# Cell Phone Usage Behavior Detection Based on Primitive Recognition for the IVBSS and SHRP2 Datasets

Summer 2020

MDP-DEVIATE, Research Subteam, Summer 2020: Jenil Shah, Shengan Zhang
Researchers: Carol Flannagan, Jared Karlow, Kathleen Klinich UMTRI Group: CMISST

## ABSTRACT

Detection of cell phone usage behaviour while driving is an important unsolved problem in the domain of transportation. Our goal differs from earlier attempts by developing a primitive-based machine learning algorithm for cell phone behaviour detection. Primitives in our context include facial landmarks, eye gaze, head orientation, cell phone, and hand position to identify behaviours such as texting and talking. The first layer of the algorithm identifies primitives in videos which are provided to the secondary algorithm for detection of cell phone behaviours. The IVBSS and SHRP2 datasets are adopted for training and testing of the algorithm. Currently, we are developing the first layer for primitive detection in videos. The research presents novel algorithms and methods adopted for the identification of each primitive, namely the OpenFace algorithm for face detection and HandTracking for hand detection at this stage. Once the ground truth is created for both datasets, the research further aims to test and improve accuracy of primitive detection and move on to the creation of the secondary algorithm.

## BACKGROUND

Though cell phone usage is discouraged while driving, methods to detect driver phone use in vehicles and provide disincentives for doing so is not currently possible due to a lack of effective detection algorithms. The NHTSA estimates that there were approximately 3,450 fatalities due to distraction-affected crashes in 2016; creation of systems to detect the cell phone usage behaviour and trigger warning can reduce these fatalities. Hence, we aim to develop a primitive-based machine learning algorithm for cell phone behaviour detection to address the problem.

## OBJECTIVES

1. Create or adopt algorithms to identify primitives such as facial landmarks, eye gaze, head orientation, cell phone, and hand position.
2. Validate the algorithms against ground truth and achieve high accuracy on IVBSS and SHRP2 datasets.
3. Develop a secondary machine learning algorithm for cell phone behaviour detection.

## METHODS

The IVBSS Dataset contains 2 perspective: face and cabin. As these are separate videos, we aim to individually identify different primitives from each. From the face perspective we aim to identify facial landmarks, head position and orientation, and gaze. From the cabin perspective we aim to identify hand position, steering wheel and cell phone.

## METHODS

The following diagrams outline the methods and algorithms that were used to determine primitives.

For face primitives, we have tried several object detection models including YOLO, AlphaPose, OpenPose, and finally decided to operate using the OpenFace Object Detection on the IVBSS Face View videos. Openface was chosen due to its high accuracy, robustness and utility.

The methods we have tried for hand detection include OpenPose and HandTracking, and we decided to use HandTracking. After our experiments, we found that OpenPose is not suitable because its detection requires the presence of a human's body part in the video, which is not often the case for SHRP2 Dataset.

**1. FaceView Primitive**

- OpenFace uses Conditional Local Neural Fields (CLNF) for detection of Facial Landmarks (Fig. 1)
  - Group detected facial landmarks into regions of eye, nose, brows & face outline which are going to be used as primitives.
    - Using facial landmarks, estimate the head pose (3D position of head with respect to camera)
      - Estimate head orientation (Rotation in terms of Yaw, Pitch, and Roll)
  - CLNF is a general landmark detector and hence it is also used for detection of 55 eye landmarks (Fig. 2)
    - Using eye landmarks, estimate the gaze primitives namely, gaze vector (3D Eye vectors) & gaze angle (angle of eye relative to straight head position & camera).

**2. Hand Position Primitive**

- HandTracking adopts Deep Neural Network (DNN) to detect the position of drivers' hands. With DNN
  - DNN is a robust model minimizing the effects of lighting condition, noisy environments, diverse viewpoints and so on
    - Retrain the DNN model adapted to IVBSS / SHRP2 Datasets to achieve a higher accuracy
      - Estimate the position of hand with a high accuracy
  - Background-removal algorithm based on Dense Optical Flow (motion tracking with frame difference)
    - Use the result of Optical Flow as a mask, together with the idea of continuous motion and pixels constituting the same object must be of similar situations
      - Estimate the position of hand with a high accuracy
      - This method of background removal can be applied to other general moving object detection algorithms

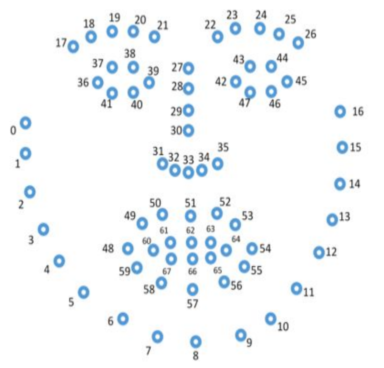

Fig 1. Facial Landmarks

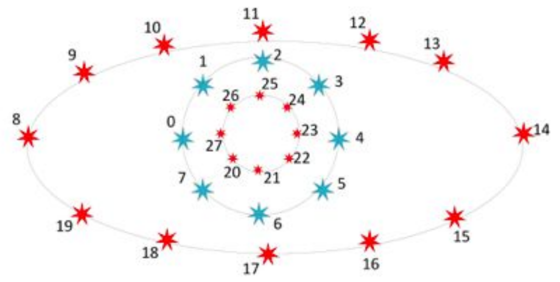

Fig 2. Eye Landmarks

## RESULTS & DISCUSSION

For hand detection, our experiments found that although the HandTracking algorithm works well on typical videos, its accuracy on SHRP2 dataset is not very high due to the large distinction between SHRP2 videos and others. The reason is that the model provided by the author is trained by EgoHands Dataset, where the hand pattern greatly differs from that found in SHRP2 in clarity and color. Though background-removal greatly reduces misdetection by removing the unimportant parts in the video, to radically solve this problem, we need will retrain the model based on SHRP2 Dataset.

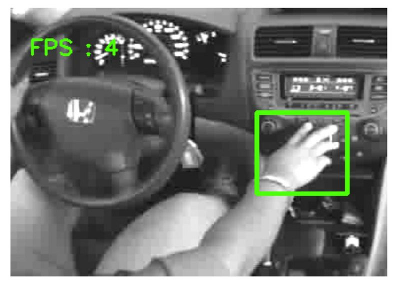

Fig 3. True Detection of Hand

Fig 4. False Detection of Hand

Fig 5. SHRP2 Frame after Background Removal

Even though the validation ground truth isn't available yet, the results from a normal eye perspective seem promising for detected primitives in the face view of IVBSS Dataset. The detection of facial landmarks is usually accurate, but the algorithm sometimes fails when the face is occluded. The eye gaze detection algorithm works well; the gaze vectors and gaze angle can be cumulatively used to track the movement of eyes such as left, right, up, and down. The calculated head orientation and head pose are relative to the camera, and can be improved by using the actual IVBSS camera parameters instead of default ones. All the above problems can be solved by retraining the OpenFace algorithm on the labeled IVBSS Dataset.

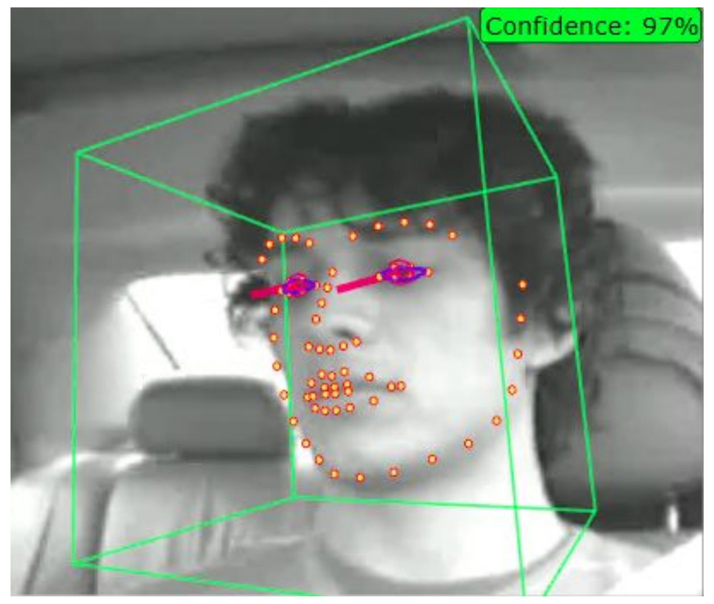


Figure 6. Face View Performance
The orange keypoint represent facial landmarks. The pink 3D vectors represent gaze vectors and the green 3D bounding box represnts head orientation & pose.

## NEXT STEPS

At this stage, we have found promising algorithms for primitive detection of both Face and Hand in Cabin view of the IVBSS Dataset. Once the labeled dataset is available, we aim to retrain the models and improve accuracy of primitive detection. Once achieved, we would test the accuracy of detection of each primitive against the validation dataset. Finally, after successfully detecting the primitives and validating the results, we aim to create the secondary cell phone detection algorithm.

### References

https://github.com/victordibia/handtracking

https://github.com/TadasBaltrusaitis/OpenFace

https://docs.opencv.org/3.4/d4/dee/tutorial_optical_flow.html

### Acknowledgements

Thanks to Dr. Carol Flannagan, Dr. Kathleen Klinich, Jared Karlow, Kevin Wang, Hanwen Miao for their guidance and help throughout the development of this project.