Enhancing bus safety: A modular driver monitoring system

Viktor Nagy[1*], Patrik Bézi[1], Gábor Kovács[1,2]

[1]Széchenyi István University, Győr, Hungary
[2] Institute of the Information Society, Ludovika University of Public Service, Budapest, Hungary
[*]Corresponding author, email: nviktor@sze.hu

Received: 1 December 2023; accepted: 27 February 2024, published online: 7 May 2024

Summary

In the area of road safety, the development of a Modular, Machine Vision-Based, Custom-Built Driver Monitoring System (DMS) for bus drivers has become imperative. This research presents a comprehensive system capable of detecting drowsiness, blinking patterns, and various forms of distraction, including the use of mobile phones, and one-handed driving. Leveraging the power of Mediapipe and YOLOv7 for real-time image analysis, as well as ROS2 for seamless data transfer, our system not only ensures the immediate safety of bus passengers but also offers expandable functionality, such as eye tracking and skeleton detection.

Keywords: Driver Monitoring System, drowsiness detection, machine vision, Mediapipe, YOLOv7

Biztonságosabb buszközlekedés: Moduláris járművezető monitoring rendszer

Nagy Viktor[1], Bézi Patrik[1], Kovács Gábor[1,2]

[1]Széchenyi István Egyetem, Győr, Magyarország
[2]Nemzeti Közszolgálati Egyetem, Információs Társadalom Kutatóintézet, Budapest, Magyarország

Összefoglalás

A biztonságos tömegközlekedés iránti igény fokozza a járművezetőkre nehezedő nyomást a növekvő forgalmi torlódások miatt. Ezt súlyosbítja a közlekedési rendszer összetettsége és a fokozódó külső ingerek hatása, különösen városi környezetben. A közösségi közlekedésben alkalmazható, a közlekedésbiztonságot fokozó moduláris, gépi látáson alapuló, egyedi fejlesztésű járművezető-felügyeleti rendszer kifejlesztése az autóbuszvezetők monitorozására elengedhetetlen. Ez a kutatás egy olyan átfogó rendszert mutat be, amely képes érzékelni az álmosságot, a pislogási mintákat és a figyelemelterelés (disztrakció) különböző formáit, beleértve a mobiltelefon-használatot, és az egykezes vezetést. A Mediapipe és a YOLOv7 valós idejű képelemzésre, valamint a ROS2 adatátvitelre való felhasználásával rendszerünk nemcsak a busz utasainak biztonságát garantálja, hanem olyan bővíthető funkcionalitást is kínál, mint például szemkövetés és csontvázfelismerés (szkeleton).

A rendszer alapvető célja az, hogy a szemmozgás, fejtartás és testtartás elemzésével pontosan azonosítja a járművezető fáradtságát, figyeli a pislogási mintákat az álmosság jeleit, és felismeri a közúti biztonságot veszélyeztető disztrakciókat. Továbbá a rendszer moduláris felépítése lehetővé teszi további érzékelők, például szemmozgás-követő rendszer, telemetriai eszközök vagy 5G-adapterek egyszerű integrálását, ami átfogó megfigyelést és adatfúziót tesz lehetővé a valós környezetbe történő adaptálás elősegítésére.

A fejlesztett Járművezető Monitoring Rendszer a ROS2 keretrendszer segítségével integrált megoldást kínál a buszvezetők megfigyelésére. A rendszer alapvető képessége a fedélzeti kamerák által rögzített felvételeken a járművezető vizuális felismerése. Ezen túlmenően képes követni a vezető testén lévő kulcspontok, mint a fej, a törzs és a karok pozícióját, ami létfontosságú az ő testtartásának és mozgásának megértésében. Az egyediséget a vezetőfülke mérete és a buszvezető személygépjárműhöz mérten dinamikus mozgásképe adja. A rendszer az emberi arcot is részletesen elemzi, kiemelve a fontos arcpontokat, mint a szemek, orr és száj. Ez lehetővé teszi a tekintet irányának, arckifejezéseknek, valamint a fáradtság vagy stressz jeleinek azonosítását. Az adatok könnyebb értelmezése érdekében a rendszer egy vizuális ábrázolást is nyújt az észlelési folyamatról. Az adatok kezelése és kommunikációja a ROS2 keretrendszeren keresztül történik, amely strukturált módon rendezi az adatokat és támogatja a valós idejű feldolgozást és elemzést. Az összegyűjtött adatok tárolására a .rosbag fájlformátumot használjuk, amely lehetővé teszi az adatok hatékony rögzítését és későbbi felhasználását elemzésekhez és felülvizsgálatokhoz.

A tanulmány a moduláris járművezető-felügyeleti rendszer felépítését, megvalósítását és tesztelését mutatja be, részletesen közli az alkalmazott algoritmusokat és technológiákat. A valós körülmények között végzett kísérletek eredményei bizonyítják a rendszer hatékonyságát, valamint a rendszer szélesebb közlekedési ökoszisztémákba való integrálhatóságát.

A buszvezető monitorozása kapcsán kapott adatok hozzájárulhatnak a jármű és utasai biztonságának fokozásához. A járművezető figyelmének nyomon követése és a kognitív terhelés elemzése lehetőséget kínál a munkakörülmények optimalizálására és a balesetmegelőzési megoldások javítására.

Kulcsszavak: Járművezető Monitoring Rendszer, álmosságérzékelés, gépi látás, Mediapipe, YOLOv7

Forewords

In his doctoral research, the KDP PhD student has embarked on a significant exploration into the critical field of driver distraction and cognitive load analysis, particularly emphasizing public transportation. The importance of this research is immense, given the critical need for the safety of buses and their occupants in our ever-changing urban environments. The doctoral candidate has tackled this issue with an impressive combination of scholarly thoroughness and practical acumen. In this study, sophisticated monitoring tools such as eye-tracking devices and front-facing cameras, have been utilized. These instruments have facilitated a deep understanding of driver profiles and behaviors, shedding light on the cognitive mechanisms driving their actions. The impact of this research extends far beyond the academic realm, influencing public safety and transportation policies. It represents a groundbreaking effort in comprehending and reducing driver distraction, a vital aspect in the assurance of safe and efficient public transport systems.

Prof. Dr. Gábor Kovács
Scientific Supervisor
Ludovika University of Public Service,
Széchenyi István University

The convergence of technology and transportation offers tremendous opportunities, especially in the realm of enhancing road safety. This doctoral thesis exemplifies the strength of cooperative innovation, effectively linking academic research with industrial application. Our collaboration has been pivotal in elevating safety standards in public transportation, particularly in the manufacturing and operational facets of bus transportation. The research has resulted in the creation of a bespoke monitoring system, designed specifically for integration with existing on-board telemetry systems. The aim here is not just to observe, but to deeply analyze and foresee safety characteristics in bus drivers' behavior. The monitoring system is designed to assist bus and coach companies in supervising and certifying driver safety, minimizing risks, and mitigating identified hazards. The PhD student's research marks a significant progression in public transport safety and reliability, focusing on the creation of a complex bus driver analysis system. This system includes also a psychological questionnaire and an advanced driving simulator equipped with interface development tools.

Prof. Dr. György Pátzay
Industrial Supervisor
Inter Traction Electrics Ltd. (ITK Holding)

Introduction

The uptake of camera-based DMSs in the automotive industry is significant and their role in accident prevention is unquestionable. The aim of this research project is to design and develop an intelligent monitoring system adapted to public road transport. The complex system is functionally capable of identifying human face and body elements and can be integrated into a modular monitoring system based on the ROS2 framework. The design of the bus DMS will consider the unique characteristics of the driving environment and traffic conditions.

The requirement for secure public transportation is escalating the pressure on drivers due to rising traffic congestion. This is compounded by a diverse array of traffic components and swift external stimuli, particularly prevalent in urban settings. Advanced Driver Assistance Systems (ADAS) can partially intervene in emergency situations, such as braking, but it remains the driver's task and responsibility to safely control the vehicle *(Blades et al. 2020)*. ADAS aim to maximise safety, but such complex systems have side effects. Adoption and use can be difficult, requiring a learning process and confidence building *(Nylen et al. 2019)*. Monitoring driver attention and analysing cognitive load offers the potential to optimise working conditions and improve accident prevention solutions.

Public road transport penetration in Europe is still high. Ensuring the safety of passengers is a priority as vehicles carry more and more passengers in an increasingly congested environment. Although buses and coaches (along with trams and trolleybuses) represent an increasingly smaller share of all modes of transport, it

Table 1 | Buses, coaches and trolleybuses in the European Union – 27 countries

Year	2013	2014	2015	2016	2017	2018	2019	2020	2021
Share (%)	10.4	10.0	10.0	10.0	9.6	9.5	9.5	7.4	N/D
Passenger-kilometer (million km)	238,368	230,616	246,661	240,574	224,813	225,958	231,099	135,427	138,262

Source: Eurostat, 2023

was still 7.4% in the EU in 2020 as shown in *Table 1 (Slootmans 2021)*. An examination of historical accident data reveals that, on average, there were 1.3 fatalities per million inhabitants in bus and coach accidents across EU countries during the period of 2017 to 2019. The victims of fatal accidents are mostly pedestrians, motorists and bus passengers. Accident injury statistics (2010–2019) show that 2% of bus and coach accidents result in serious injuries *(Slootmans 2021)*.

Data analysis has identified several factors that influence the occurrence of accidents: negligence of bus operators, driver negligence and external factors (weather and road conditions) *(Eurostat 2023)*. Driver error can be broken down into several factors, typically depending on the current mental and physical state, so monitoring driver behaviour and subsequent passive or active interventions can ensure safer bus travel *(Goh et al. 2014)*. Distractors refer to visual, auditory, physical or cognitive stimuli that disrupt driving activities including secondary activities such as mobile phone use, conversing with passengers, and daydreaming. Fatigue, monotony and sleep deprivation increase the risk of accidents, as reduced attention can lead to reduced information processing and indecision, leaving the driver unable to react in an emergency situation *(Young–Regan–Hammer 2003)*. Most research on alertness focuses on sleep deprivation *(Otmani–Rogé–Muzet 2005; Huhta–Hirvonen–Partinen 2021; Felix et al. 2022)*. Accident data and simulation studies, however, suggest that a loss of alertness can also occur during the day, particularly on long, monotonous journeys *(Thiffault–Bergeron 2003)*. Driver errors are more frequent while driving monotonously because the required tasks and stimulus levels are low, leading to a reduction in attention.

Various monitoring systems have been developed and used to observe or detect driver behaviour. Different approaches use psychological tests and various physiological sensors, data collection and analysis techniques, as follows: (1) vehicle data-based measures; (2) behavioural measures and (3) physiological measures *(Sahayadhas–Sundaraj–Murugappan 2012)*. Vehicle data-based tests have been used in several studies to identify driver behaviour. Among others, a method has been developed to detect driving habits using single turn vehicle sensor data *(Hallac et al. 2016)*. Human–vehicle interaction analysis systems can monitor fuel consumption, CO_2 emissions, driving style and driver health in real-time with high predictive reliability *(Campos-Ferreira et al. 2023)*.

Several behavioural analyses, mainly using the Driver Behaviour Questionnaire, have shown that professional drivers engage in less risky behaviour than non-professional drivers, but are more likely to be involved in traffic accidents due to longer driving time *(Maslać et al. 2018)*. Some studies have shown that driver behaviour is related to driving conditions. A large sample of intercity bus drivers has been analysed using questionnaires, and psychometric results have highlighted the importance of diagnostics to distinguish safe from unsafe drivers *(Karimi–Aghabayk–Moridpour 2022)*.

Detection of driver fatigue and distraction can be based on biometric cues, steering perception, or observation of the driver's face *(Sigari et al. 2014)*. Some researchers have found that higher average heart rate (HR) indicates that the driver is performing more demanding or secondary tasks *(Biondi et al. 2016)*. Others have conducted heart rate variability analyses validated by electroencephalography to detect driver drowsiness *(Fujiwara et al. 2019)*. Several authors have used wearable Galvanic Skin Responses (GSR) wristband to identify driver distraction *(Dehzangi–Rajendra–Taherisadr 2018)*. Electroencephalogram (EEG) is widely used to detect driver behaviour, but the signals are non-stationary and the procedure is confusing for study participants, especially when using multi-sensor systems *(Li–Chung 2015)*. Recently, single-channel EEG-based drowsiness detection systems have been tested, where the analysis of a single feature parameter allowed for lower processing and storage capacity, leading to simpler implementations in embedded systems *(Balam–Chinara 2021)*. Other physiological sensors, such as electromyogram (EMG) measuring driver muscle fatigue, have also been used and the corresponding measurement methodology has been described *(Rahman et al. 2022)*.

The most common solutions for DMS are non-wearable, camera-based systems that use machine vision for facial recognition. This approach is convenient for drivers as it does not require the use of wearable devices, allowing for non-disruptive monitoring and can be perfectly integrated into production vehicles. According to Regulation 2019/2144/EU, all new European vehicles from 2024 will be required to use "driver drowsiness or distraction alerting" *(European Parliament 2019)*. Visual sensors with different viewing angles, RGB, IR or other cameras facing the driver are used to collect driving data – including driver interaction, behaviour and environment (driver's cab). Typically, the following categories of data are detected: hand movement, body position estimation, facial recognition and distraction/drowsiness *(Koay at al. 2022)*. Face recognition methods can generally be classified into the following categories: feature-based and learning-based methods *(Madhav–Vishwakarma 2015)*. Learning-based methods are generally more robust than feature-based methods, but they usually require more computational resources. However, these methods can achieve a recognition rate of more than 80% under laboratory conditions *(Chaves et al. 2020)*.

This research focuses on the development of a modular bus DMS, employing cutting-edge computer vision and algorithms specifically customised for public transport applications. The system utilises camera-based body, face, and eye recognition technologies to detect mobile phone usage and driver drowsiness. Additionally, it is designed for scalability, allowing integration with further research tools such as eye-tracking devices. The system employs ROS2 for data distribution, facilitating universal data acquisition and streamlined time synchronisation including the connectivity to on-board vehicle data modules, and telemetry systems.

System design and methodology

The development of the bus DMS incorporates an innovative dual-camera setup and sophisticated eye-tracking technology to enhance driver safety and performance. The frontal camera, strategically positioned directly in front of the driver, is pivotal for capturing facial expressions and potential distractions, with a specific focus on monitoring signs of fatigue or inattention. Complementing this, the isometric camera offers a broader perspective, encompassing the driver's hands, posture, and overall body language, which is crucial for detecting behaviors like mobile phone usage. To enhance the system's capability and for validation purpose Pupil Labs Core eye-tracking device was used in the development phase of the research *(Fig. 1)*. This advanced device meticulously captures eye movement data, including gaze direction, pupil dilation, and blink rate. These metrics are vital for assessing the driver's alertness and identifying early signs of fatigue or distraction.

The system's objectives are multifaceted. A primary goal is to utilise advanced imaging for real-time monitoring of the driver's facial features and eye movements, particularly focusing on blink rate and eye closure as indicators of fatigue. Additionally, the system is designed to recognize and track the driver's body posture and movements, identifying key body points like the head, torso, and arms to assess body language and detect potential distractions such as mobile phone usage.

Our method uses convolutional neural networks, which are primarily designed for image processing. The simplified working principle of convolutional neural networks is as follows: the structure of the networks consists of convolutional layers, through which features are continuously filtered; one of the most important of these layers is the so-called pooling layer, which pools redundant data from the matrices describing the images and thus reduces the number of individual parameters *(Ghazal et al. 2018)*.

The system integrates several state-of-the-art technological solutions. YOLOv7, renowned for its speed and accuracy, is employed for its exceptional ability to identify and track the human body in real-time, mapping out key points and movements *(Redmon et al. 2015)*. Complementing YOLOv7, Google's MediaPipe framework provides detailed body point tracking, enhancing the detection of posture, gestures and facial feature tracking, focusing on the eyes and other facial components to

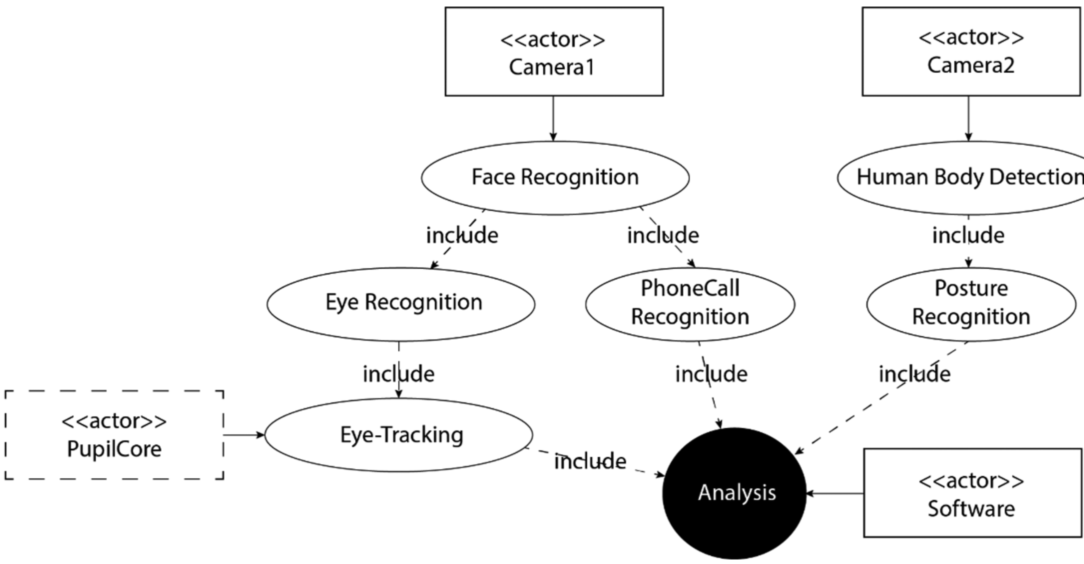

Figure 1 Use case framework of the system

Source: authors's design

analyse blinking patterns and potential signs of fatigue *(Lugaresi et al. 2019)*. The system leverages an integrated approach combining YOLOv7 and Google's MediaPipe with a standard dataset for training, optimising the detection of human behaviours and objects relevant to driving safety. YOLOv7 is selected for its rapid and accurate object detection capabilities, crucial for identifying potential distractions like mobile phone usage. In parallel, MediaPipe enhances the system's ability to track detailed facial features and gestures, enabling precise monitoring of the driver's attention and fatigue levels. The employment of a standard dataset ensures the algorithms are finely tuned across a wide array of driving scenarios, significantly improving the system's accuracy and reliability in real-world applications.

By correlating data from these diverse sources, the system offers a robust solution to monitor and enhance driver safety effectively in the research and development work period. The appropriate solutions have been selected for the research objective, so MediaPipe will be used for face and eye recognition, while YOLOv7 will be used for object recognition, in particular the recognition of mobile phones.

System Requirements

In the context of developing a sophisticated DMS, several functional requirements have been identified and implemented to ensure comprehensive and effective monitoring. The system is designed with the following capabilities:

1. Visual Detection of the Driver on Recorded Footage: The system is equipped to perform visual detection of the driver using the footage recorded by the onboard cameras.
2. Tracking the Position of Key Points of a Recognized Shape: An integral part of the system's functionality is its ability to track the position of key points on the driver's body, such as the head, torso, and arms.
3. Human Face Recognition and Examination of Its Key Points: It examines key points on the driver's face, such as the eyes, nose, and mouth, to gather data on the driver's gaze direction, facial expressions, and potential signs of fatigue or stress.
4. Visual Representation of the Detection: To facilitate easy understanding and analysis of the monitored data, the system provides a visual representation of the detection process.
5. Using the ROS2 Framework to Send Individual Data to a Topic: The system leverages the ROS2 (Robot Operating System 2) framework for efficient data handling and communication. This structured approach ensures organized data flow and supports real-time processing and analysis.
6. Recording Files: For data recording and storage, the system utilises the .rosbag file format.

These functional requirements collectively contribute to the robustness and reliability of the DMS, enabling comprehensive analysis of driver behavior and enhancing overall road safety.

System Setup

Figure 2 illustrates the interconnectedness of individual system elements, called nodes within our monitoring system, highlighting their communication via published

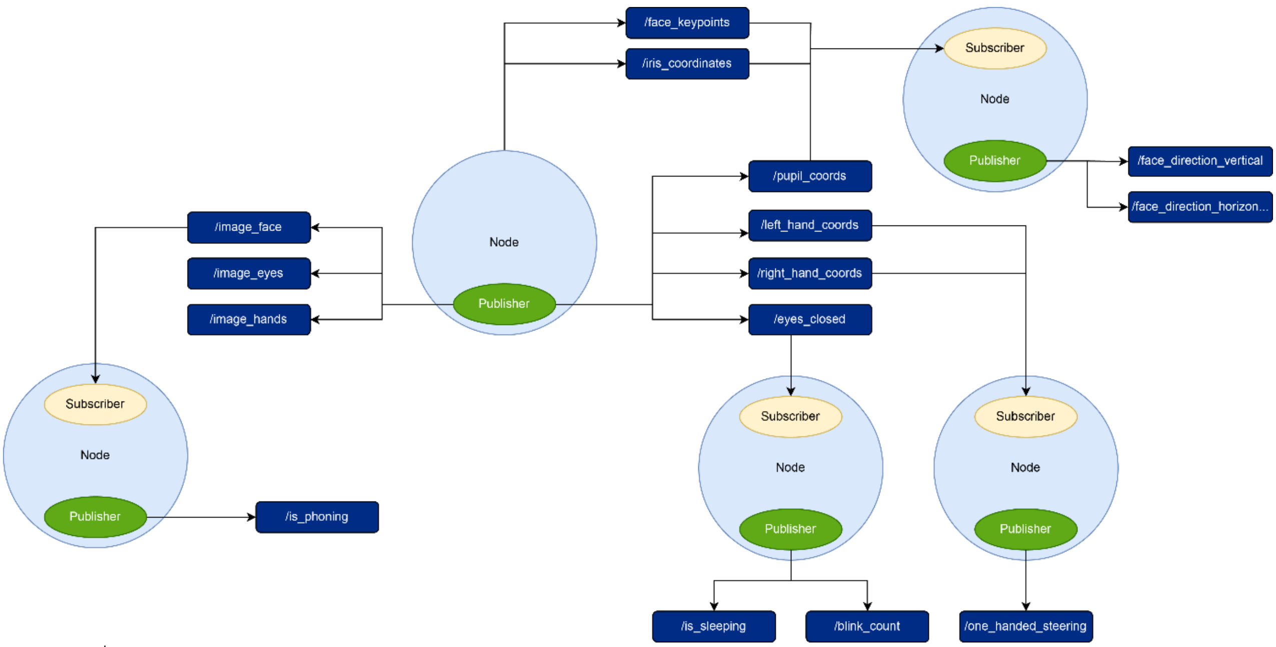

Figure 2 Connection of software nodes

Source: authors's design

messages and the relevance of these messages for executing various tasks. The primary input is an RGB image from which the system initiates by recognizing the human shape. This node operates independently, only publishing data without subscribing to any other input. Three of the published image data are for visualisation purposes, where the main key points of the recognised shape such as eyes, face and hands can be observed. However, the more critical aspect of the system is the publication of data related to individual body parts. This data, to which other nodes subscribe, is crucial for monitoring the gaze, determined by the head position and pupil coordinates. The system not only tracks eye movement but also discerns whether the eyes are open or closed. This feature is vital for monitoring the driver's state, distinguishing between blinking and potential sleepiness while driving. Blink rate analysis, derived from individual recordings, further offers insights into the driver's focus and alertness. Additionally, by analysing the coordinates of the driver's hands, the system assesses hand positioning on the steering wheel and frequency of one-handed steering, providing a comprehensive overview of the driver's engagement and driving habits. Face recognition has a mobile phone detection node which is monitoring if the driver is holding a device in his/her hand or not.

Driver recognition

The first task of the analysis software is to recognise the driver, select the individual key points and track their image coordinates. This task is performed by a separate node in the software package. MediaPipe is used to detect eyes, facial key points, and the entire body. Preliminary solution was used as a body detection and skeleton visualisation tool.

Blinking recognition

Blink detection is implemented through a straightforward yet informative approach, focusing primarily on the eyelid and its immediate area. Utilising MediaPipe's facial recognition technology, we are able to identify a multitude of key points along with their coordinates. The initial step in this process involves locating the eyelid's top, bottom, and sides, which are marked using specific key points. MediaPipe provides these key point coordinates in a normalised format, ranging from 0.0 to 1.0, relative to the dimensions of the input image. To effectively utilise this data, we convert these normalised values into actual pixel points on the image by multiplying them with the image's width and height.

For blink detection, the coordinates are used to determine the positions of specific points around the eye. For instance, the point $P_l(x_l,y_l)$ represents the left side of an eye, and $P_r(x_r,y_r)$ the right side. The distance between these points is calculated as follows:

$$\boldsymbol{d_h} = \sqrt{(\boldsymbol{x_r} - \boldsymbol{x_l})^2 + (\boldsymbol{y_r} - \boldsymbol{y_l})^2}.$$

Similarly, the distance between the upper eyelid $(P_t(x_t,y_t))$ and the lower eyelid $(P_b(x_b,y_b))$ is computed as:

$$\boldsymbol{d_v} = \sqrt{(\boldsymbol{x_h} - \boldsymbol{x_t})^2 + (\boldsymbol{y_h} - \boldsymbol{y_t})^2}$$

These distances, d_h and d_v, represent the width and height of the eye area in pixels, respectively. The system then uses the ratio of these distances (d_h/d_v) to determine if a blink is occurring. By default, a threshold value of 8.0 is used to assess this ratio. The program independently monitors the closure of each eye, enabling it to deduce whether the driver is focusing ahead, experiencing eye irritation, or potentially falling asleep. In such scenarios, the duration for which the eyes remain closed continuously is also observed, providing additional insights into the driver's state. In blink counting, it is important to note whether the subject closes both eyes simultaneously. If only one eye is closed for an extended period, it likely indicates external disturbance rather than a blink. Therefore, our software registers a blink only when both eyes close together.

Fatigue analysis

The noise and monotony of a bus can make you sleepy in the long term – especially in suburban traffic. The driver's alertness is a prerequisite for safe driving, so a system is being developed to distinguish between a person who is asleep and someone who is distracted. It should be noted that being unable to open one's eyes because of an external influence does not mean that one has fallen asleep, but closing one eye certainly does not cause one to fall asleep. This logic can be simply illustrated by a truth table *(Table 2)*.

Table 2 | Truth table for fatigue analysis

A	B	Disturbed	Sleep
1	1	1	1
1	0	1	0
0	1	1	0
0	0	0	0

Source: authors's design

Where A is left eye closed for t time, B is right eye closed for t time.

The software monitors the number of eyes closed and the number of consecutive frames in which this happens. So

$$closed_{left} = \{0, \quad if\ left\ \ eye\ is\ open\ closed_{left} + 1$$

where $closed_{left}$ is the number of consecutive frames in which the driver's left eye is closed and hence

$$closed_{left} \geq minclosedone_{frame}$$

where $minclosedone_{\text{frame}}$ is the minimum number of frames for this, then the left eye of the observed person is obstructed by something that prevents him from opening it. The logic is the same for the right eye. A similar principle is used to determine sleep, but here the condition is that both eyes must be closed for a certain period of time, which is used to say that the driver has fallen asleep. This time is also the number of consecutive frames in which both eyes are closed.

Gaze-monitoring

Gaze-monitoring function to be introduced is also part of the previous module. This function is the gaze tracker, which uses head rotation and irises to estimate which direction the driver is facing. The justification for gaze tracking is that the driver's attention can easily be distracted by anything inside the vehicle. For example, if the driver is talking to other people, to passengers and turns towards them, or in more extreme cases, if he is pressing his phone. The input to this system is therefore the head position or the image coordinates of the iris, and the output is one of the viewing directions that may be: right, left, up, down, front-facing.

The problem was solved using OpenCV's SolvePNP method, which takes as input the camera matrix, the distortion coefficients, and the 3D coordinates of the eye and the 2D coordinates of the facial keypoints *(D'Orazio et al. 2004)*. The 3D coordinates of the facial key points were determined through a comprehensive analysis designed to represent the majority of the population. This involves:

- Data Collection: Gathering a diverse dataset of facial structures from various demographic groups.
- Modeling: Creating a 3D model that encapsulates the average distances and arrangements of facial features.
- Validation: Testing the model against a control group to ensure its accuracy and applicability across different individuals.

This approach ensures that the 3D model used for the SolvePNP method is both representative and versatile, catering to the general population with minimal error margin.

In the tests, the 3D coordinates of the facial key points were determined to fit most people in general. The key points used are the tip of the nose, the chin, and the outer corners of the eyes and mouth. The 2D coordinates are advertised by the node responsible for driver recognition on the corresponding topics. To use the SolvePNP function, the following inputs are required: 3D coordinates of the left/right eye; 2D coordinates of the whole face; Camera matrixes; Distortion coefficients. The outputs are the rotation and translation vectors used to determine 1-1 point relative to the corners of the eyes. To control the sensitivity of the head movement, two variables are introduced for each eye, which are a kind of "scorers" and two multipliers that move the endpoints of the lines describing the eye direction in the horizontal and vertical directions. The vertical and horizontal scores for the left eye are *ly_score* and *lx_score*, and for the right eye *ry_score* and *rx_score*. They are calculated as follows:

$$rx\_score = \frac{x_{right\_iris} - x_{right\_eye\_inner}}{x_{right\_eye\_outer} - x_{right\_eye\_inner}}$$

$$ry\_score = \frac{y_{right\_iris} - y_{right\_eye\_bottom}}{y_{right\_eye\_bottom} - y_{right\_eye\_top}}$$

Where $(x_{\text{right\_iris}}, y_{\text{right\_iris}})$ are the coordinates describing the iris of the right eye in the image, $x_{\text{right\_eye\_outer}}$ is the x coordinate of the outer corner of the right eye, and $x_{\text{right\_eye\_inner}}$ is the x coordinate of the inner corner of the right eye. $y_{\text{right\_eye\_bottom}}$ and $y_{\text{right\_eye\_top}}$ are the y coordinates of the top and bottom of the eye. The formulae are analogous for the left eye. If the last score and the newly calculated score differ only within a certain threshold, the following formula is used to overwrite the score for the right eye:

$$rx\_score = (rx\_score + last\_rx)/2$$

$$ry\_score = (ry\_score + last\_ry)/2$$

For the left eye, the formulas look similar as follows:

$$lx\_score = (lx\_score + last\_lx)/2$$

$$ly\_score = (ly\_score + last\_ly)/2$$

The formulas above can be used to estimate the eye direction using the multipliers defined by the expert. These are *x_score_multiplier* and *y_score_multiplier*. The endpoint of the line describing the direction of the right eye is thus obtained as follows:

$$x_{right\_corner} = x_{right\_corner} - (lx\_score\text{-}0.5) * x\_score\_multiplier$$

$$y_{right\_corner} = y_{right\_corner} - (ly\_score\text{-}0.5) * y\_score\_multiplier$$

The above data may be useful for further analysis in the future to estimate the head orientation. It is sufficient to project the obtained points onto the flat camera image. If we know that the outer corners of the eyes are *l_corner(x,y)* for the left eye and *r_corner(x,y)* for the right eye, it is worth examining the length of the direction vectors for both eyes, as well as the direction of the direction vectors *x* and *y*. The length of the vectors can be calculated using the classical Euclidean algorithm. Then, if the length of the vectors exceeds a minimum, we can be sure that the person is looking left if both direction vectors point left. The same applies to the right, and in all other cases we can assume that the driver is looking horizontally to the centre. The advertised topics are *face_direction_vertical* for vertical and *face_direction_horizontal* for horizontal.

Mobile phone recognition

To recognise the phone, YOLOv7 let is used, since the network yolov7-tiny.pt, trained on the COCO dataset, already contains the mobile phone class *cell_phone (Lin et al. 2014)*. To use this, it is necessary to know the area of the detected object. If its size does not exceed a certain size, it should be ignored as it is likely to be far away from the camera. Equally important is that false positive detections that flash in must be dealt with in some way, as changing conditions can, albeit rarely, fool the YOLO. Given a suitable image occupied area, the coordinates of the edges and corners of the occupying geometry are published in a topic called *cell_phone_coordinates*, which contains arrays of integers. The first element of each array is the phone identifier, since it is assumed that several phones may appear in a frame, for example when passengers board.

Results

The test results show that object detection from both the MediaPipe and YOLOv7 sides worked well on the recordings. Mostly false positives occurred during phone detection, but this problem is relatively easy to solve with the presented technique. Thanks to the accuracy of MediaPipe, the fatigue module works reliably even for people wearing glasses. Sample images of the monitoring result are shown in *Figure 3*.

Figure 3 Monitoring results: a)–b) Pose estimation (preliminary results) for manual distraction detection; c)–d) Face and gaze detection for visual attention detection; e)–f) Mobile phone detection for distraction detection

Source: authors's images

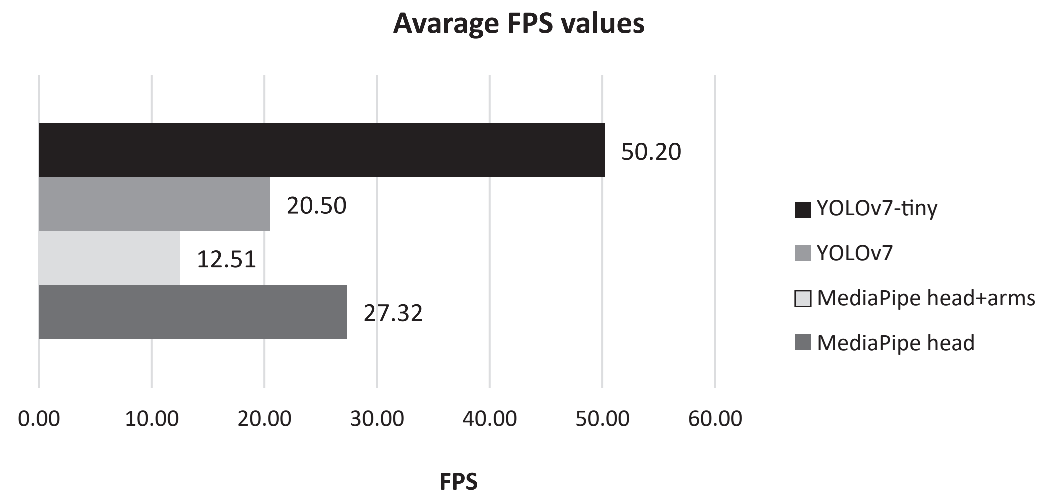

Figure 4 | Performance of the individual recognizers on 720p video

The current configuration of our DMS shows optimal performance when the camera is positioned directly facing the driver. This setup reduces the likelihood of misinterpretations by the posture detector, which can occur with an isometric view. A short-term remedy involves manually adjusting the system's parameters. For a more permanent solution, establishing a baseline for detecting when the driver is looking straight ahead is being considered. Preliminary results suggest that testing the entire body and its key points in an isometric view could enhance system reliability.

Detection problems only occur when light is reflected from the lens of the glasses. Intensive light, such as sunlight, directly hitting the camera lens can also pose an issue, potentially leading to errors in the system's detection and analysis. For the software to function effectively, it is crucial that it visualises the key points being analysed. If these points are not visible, empty coordinates will be sent to each topic, rendering the individual analysis modules inoperative. To mitigate this issue, it is advisable to position the camera in a location where obstruction or visibility issues occur in very few frames. The accuracy of detection can be improved by retraining the neural network.

The collected data results were stored in ROS2 .bag files. This format is advantageous as it enables the recording of results with associated timestamps. Essentially, this allows for the messages from selected topics, or potentially all topics, to be archived in a database. These recorded messages can then be replayed in subsequent tests and trials, facilitating thorough analysis and review.

The MediaPipe has been tested with 720p and Full HD video where the frame rate was approximately 30 FPS, and 20 FPS. Since the detection determines the operation of the other software elements, the data is sent out to each topic on the same frequency. The performance of the recognition algorithms is illustrated in Figure 4, where the testing was done on a 720p resolution video.

The tests were performed on two different machines, a PC and a laptop. Laptop parameters: CPU: AMD Ryzen R7-4800H, RAM: 16 GB DDR4, GPU: Nvidia GTX 1650 4 GB. PC parameters: CPU: AMD Ryzen 5 3600, RAM: 16 GB DDR4, GPU: Nvidia GTX 1660 Super OC 6 GB.

The computational load can be notably reduced by employing multiple neural networks concurrently on a single machine. Consequently, developing a distributed system with two embedded systems, one running YOLO and the other MediaPipe separately, could be a beneficial approach.

Conclusions

Our bus DMS, developed using ROS2, presents a highly modular and scalable solution, adept at integrating various advanced methods. Its design facilitates the seamless addition of new functionalities, such as the proposed integration of YOLOv7 and MediaPipe for enhanced mobile phone detection and key human point identification. Particularly suited for the bus cab environment, the system effectively manages the challenges posed by larger cabin spaces and the driver's dynamic movements. The dual-camera setup enhances pose tracking accuracy, with a front-facing camera being crucial for monitoring potential drowsiness, distraction, and driving behaviours. The system excels in detecting driver fatigue, monitoring drowsiness indicators, and distinguishing distractions, thanks to its machine vision technology. Its modular architecture allows easy integration with additional sensors, offering a broad data fusion scope.

In this research, our focus is on three key areas essential for bus transportation safety, with the understanding that our system is still in the development phase with these functions/dimensions. These include: the impact of the driver's condition, such as fatigue, nervousness, and illness; the effects of incidents occurring within the bus; and the influence of external events on the safety of bus operations. Each dimension is critical to our comprehensive strategy for advancing overall safety in public bus transportation.

In future developments, we aim to utilise ROS2 for enhanced communication and explore the distribution of neural network operations across multiple hardware units to improve accuracy and frame rate. Potential expansion into night-time performance evaluation is also anticipated. A key future goal is to integrate this system into a comprehensive network that includes telemetry and OTA connectivity, with further enhancements like an on-board alert system with audiovisual cues for driver safety.

This paper has detailed the system's architecture, implementation, and real-world trial results, highlighting its effectiveness in enhancing bus driving safety and its potential applicability in wider transportation contexts. By addressing driver distractions and fatigue, the system substantially contributes to road safety and passenger welfare.

Acknowledgement

The research was carried out as a part of the Cooperative Doctoral Program supported by the National Research, Development and Innovation Office and Ministry of Culture and Innovation.

National Research, Development and Innovation Office Hungary

PROJECT FINANCED FROM THE NRDI FUND

References

Balam, V. P., & Chinara, S. (2021) Development of single-channel electroencephalography signal analysis model for real-time drowsiness detection: SEEGDD. Phys Eng Sci Med, Vol. 44. pp. 713–726. https://doi.org/10.1007/s13246-021-01020-3

Biondi, F., Coleman, J. R., Cooper, J. M., & Strayer, D. L. (2016) Average heart rate for driver monitoring systems. Int J Hum Factors Ergon, Vol. 4. pp. 282–291. https://doi.org/10.1504/IJHFE.2016.083521

Blades, L., Douglas, R., Early, J., Lo, C.Y., & Best. R. (2020) Advanced driver-assistance systems for city bus applications. SAE Technical Papers, April. https://doi.org/10.4271/2020-01-1208

Campos-Ferreira, A. E., Lozoya-Santos, J. de J., Tudon-Martinez, J. C., Mendoza, R. A. R., Vargas-Martínez, A., Morales-Menendez, R., & Lozano, D. (2023) Vehicle and driver monitoring system using on-board and remote sensors. Sensors, Vol. 23. https://doi.org/10.3390/s23020814

Chaves, D., Fidalgo, E., Alegre, E., Alaiz-Rodríguez, R., Jáñez-Martino, F., & Azzopardi, G. (2020) Assessment and estimation of face detection performance based on deep learning for forensic applications. Sensors (Switzerland), Vol. 20. pp. 1–21, https://doi.org/10.3390/s20164491

Dehzangi, O., Rajendra, V., & Taherisadr, M. (2004) Wearable driver distraction identification on-the-road via continuous decomposition of galvanic skin responses. Sensors (Switzerland), Vol. 18. pp. 1–16. https://doi.org/10.3390/s18020503

D'Orazio, T., Leo, M., Cicirelli, G., & Distante, A. (2004) An algorithm for real time eye detection in face images. Proceedings of the 17th International Conference on Pattern Recognition, 2004, ICPR 2004. Cambridge, UK. Vol. 3. pp. 278–281. https://doi.org/10.1109/ICPR.2004.1334521

European Parliament (2019) European Parliament REGULATION (EU) 2019/2144 of the European Parliament and of the Council. Official Journal of the European Union.

Eurostat (2023) Eurostat Passenger Transport by Buses and Coaches by Type of Transport - Vehicles Registered in the Reporting Country. 30/03/2023 23:00 2023.

Felix, M., Intriago Alvarez, M. B., Vanegas, E., Farfán Bajaña, M.J., Sarfraz, Z., Sarfraz, A., … & Cherrez-Ojeda, I. (2022) Risk of obstructive sleep apnea and traffic accidents among male bus drivers in ecuador: is there a significant relationship? Annals of Medicine and Surgery, Vol. 74. https://doi.org/10.1016/j.amsu.2022.103296

Fujiwara, K., Abe, E., Kamata, K., Nakayama, C., Suzuki, Y., Yamakawa, T., … & Kadotani, H. (2019) Heart rate variability-based driver drowsiness detection and its validation with EEG. IEEE Trans Biomed Eng, Vol. 66. pp. 1769–1778. https://doi.org/10.1109/TBME.2018.2879346

Ghazal, M., Abu Haeyeh, Y., Abed, A., & Ghazal, S. (2018). Embedded fatigue detection using convolutional neural networks with mobile integration. Proceedings – 2018 IEEE 6th International Conference on Future Internet of Things and Cloud Workshops, W-FiCloud 2018, August. pp. 129–133. https://doi.org/10.1109/W-FiCloud.2018.00026

Goh, K., Currie, G., Sarvi, M., & Logan, D. (2014) Factors affecting the probability of bus drivers being at-fault in bus-involved accidents. Accid Anal Prev, Vol. 66. pp. 20–26. https://doi.org/10.1016/j.aap.2013.12.022

Hallac, D., Sharang, A., Stahlmann, R., Lamprecht, A., Huber, M., Roehder, M., … & Leskovec, J. (2016) Driver identification using automobile sensor data from a single turn. IEEE Conference on Intelligent Transportation Systems, Proceedings, TSC 2016. pp. 953–958. https://doi.org/10.1109/ITSC.2016.7795670

Huhta, R., Hirvonen, K., & Partinen, M. (2021) Prevalence of sleep apnea and daytime sleepiness in professional truck drivers. Sleep Med, Vol. 81. pp. 136–143. https://doi.org/10.1016/j.sleep.2021.02.023

Karimi, S., Aghabayk, K., & Moridpour, S. (2022) Impact of driving style, behaviour and anger on crash involvement among iranian intercity bus drivers. IATSS Research, Vol. 46. pp. 457–466. https://doi.org/10.1016/j.iatssr.2022.07.003

Koay, H. V., Chuah, J. H., Chow, C. O., & Chang, Y. L. (2022) Detecting and recognizing driver distraction through various data modality using machine learning: A review, recent advances, simplified framework and open challenges (2014–2021). Eng Appl Artif Intell, Vol. 115. 105309. https://doi.org/10.1016/j.engappai.2022.105309

Li, G., & Chung, W. Y. (2015) A context-aware eeg headset system for early detection of driver drowsiness. Sensors (Switzerland), Vol. 15. pp. 20873–20893. https://doi.org/10.3390/s150820873

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., … & Zitnick, L. (2014) Microsoft COCO: Common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., & Tuytelaars, T. (eds) Computer vision – ECCV 2014. ECCV 2014. Lecture notes in computer science, Vol. 8693. Springer. https://doi.org/10.1007/978-3-319-10602-1_48

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., … & Grundmann, M. (2019) MediaPipe: A framework for building perception pipelines. arXiv:1906.08172

Maslać, M., Antić, B., Lipovac, K., Pešić, D., & Milutinović, N. (2018) Behaviours of drivers in Serbia: Non-professional versus professional drivers. Transp Res Part F Traffic Psychol Behav, Vol. 52. pp. 101–111. https://doi.org/10.1016/j.trf.2017.11.020

Nylen, A. B., Reyes, M. L., Roe, C. A., & McGehee, D. V. (2019) Impacts on driver perceptions in initial exposure to ADAS technologies. Transp Res Rec. https://doi.org/10.1177/0361198119847975

Otmani, S., Rogé, J., & Muzet, A. (2005) Sleepiness in professional drivers: Effect of age and time of day. Accid Anal Prev, Vol. 37. pp. 930–937. https://doi.org/10.1016/j.aap.2005.04.011

Rahman, N. A. A., Mustafa, M., Sulaiman, N., Samad, R., & Abdullah, N. R. H. (2022) EMG signal segmentation to predict driver's vigilance state. Lecture Notes in Mechanical Engineering, pp. 29–42. https://doi.org/10.1007/978-981-16-4115-2_3

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015) You only look once: Unified, Real-time object detection. arXiv: 1506.02640.

Sahayadhas, A., Sundaraj, K., & Murugappan, M. (2012) Detecting driver drowsiness based on sensors: A review. Sensors (Switzerland), Vol. 12. pp. 16937–16953. https://doi.org/10.3390/s121216937.

Sigari, M.-H., Pourshahabi, M.-R., Soryani, M., & Fathy, M. (2014) A review on driver face monitoring systems for fatigue and distraction detection. International Journal of Advanced Science and Technology http://sersc.org/journals/index.php/IJAST/index, Vol. 64. pp. 73–100. https://doi.org/10.14257/ijast.2014.64.07

Slootmans, F. (2021) European road safety observatory - Facts and figures - Light trucks. https://road-safety.transport.ec.europa.eu/system/files/2021-12/F%26F%20Light%20trucks%2020210712.pdf [Downloaded: 04/12/2023]

Thiffault, P., & Bergeron, J. (2003) Monotony of road environment and driver fatigue: A simulator study. Accid Anal Prev, Vol. 35. pp. 381–391. https://doi.org/10.1016/S0001-4575(02)00014-3

Young, Ch., Regan, M., & Hammer, M. (2003) Driver distraction: A Review of the literature (Report). Monash University Accident Research Centre, 66.