Driver Monitoring System: Enhancing In-Cabin Safety with AI-Powered Vision

Mohamed S. Abdallah, Huynh Thai Hoa, Chernozhukov Maksym, 박건기, Sadat Hossain, 은관주, Asim Usman, 김수훈

AI Lab, DeltaX Co., Ltd., 5F, 590 Gyeongin-ro, Guro-gu, Seoul 08213, Republic of Korea

sameer@deltax.ai , hoaht@deltax.ai , maksym.labs@deltax.ai, pgg113579@deltax.ai , sadat@deltax.ai , nnn1520@deltax.ai , usman@deltax.ai , stephenkim@deltax.ai

ABSTRACT

This paper introduces a robust Driver Monitoring System (DMS) that leverages cutting-edge AI and computer vision technologies to address critical in-cabin safety challenges. The system is designed to monitor driver behavior, detect drowsiness, assess distraction levels, and ensure compliance with safety measures such as seatbelt usage and hands-on steering wheel policies. Utilizing a single in-cabin camera, the DMS captures high-resolution images and processes them in real-time using advanced models like YOLOX, 3D landmarks, and eye gaze estimation algorithms. Key features include the detection of driver drowsiness through facial landmark analysis, behavior classification for actions like phone usage and smoking, and gaze-based distraction detection. Robust methodologies and comprehensive datasets ensure system reliability across diverse lighting conditions, occlusions, and edge cases. Evaluation results demonstrate high accuracy rates.

1. Introduction

This paper highlights the system's architecture, methodologies, and performance benchmarks, paving the way for safer automotive environments. The automotive industry is increasingly focusing on enhancing driver safety by developing advanced Driver Monitoring Systems (DMS).

Distracted and drowsy driving significantly contributes to road accidents worldwide. Addressing these challenges requires a system capable of real-time monitoring and alerting, using cutting-edge AI and computer vision technologies. This paper introduces a DMS that employs a single camera setup to detect driver states and behaviors, contributing to safer driving environments.

Driver Monitoring Systems (DMS) have garnered significant attention due to the growing demand for enhanced in-cabin safety and autonomous vehicle technologies. Previous studies have explored various aspects of driver monitoring, including drowsiness detection, distraction estimation, and behavior analysis.

Drowsiness detection has been a focal area, with approaches ranging from analyzing eyelid closure and yawning to monitoring brain activity using electroencephalography (EEG) to a computer vision system to monitor eyelid movements for fatigue detection.

Early distraction detection systems relied on head pose estimation using 2D landmarks. Recent work emphasizes the use of gaze estimation models, which leverage deep learning to predict driver attention in dynamic environments.

Seatbelt detection systems have largely employed traditional image-processing techniques combined with machine learning classifiers (Shivakumar et al., 2018). However, advancements in deep learning, such as YOLOX, have enhanced the ability to detect seatbelt states and hand positions on the steering wheel with higher precision.

2. System Architecture and Methodology

The proposed DMS (Fig 1) integrates both hardware and software components to enable real-time monitoring of driver behavior and safety features. The system's architecture includes an in-cabin camera placed near the rear-view mirror, connected to an edge device like a TI board for processing. The architecture supports live data analysis, ensuring flexibility and robustness across diverse operational conditions.

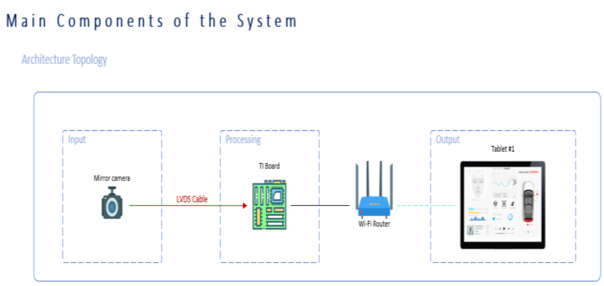

Fig 1. System Architecture

Our system aims to provide better safety for the automotive industry by using AI Computer Vision technology and a single camera to sense what happens in the cabin and trigger actions accordingly. The next generation of vehicles should be equipped with this technology.

The features covered by our system fall into the Driver Monitoring System category which is:

1. Drowsiness Detection.
2. Behavior Detection
3. Distraction Detection
4. Seat Belt Detection
5. Steering Wheel Interaction

2.1 Drowsiness Detection

The Drowsiness Detection is a feature that detects if the driver is in the capacity of driving or gets sleepy while driving.

The drowsiness detection feature uses facial landmark models to assess the driver's alertness. It tracks eye openness and calculates thresholds for categorizing the driver as awake, drowsy, or asleep. This involves real-time analysis of facial landmarks and time-series data.

The drowsiness detection feature processes eye landmark data to determine the state of the driver's eyes as either open or closed based on the distance between the top and bottom eye landmarks relative to a defined threshold. Time-series analysis is then applied to classify the driver's state into one of three categories: awake, drowsy, or asleep. Specifically, the system identifies the driver as awake if their eyes remain open continuously for at least 1 second, drowsy if their eyes remain closed for 1 second, and asleep if their eyes are closed continuously for 3 seconds. This approach ensures accurate and timely detection of drowsiness levels (Fig 2).

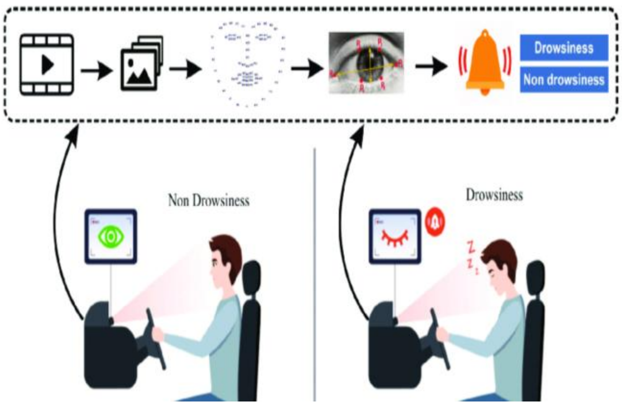

Fig 2. Facial Landmark

The performance (Metric) of Drowsiness Detection was evaluated using Normalized Mean Error (NME) and feature Accuracy. We achieved 1.7715 % and 97.733 % respectively.

2.2 Behavior Detection

Behavior detection identifies actions such as smoking, phone usage, drinking, and eating by analyzing the driver's hand regions. Using the YOLOX object detection model, the system detects relevant objects, including phones, cups, or food, within the predicted hand areas derived from body keypoints. This enables accurate classification of behaviors such as texting, scrolling, or making calls, ensuring comprehensive monitoring of driver activities. (Fig 2, 3).

Calling

Texting

Holding

Fig 3. Driver Behavior (Phone use)

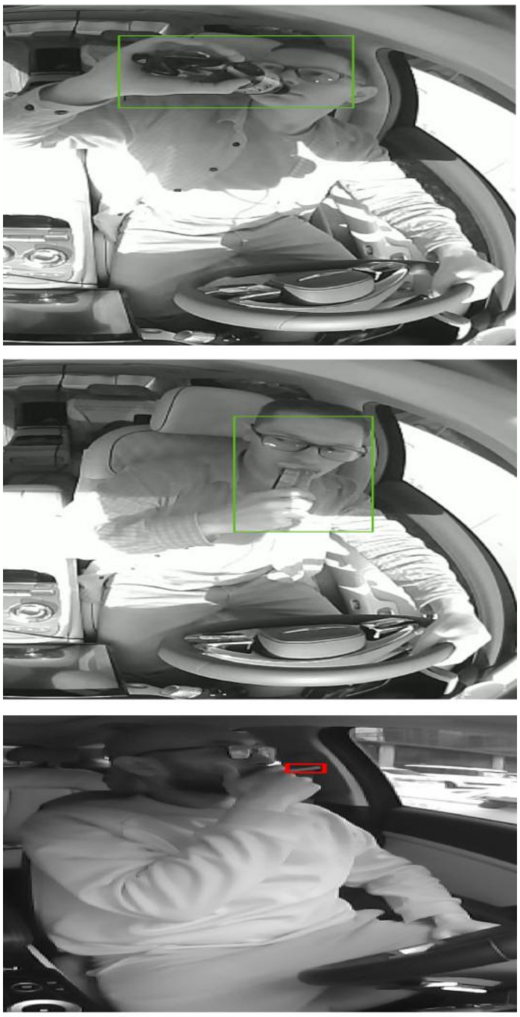

Fig 4. Driver Behavior (Drink-Eat-Cigarettes)

The performance (Metric) of Body Keypoint Model was evaluated using Average Precision (AP) and Average Recall (AR). We achieved 0.810, 0.837, 0.837, and 0.865 for AP, AP50, AP75, and AR respectively.

2.3 Distraction Detection

Distraction Detection or Forward Looking is a feature that detects if the driver is keeping his focus on the road or if he's being distracted, looks elsewhere, and creates a potential situation of danger around him.

Distraction detection involves gaze estimation and head pose analysis. An eye estimation model replaces traditional head pose thresholds, providing improved accuracy in identifying forward-looking focus.

For this feature, we initially employed a facial landmark model to estimate the driver's head pose, analyzing yaw, pitch, and roll angles and comparing them against predefined thresholds to determine if the driver was looking forward. To enhance accuracy, we later transitioned to an eye gaze estimation approach, using a model to predict the driver's gaze direction (pitch and yaw) and comparing these values against specified thresholds for forward-looking assessment (Fig 5, 6).

Fig 5. Facial Landmark

Fig 6. Eye Gaze Estimation

2.4 Seat Belt Detection

Seatbelt Detection is a feature that must detect the driver seat of a vehicle whenever the driver is sitting to see if he fastened their seatbelt or not.

The system relies on YOLOX and ROI settings for precise localization. If the driver is detected, we will determine the driver's seatbelt status based on the detection result of two classes: seat belt on, seat belt off (Fig 7).

The performance (Metric) of Seat Belt Detection was evaluated using Mean average precision (mAP). We achieved 95.7 and 93.2 for seatbelt_on, seatbelt_off respectively.

Fig 7. Seat belt detection

2.5 Steering Wheel Interaction

The Steering Wheel Manipulation is a feature that detects if the driver is keeping at least one hand on the steering wheel while driving. The Steering Wheel

Manipulation status is determined based on the occupancy status of the driver, If the driver is detected, we will classify hand positions as both hands on, left-hand on, right-hand on, or hands off. (Fig 8).

The performance (Metric) of Steering Wheel Interaction was evaluated using Mean average precision (mAP). We achieved 94.0, 93.4, 94.0 and 94.0 for hands_both_on, hands_off, hands_left, and hands_right respectively.

Holding with both hands

Holding with right hand

Holding with left hand

Fig 8. Holding Steering Wheel

3. Cameras and Hardware setup

Our solution is built with the camera positioned beneath the rearview mirror. The camera needs to be adjusted to ensure coverage of the entire interior of the car, capturing all seats and passengers.

We used different types of cameras like ELP and Leopard with different resolutions, chromaticity (RGB/IR), and FOV(H/V). Table 1 compares these cameras models regarding model name, Resolution, FPS, Chromaticity, and Field of View.

Table 1. Camera Specs

	ELP	Leopard
Model name	CMOS OV2710	LI-OX05B1S-VCSEL-OMS-9295-200H
Resolution	1920 x 1080 2MP	2592 x 1944 5MP
FPS	30	60
Chromaticity	RGB / NIR	RGB - IR
Field of View (FoV)	180˚ horizontal.	200˚ horizontal

Board details: The TDA4VH is an automotive System-on-Chip (SoC) from Texas Instruments' Jacinto™ 7 processor family, tailored to meet the sophisticated demands of Advanced Driver Assistance Systems (ADAS) and autonomous vehicle technologies. It features a variety of automotive interfaces, including multiple Camera Serial Interface (CSI) ports, Ethernet interfaces, and Controller Area Network (CAN) modules, enabling seamless integration with automotive sensors and communication networks.

For development and testing, Texas Instruments provides the J784S4XEVM evaluation module, compatible with the TDA4VH-Q1 and other Jacinto™ 7 processors. This evaluation board serves as a comprehensive platform for developers to design, test, and optimize applications, fully utilizing the capabilities of the TDA4VH-Q1 SoC.

To enhance dataset diversity, we utilized various vehicles, including the Genesis G90, Hyundai IONIQ, and KIA Sportage.

4. Datasets

Our solution is developed using our own datasets. The authors considered that the dataset should vary in illumination, skin color, clothes, and car brand to increase the model’s robustness.

The data collection process is designed to ensure diverse and comprehensive scenarios, capturing various conditions and behaviors. Data is gathered across different locations, including on the road, parking lots, and basements, under varied weather conditions such as sunny, cloudy, and nighttime settings.

To ensure diversity, the data incorporates people of different ages, genders, and body sizes, wearing clothes of various colors, types, textures, and fabrics. Additionally, it captures in-vehicle behaviors such as body movement, rotation, and bowing, creating a rich dataset for robust model training and evaluation. Over 100,000 images were collected in total, and they were annotated using various annotation tools.

The data collection process follows specific guidelines to ensure consistency and quality across scenarios. Drivers are required to perform the same action continuously with smooth, uninterrupted movements in various poses and locations. Equal durations of recorded videos are recommended for each scenario, maintaining uniform FPS and resolution. For static scenarios, additional lighting, such as lamps, is advised to vary illumination conditions. Objects, particularly cigarettes, must remain visible to the camera throughout the recordings. Furthermore, the positioning of items such as phones, cigarettes, bottles, cups, and snacks should be adjusted within the imaging space (e.g., moving up-down, left-right, or zooming in-out) to enhance dataset diversity.

5. Conclusion

This paper presents a comprehensive Driver Monitoring System (DMS) that leverages advanced AI and computer vision techniques to enhance in-cabin safety and reduce road accidents. By integrating features such as drowsiness detection, behavior analysis, distraction detection, seatbelt compliance, and steering wheel interaction monitoring, the system addresses critical safety concerns in real-time. The utilization of robust models like YOLOX and gaze estimation algorithms, coupled with a single in-cabin camera setup, ensures high accuracy and efficiency across diverse operational scenarios. Rigorous evaluations demonstrate the system's ability to handle challenges such as varying lighting conditions, occlusions, and edge cases.

Future work will focus on incorporating multi-camera systems, expanding datasets to improve generalization, and integrating additional safety features to further enhance system reliability and functionality.

While DMS primarily focuses on monitoring the driver’s behavior and state, In-Cabin Monitoring Systems (ICMS) extends its capabilities to include monitoring all occupants and their interactions within the vehicle cabin. This holistic approach enhances safety by addressing scenarios such as passenger seatbelt compliance and detecting potentially unsafe behaviors among passengers, such as distracting the driver. Furthermore, ICMS provides better situational awareness, paving the way for seamless integration with autonomous driving systems and advanced safety mechanisms.

The potential solution ICMS introduces a novel, unified framework for monitoring both the driver and passengers, leveraging state-of-the-art computer vision and AI techniques. Unlike traditional DMS, which focuses solely on the driver, ICMS offers an integrated solution for comprehensive cabin monitoring. This approach not only improves safety but also contributes to the development of intelligent in-cabin experiences, setting a new benchmark for modern automotive systems. By seamlessly combining multiple features in a single system, the ICMS represents a significant advancement in automotive safety and convenience.

Acknowledgement

A heartfelt thank you to all the talented researchers at DeltaX for their dedication and contributions!

References

[1] Dasgupta, A., George, A., & Routray, A. (2013). A Vision-based System for Monitoring Drowsiness and Distraction. Proceedings of the IEEE Conference on Intelligent Vehicles.
[2] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[3] Liu, W., Anguelov, D., Erhan, D., et al. (2016). SSD: Single Shot MultiBox Detector. Proceedings of the European Conference on Computer Vision (ECCV).
[4] Chen, X., Pan, J., Li, Y., & Huang, Y. (2017). Driver Behavior Recognition Based on Deep Learning. Proceedings of the International Conference on Neural Information Processing.
[5] Murphy-Chutorian, E., & Trivedi, M. M. (2009). Head Pose Estimation in Computer Vision: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence.
[6] Kellnhofer, P., Recasens, A., Matusik, W., et al. (2019). Gaze360: Physically Unconstrained Gaze Estimation in the Wild. Proceedings of the IEEE International Conference on Computer Vision (ICCV).
[7] Shivakumar, K., Balasubramaniam, S., & Raj, A. (2018). Automated Seatbelt Detection System for Enhanced Vehicle Safety. International Journal of Computer Applications.
[8] Fan, X., Wang, L., & Xiong, J. (2019). Eye Tracking for Vehicle Driver Monitoring: A New Dataset, Baselines, and Evaluation. IEEE Transactions on Intelligent Vehicles, 4(4), 539–549.
[9] Chen, H., Wang, L., & Xu, H. (2020). Multi-Task Learning for Driver Behavior Analysis via Convolutional Neural Networks. Proceedings of the IEEE International Conference on Multimedia Big Data (BigMM).
[10] https://www.elpcctv.com/
[11] https://leopardimaging.com/