Report: Technologies for Autonomous Vehicles - Driver Monitoring Systems using AI

Basmo, Peder (Matr. 321299)
Skinnemoen, Erlend (Matr. 321135)

May 8, 2024

Abstract :

This report presents the development and implementation of a Driver Monitoring System utilizing Python and Google's MediaPipe FaceMesh to enhance road safety by monitoring driver attentiveness. The system calculates the Eye Aspect Ratio and the percentage of eye closure to detect drowsiness, while also tracking eye and head gaze positions to identify driver distraction. By leveraging these metrics, the system provides real-time alerts to drivers, prompting corrective actions to maintain focus on the road. This approach aims to enhance road safety.

1 Introduction

Driver Monitoring Systems (DMS) are essential for improving road safety by continuously assessing a driver's alertness. Using artificial intelligence and computer vision, it's possible to detect symptoms of drowsiness and distraction in 'real-time'. This report outlines the development and implementation of a DMS utilizing Python and Google's MediaPipe FaceMesh. The system calculates the Eye Aspect Ratio (EAR) and percentage of eye closure (PERCLOS) in order to assess the drivers drowsiness. Additionally, it combines eyes and head gaze positions to detect if the driver is distracted, and not looking at the road. By combining these metrics, the DMS provides a solution to enhance driver safety through monitoring and alerting mechanisms.

2 Background

Eye Aspect Ratio (EAR): EAR is a geometric measure used to determine the openness of the eyes, which helps in detecting blinks and estimating eye closure levels. The calculation is based on the vertical and horizontal distances between specific facial landmarks around the eyes. The script calculates EAR by measuring the distances between these landmarks, for each eye, six key points are used. Two points at the eye corners (one at the inner corner and one at the outer corner). Four points on the top and bottom eyelids. When the eye is fully open, the EAR value is higher because the distance between the upper and lower eyelids is maximum. When the eye closes, the EAR decreases significantly due to the reduction in the vertical distances. This relationship is shown in eq.(1).

$$EAR = \frac{bottom\_of\_eye - top\_of\_eye}{left\_eye\_corner - right\_eye\_corner} \tag{1}$$

PERCLOS Calculation: The script uses the EAR meaasurements to calculate PERCLOS, a metric that quantifies the percentage of time the eyelids cover a substantial part of the pupil over a defined time frame. One calculates it by measuring the EAR continuously, establishing a threshold for when the eyes are "closed," and calculating the percentage of time the EAR falls below this threshold during the observation period. This threshold helps in determining moments of significant eye closure, indicative of drowsiness. For the purpose of this system, PERCLOS is specifically defined as the proportion of time that the eyelids are at least 80% closed relative to their fully open state.

3 Methods

3.1 Facial Landmark Detection and data processing

The first half of the script utilizes the MediaPipe FaceMesh model configured to track facial landmarks. It captures video from the laptop's camera, processes each frame to detect landmarks, and calculates their 2D and 3D coordinates without saving the frames, which enhances performance. Utilizing the MediaPipe drawing utilities, the script configures the necessary specifications for a visualization on the real-time video which is presented .

As frames are processed, specific landmarks determine the positions of the eyes and pupils as well as other significant points, such as the nose. The script uses these landmarks to solve the Perspective-n-Point (PnP) problem, obtaining rotational vectors converted to Euler angles for understanding head orientation.

3.2 Drowsiness detection

The second half of the script calculates and evaluates conditions like excessive eye closure and deviation from a forward gaze, triggering alerts if the driver is perceived as either drowsy or distracted. When the systems starts, the script calibrates the EAR by instructing the driver to keep their eyes open and head forward, collecting EAR data over a three-second period to establish a baseline for the eyes open state. Subsequently, the script prompts the driver to close their eyes to calculate the baseline EAR for the eyes closed state, determining a threshold EAR that indicates significant eye closure. Post-calibration, the script continually checks if the EAR falls below the calculated threshold, indicating that the eyes are predominantly closed. If the EAR is consistently below this threshold, it is recorded as an instance of drowsiness. If this is the main behavior over a 10-second period (80% of the time) the driver is alerted with a "Drowsy" message on the display and an audio alert is triggered.

3.3 Distraction detection

The script calculates the pitch and yaw for both eyes from the facial landmarks and compares these with the baseline head orientation established during the calibration phase. The calibration captures the initial head vertical (pitch) and horizontal (yaw) angles, which represents the driver's forward-facing position. If the combined average deviation in pitch or yaw of the eyes, plus any change in head orientation from the baseline, the script classifies this as a distraction event[1]. Upon detecting such a deviation the script displays a "DISTRACTED" alert on the display. This dual approach, monitoring both eye and head movements, enables the system to identify moments when the driver's attention may not be focused on the road.

The calculation of eye gaze direction is based on the movement of the iris. The pitch and yaw movements of the iris are normalized by half the height and half the width of the eyes, respectively. This normalization is then scaled to the maximum angles that the human iris can move within the eye socket[2], which are circa 30 degrees for constrained vertical movement and 45 degrees for typical horizontal movements (Abdullah et al., 2022). For instance, if the vertical component of the gaze vector indicates an upward or downward movement, it is normalized by half the eye height and multiplied by 30 degrees, representing the maximum vertical movement of the iris. Similarly, horizontal movements are scaled by 45 degrees. This method allows for an accurate estimation of gaze direction by quantifying how much the iris deviates from the center of the eye in both dimensions, which is important for determining where a person is looking relative to their head position.

4 Results and discussion

The implementation of the eye gaze tracking in the DMS has proven to be effective in assessing attentiveness and detecting distractions for a driver. By calculating the pitch and yaw of each eye, the system is able to determine the direction of the driver's gaze relative to their head position.

To further enhance the precision and effectiveness of the DMS, several improvements can be implemented. By upgrading to a higher-quality camera the system can become more sensitive to small changes in the driver's facial expressions and eye movements. Additionally, tuning the parameters used in facial landmark detection could improve the system's responsiveness to various environmental conditions and individual facial features, thereby enhancing overall precision and reliability. These advancements would collectively contribute to a more robust and effective monitoring system.

The scripts initiates each session by establishing a baseline for EAR and head orientation. This calibration accounts for individual variations in eye morphology and typical gaze patterns, tailoring the system to recognize deviations from each driver's norm. This makes the system more robust against both Type I & II errors due to individual differences. It makes the program somewhat more complex, but is made up for in precision.

The selection of 30 and 45 degrees as the maximum constrained movements for the eye is based on the limitations of the human eye socket. These thresholds are necessary to accurately interpret eye movements relative to normal and alert states. However, this is a big simplification and can be improved by implementing a more sophisticated model for describing constrained eye movements. Hence, with the current system, tuning the maximum constrained movement values is important, as setting them appropriately ensures that the system neither underestimates nor overestimates the range of normal eye movements, which could lead to false alarms or missed detections.

5 Conclusion

The implementation of the Driver Monitoring System has demonstrated the possibility to detect driver drowsiness and distraction through AI using facial landmark detection. The system's ability to calculate EAR and PERCLOS provides a way to measure drowsiness, while the assessment of eye and head orientation is able to identify distractions. The initial calibration for individual drivers helps tailor the system to the specific driver's characteristics, improving the systems ability to adapt. To further improve the system, enhancements such as using a higher-quality camera and refining detection algorithms will help on the system's sensitivity and precision. Overall, the DMS contributes to a monitoring system that contributes to road safety by ensuring drivers remain attentive and alert.

References

Abdullah, Haseena, Pavan Kumar Verkicharla, and Shonraj Ballae Ganeshrao (2022). "Extent of foveal fixation with eye rotation in emmetropes and myopes". In: *Journal of Optometry* 15.4, pp. 293–298. ISSN: 1888-4296. DOI: `https://doi.org/10.1016/j.optom.2021.12.001`. URL: `https://www.sciencedirect.com/science/article/pii/S1888429621000881`.

[1] Roll motion is not considered as it's possible for the face to perform such a motion while keeping the eyes on the road

[2] These factors are tuneable