.
[53] A. Zadeh, Y. Chong Lim, T. Baltrusaitis, and L.-P. Morency, "Convolutional experts constrained local model for 3d facial landmark detection," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017, pp. 2519–2528.
[54] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*. IEEE, 2018, pp. 59–66.
[55] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
[56] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," *CoRR*, vol. abs/1603.02754, 2016. [Online]. Available: http://arxiv.org/abs/1603.02754