

Technologies for Safe & Efficient Transportation

THE NATIONAL USDOT UNIVERSITY
TRANSPORTATION CENTER FOR SAFETY

Carnegie Mellon University

UNIVERSITY of PENNSYLVANIA

Multimodal Detection of Driver Distraction

FINAL RESEARCH REPORT

Maxine Eskenazi, Louis-Phillippe Morency, Alan W Black, Yulun Du

Contract No. DTRT12GUTG11

Distracted driving has become a major cause of crashes and loss of life. While there is legislation prohibiting the use of cellphones while driving, people continue to use them. We reason that if we can create a system that automatically detects when a person is distracted and warns them (even shutting down an application if necessary), then some serious accidents could be prevented.

The goal of this project was to expand on our previous work on automatically detecting driver distraction by using more modalities than simply speech. For this we needed to collect a new dataset where the car information (gas pedal, etc) and video driver information from a good quality back-facing camera were collected. We collected data from 30 subjects driving the OpenDS simulator we had previously created. In the past, we had used third party observation to annotate the places where the driver was believed to be distracted, for this dataset. For this dataset, we asked each subject to watch the video of their session and to stop the recording at places where they felt they “noticed something else” or that “they may have been less attentive”. They were asked to mark it (a modification of the NASA Task Load Index) and this annotation was logged and synced with the rest of the data. The route conditions were modified slightly – to the hairpin turns and a stop sign in the earlier database we added different placements of the turns and a stop light as well as street signs that could be seen from a distance. We also added small packages along the roadway to represent minor visual distractions that should not have an effect on the safety of the driver but could distract their attention.

Figure 1 shows an example of the data: video of the driver, the route they saw and the three types of modalities we detected (facial information, speech, car information).

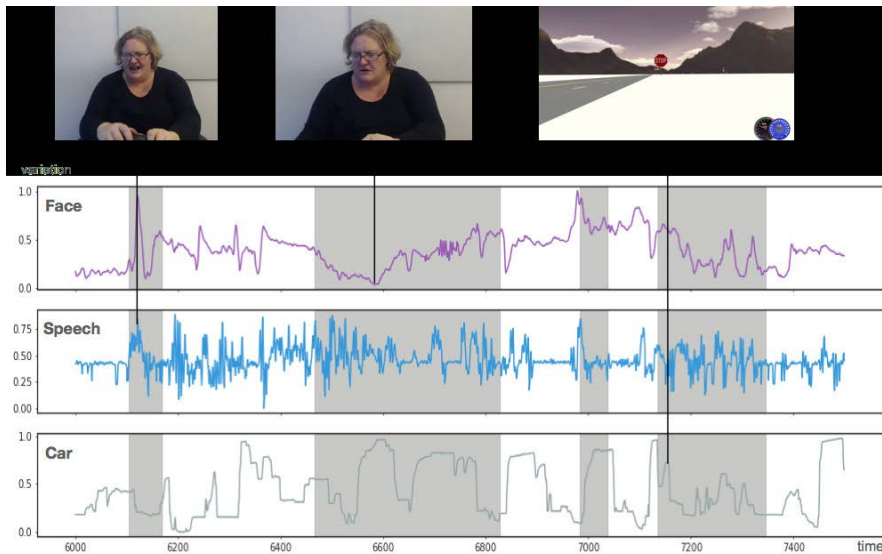


Figure 1: A case-in-point visualization of distraction for three modalities: Face, Speech, and Car. Each feature modality is reduced and normalized to a one-dimensional space and projected onto a continuous time axis. The grey area denotes the time period when the driver said she was distracted time period when the driver said she was distracted.

In this project we used multimodal polynomial fusion (MFP) over the three modalities to

detect distraction. The description of the algorithms used can be found in the Interspeech 2018 paper, Multimodal Polynomial Fusion for Detecting Driver Distraction. The facial features of facial landmarks, head pose turns, glances, eye gaze and facial action units were used. We posited that with this additional information we would obtain more accurate distraction detection than by simply using the speech information alone as we had done in the past.

Results show that MPF performs better than baseline models (all using three modalities), where the MPF model achieves an AUC of 0.7152, an EER of 0.3416, and an F-1 score of 0.5641 on the test set, while the best baseline NN-Cube achieves 0.7048, 0.3488, and 0.5453 respectively. We found that using more modalities achieves the best false positive/false negative tradeoff and the best overall detection.

Besides the successful detection results, we have run tests to determine how *early* MFP could detect distraction. While much of the literature in distraction detection shows that at the time the driver is distracted some algorithms can detect the distraction, this is not practical for use in a real driving situation. If we only detect when the driver is in a dangerous situation and is distracted, it is too late to avoid the consequences. It has been shown that a driver would need to be warned about 8 seconds before the dangerous situation. In preliminary experiments with MFP, we have found that we can reliably *predict* a dangerous distraction situation at 6 seconds previous to the situation and fairly reliably predict it at 8 seconds before the situation. This demonstrates the powerful predictive nature of this machine learning approach. We believe that with more training data, which would help eliminate more individual differences, we could have even better results at 8 seconds before the situation.

Milestones

- Created a new multimodal database with 30 drivers
- Gathered multiple modalities of data
- Implemented multimodal polynomial fusion for distraction detection
- Published results

Products

- Database of 30 drivers with multimodal information: Number of drivers: 30
- Average driving time: 15:13 minutes per subject. The database is publicly available on github: <https://github.com/DialRC/DistractionDetectionDataset>
- Distraction detection algorithms (which are also available on github)
- Publication: Multimodal Polynomial Fusion for Detecting Driver Distraction, Yulun Du, Alan W Black, Louis-Philippe Morency, Maxine Eskenazi, 2018, Proceedings Interspeech 2018.

References

OpenDS <https://www.openhub.net/p/OpenDS>

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.