Considerations on Monitoring the Drowsiness of Drivers Through Video Detection and Real-Time Warning

Maria Claudia Surugiu(✉) and Ion Nicolae Stăncel

Department of Telematics and Electronics in Transportation, Faculty of Transportation, University "Polytechnica" of Bucharest, Bucharest, Romania
claudia.surugiu@upb.ro

Abstract. Microsomnia, decreased concentration and fatigue at the wheel are particularly dangerous and are the cause of many accidents. However, the initial signs can be detected in advance: tired, low-attention drivers perform less precise steering maneuvers and have to make minor path corrections more often.

The willingness to take over the vehicle control in driving scenarios, in autopilot mode, is an important factor for road safety. This paper presents a low-cost system for automatic recognition of driver activity by eye monitoring. Thus, an architecture based on eye movement and blink tracking data is introduced in this system, thus analyzing several features. It is estimated that this technology will help prevent acci-dents caused by drivers who become drowsy. Various studies have suggested that about 20% of all road accidents are related to fatigue.

Keywords: Video sensors · Driving attention · Fatigue detection · Algorithm

1 Introduction

Automotive electronic systems are experiencing an unprecedented age of development. For luxury vehicles, it is estimated that electronic systems account for about 23% of the total cost of the vehicle [1]. Experts also estimate that about 80% of automotive innovations are mostly in the electronic system field. The main purpose of automotive electronic systems is for driver assistance in the vehicle control, by using traction (engine), steering (electric power steering), braking system (ABS, ESP) or suspension (active suspension). In early experiments conducted by Google with the autopilot vehicle division, the company reported test drivers falling asleep while driving at a speed of 55 km/h on the highway. The possibility of vehicle control taking in certain driving scenarios, in autopilot mode, is an important factor for road users' safety. Drowsiness is one of the most common causes of road accidents [1, 2]. The effects of fatigue can be compared to those of driving under the influence of alcohol. Like alcohol, drowsiness slows down the reaction time, which leads to an increased risk of being involved in a car accident [2].

Causes of drowsiness or fatigue include.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
L. Moldovan and A. Gligor (Eds.): Inter-Eng 2021, LNNS 386, pp. 661–671, 2022.
https://doi.org/10.1007/978-3-030-93817-8_59

- Insufficient sleep;
- Fatigue accumulated over time;
- Untreated sleep disorders;
- Use of sedative drugs.

Drowsiness or fatigue can cause the following problems:

- Slowing the reaction time, visual deficiencies, judgment deficiencies;
- Difficulties in processing information;
- Decreased performance, motivation and alertness;
- May cause irritable and aggressive behavior.

The incidence of accidents caused by fatigue usually occurs between 2:00–6:00 and 14:00–16:00. There is a 20 times higher probability that a driver will fall asleep behind the wheel at 6:00 compared to 10:00. Motorways are roads where most accidents are common due to fatigue driving, because of the lack of incentives for drivers. A study conducted on professional drivers, indicated that the risk of having an accident due to fatigue begins to increase after 9–10 h of driving [1]. This risk doubles after this period. According to a study by the AAA (Foundation for Traffic Safety) on 3500 drivers in the United States, monitored over a period of six months on board mounted cameras, it was shown that they were involved in 700 road accidents. The drowsiness/fatigue condition caused 9.5% of those accidents. Depending on the country and region, statistics indicate that fatigue and drowsiness while driving can cause between 10% and 30% of all road accidents. According to the Police Headquarters in Romania, in 2017, 190 accidents resulting in 227 deaths were caused by drivers who fell asleep at the wheel. Most of the accidents occurred on weekends, on national roads. Unfortunately, Romania ranks 10th in the top of road accidents caused by drowsiness. More than a million Romanians have sleep disorders, but only 5000 are diagnosed.

2 State of the Art

In the last 5 years, new technologies are used to manage the amount of information, obtained from video tools, in the form of objects and images. Different hybrid systems in terms of how to interpret this image and object date are concatenated in homogeneous applications, such as detection systems that use the Convolutional Neural Networks (CNN) with the Scale Invariant Feature Transform (SIFT) algorithm integration [3]. New methods of implementing visual objects recognition based on deep learning with CNN are adapted to both demanding applications in terms of hardware, which has driven the development and research of low-cost systems, which consists in the use of open-source platforms, such as be Rasberry Pi [4]. Also, in terms of latency and accuracy, CNN networks are studied, from the perspective of eye indicators status which controls the detection system. The main types of networks used are Fully Designed Neural Network (FD-NN), Transfer Learning in VGG16 and VGG19 with extra designed layers (TL-VGG) [5].

Vehicle manufacturers use video cameras, sensors to track head position, steering wheel monitoring and audible alerts to ensure drivers pay attention when using advanced driver assistance systems, such as autopilot. Several car manufacturers offer drowsiness detection systems [6]. In various studies, [7] the most used techniques for detecting fatigue are based on monitoring physiological parameters: brain waves, heart rate and respiration. These methods are intrusive and require the use of sensors attached to the driver, which is inconvenient. Other techniques have also been studied to detect the fatigue of drivers by moving their eyes and blinking, [8] thus proving that if the blink rate increases, it indicates a state of fatigue.

For example, BMW uses a video camera mounted in the car to solve a critical challenge with autopilot systems, ensuring that drivers pay attention to the road. The driver-monitoring system proposed by Ford consists of a small traffic-oriented camera connected to an on-board computer. The system offered by Mercedes draws up an individual profile of the driver's driving style, which is constantly compared with the latest sensor feedback. At the center of the system is an extremely sensitive sensor, which records direction and speed of the vehicle [9]. The warning system used by Nissan presents the following solution: drivers usually make small adjustments and corrections to the steering wheel while driving. It is common for tired drivers to stop or slow down their steering wheel movements. The system used by Subaru uses a dedicated video camera and facial recognition software that tracks the driver's activity to calculate two stages of fatigue - sleepy and extremely sleepy. The system offered by VW closely monitors the driver's behavior, noticing any irregular movement of the steering wheel, the pedal use and any deviations from the lane (Lane Assist), so it can judge when the driver begins to feel drowsy and must stop and take a break [10]. If the system detects that the driver is starting to lose focus, he will be warned with a visual display on the instrument panel and a warning sound.

3 Description of Work

The proposed system includes a video camera aimed at generating images of the driver, including the eyes. It also includes a processor for processing images generated by the camcorder. The processor monitors the acquired image and determines whether the eye is in an open or closed position. The processor also determines a proportion of the closing time of the eyes as the proportion of a time interval in which the eye is in the closed position and determines a state of drowsiness when the proportion of time exceeds a certain threshold value. The sleep detector system will use a single-color video camera to capture images from the passenger compartment. The camera is mounted firmly on the driver's side to see his upper body. Image analysis will be performed using the eye tracking algorithm and motion detection.

Eye Detection

Algorithms commonly used for eye detection include the Hough transform, matching predefined patterns, principal component analysis (PCA), and the Adaboost algorithm.

In this paper, the eye *is monitored by facial cues detection* to locate important regions of the face [10].

Blinking Detection

Blinking is detected using the Eye Appearance Report (EAR), introduced by [11].

Extraction of Imaging Features

The pre-trained facial landmark detector, iBUG 300 W, from the Dlib library is used to estimate the location of 68 coordinates (x, y) that match the facial structures (see Fig. 1) [11–13].

Fig. 1. View over the 68 coordinates of the face marker in the iBUG 300 W data set [11].

The IBUG database was released as part of the first 300-W version. It consists of 135 images downloaded from the web, with large variations in expression, lighting conditions and positions [11, 12]. Face Marker Detection uses an input image and a shape detector that tries locating key points of interest along the face.

Therefore, facial detection cues are a two-step process [11, 12]: face location in the image and facial structures detection.

Face Location in the Image

For this first stage, an object detector will be applied: Oriented Gradient Histogram (HOG) and SVM (Support Vector Vectors) classification algorithm pre-trained for the face detection task.

Oriented Gradient Histogram (HOG)

HOG descriptors are feature descriptors used in image processing for object detection purposes. The technique is based on monitoring the number of occurrences of the orientation of a gradient (variation per unit length of a scalar quantity) in a certain region of the image. Below, a generated face model by HOG is presented (see Fig. 2). The main idea is for the local appearance and shape of an object in an image to be described by the gradient intensity distribution and the edge orientation distribution [11].

The calculation of the gradient values is done by applying one-dimensional bypass filters horizontally and vertically. The filtering is done with two convolution nuclei:

$$D_x = [-1\ 0\ 1]$$
$$D_y = [-1\ 0\ 1]^T$$

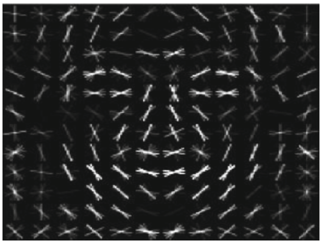

Fig. 2. Face model generated by HOG

Using a convolution operator $I_x = I \cdot D_x$ și $I_y = I \cdot D_y$ we will obtain horizontal and vertical derivatives. The magnitude of the gradient is $|G| = \sqrt{I_x^2 + I_y^2}$ and the gradient orientation is $\arctan\left(\frac{I_y}{I_x}\right)$. The second step is to create the histogram for each cell. Each pixel that forms this cell will be represented in the resulting histogram.

Fig. 3. The magnitude of the gradient

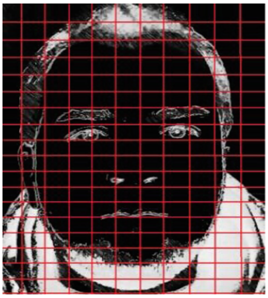

Fig. 4. Division into cells

To consider, the variations in brightness and contrast, the grouping of cells in larger structures (blocks) (see Fig. 3 and Fig. 4) will take place to locally normalize the magnitude of the gradients. The HOG descriptor will result from the normalized components of the histogram. The blocks partially overlap so that each cell contributes several times to the final descriptor [11].

Normalization of blocks

The normalization factor is calculated with one of the Eqs. (1), (2) below, [12]:

$$L_2 - norm:\ f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}} \tag{1}$$

$$L_1 - norm:\ f = \frac{v}{\sqrt{\|v\|_1 + e}} \tag{2}$$

where

$v =$ the non-normalized vector containing all the histograms of a block;

$\|v_k\| =$ is the norm k for $k = 1, 2$; $e =$ constant of very low value.

Euclidean distance between two vectors $P = (p_1 p_2, \ldots p_n)$ and $Q = (q_1 q_2, \ldots q_n)$ is:

$$d = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \ldots + (p_n - q_n)^2} \tag{3}$$

Support vector machine (SVM)

Mainly, the studied topic can be classified into two classes (sleepy or non-sleepy). SVM was chosen because it is used generically for binary classification issues and has attributes that make it fit perfectly [10, 11]. SVM is a set of supervised learning methods used to classify, regress, and detect external values. It uses a subset of decision- making training points called support vector. Different Kernel functions can be specified for the decision function. An SVM creates a hyper-plane or set of hyper-planes in a large or infinite dimensional space that can be used for classification, regression, or other objectives [9, 11]. A good separation is achieved by the hyper-plane that has the greatest distance to the nearest data points formation of any class (the so-called functional margin), because, in general, the higher the margin the lower the classifier generalization error gets.

The separation surface is described by the following Eq. (4):

$$(w \cdot x) + b = 0 \tag{4}$$

where vectors are represented by w and x; b represents a scalar.

The SVM method assumes that the separation surfaces equations are normalized. The two classes will have the limits that pass-through support vectors in Fig. 5, and Eqs. (5), (6):

$$(w \cdot x) + b = -1 \tag{5}$$

$$(w \cdot x) + b = 1 \tag{6}$$

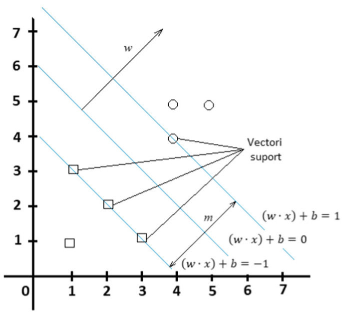

Fig. 5. Support vectors and characteristic problem notions illustration.

The actual classification involves determining w and b parameters that maximize the separation margin between classes.

Facial Structures Detection

Facial mark detectors try locating and label the following facial regions: Mouth, Nose, Jaw, Left and Right Eye, Left and Right Eyebrow [11]. The detection method begins with a training data set with facial markers manually labeled on an image. These images are part of a library that specifies the coordinates (x, y) of the regions surrounding each facial structure [11]. Using this dataset, a set of regression trees is "trained" to estimate the positions of facial marks using pixels [12, 13].

3.1 Real-Time Blink Detection Using Facial Cues

The proposed algorithm estimates the mark positions, extracts a single scalar quantity, the Eye Aspect Ratio (EAR), [11, 13] characterizing the opening of the eye in each frame. Finally, an Support vector machine (SVM) [13] classifier detects blinking as a pattern of EAR values in a short time window. Existing methods are active or passive for Eye aspect ratio (EAR), [11–13]:

- Active methods are reliable, but they use special hardware, often expensive and intrusive, such as infrared cameras and illuminators, wearable devices, glasses with a special video camera that observes the eyes;
- Passive methods are based on a standard remote camera only.

In this paper, the passive system was studied. Regarding the detection of blinking; only two sets of facial structures, namely, the eyes, were of interest. Each eye is represented by 6 coordinates (x, y), starting from the left corner of the eye and then clockwise around the rest of the region (see Fig. 6).

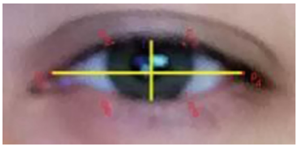

Fig. 6. The 6 facial features associated with the eye.

In these coordinates, there is a relationship between the width and height. Based on the real-time detection of the clip using facial cues, an equation can be derived that reflects this relationship called the Eye Aspect Ratio in (7), [11]:

$$EAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|} \tag{7}$$

where $p_1,\ldots, p_6$ are 2D facial landmarks.

The numerator of this equation calculates the distance between the vertical landmarks of the eye, while the denominator calculates the distance between the horizontal landmarks of the eye. Using this simple equation, we can avoid image processing techniques and simply rely on the eye reference distance ratio to determine whether a person is blinking. To make this clearer, we will demonstrate the results [11–13] in the images below (see Fig. 7.a), b)). In the Fig. 7 a) a completely open eye is presented, so the aspect ratio of the eye here is large and relatively constant over time. In the Fig. 7 b), you can see that once the person blinks, the aspect ratio of the eye decreases dramatically, approaching zero [11]:

$$EAR = \begin{cases} X > 0, & eyes\,opened \\ 0, & eyes\,closed \end{cases} \tag{8}$$

While the eyes are closing, the EAR result will be about 0, while during open eyes, the EAR can be any integer x greater than 0, [12, 13, 15].

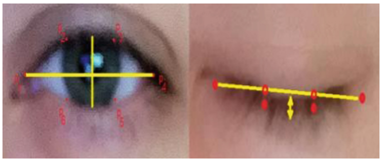

Fig. 7. Eye marks when the eye is open and closed.

3.2 Vehicle Equipment

The paper presents a low-cost system for automatic recognition of driver activity by eye monitoring. Therefore, an architecture based on eye movement and blink tracking data is introduced in this system, thus analyzing several features. The camera used in this project is an HP HD-3110. It has a video resolution of 720p, 1280 * 720 pixels, and autofocus. The camera is mounted on top of the dashboard and is connected to the Raspberry Pi control unit [14, 15]. After the assembly was completed, the drowsiness detector was built using computer vision techniques (see Fig. 8).

Fig. 8. The position of the camera in the vehicle.

OpenCV was used, an open-source library (which provides certain finished products, allowing users to modify and improve it without any obligation) for the field of image processing and not only, originally developed by Intel. The library has interfaces for C/C++, Python, Java, Matlab and runs on Windows, Linux and Mac OS X. This technology is especially based on continuous image processing. At this time the OpenCV library supports continuous time captures, object detection, application of simple filters on images.

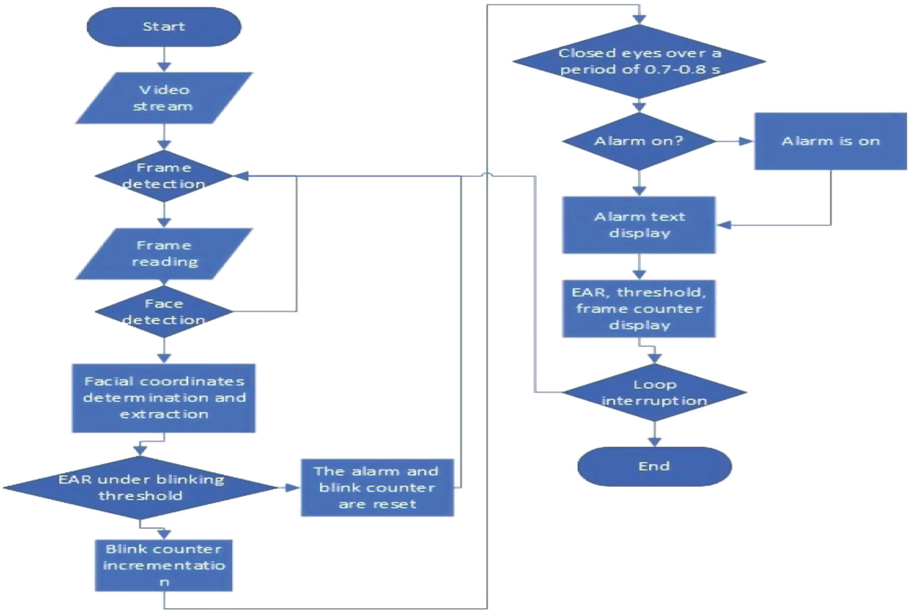

Fig. 9. Proposed system algorithm.

3.3 Drowsiness Detector Test

The general flow of the proposed algorithm for detecting drowsiness is simple. First, the camera will be configured to monitor the driver of the vehicle. When a face is identified, face detection is applied, and the eye regions are extracted. After determining the eye regions, the Eye Aspect Ratio (EAR) is calculated to determine whether the eyes are closed and if they are closed for a period, an alarm is activated [11, 12, 15] (see Fig. 10 a) and b)).

Once the software was running, the video stream began to be processed, followed by careful testing of the drowsiness detector. After completing the initial tests in a closed perimeter, the aim was to perform the test on a slightly crowded road because driving with closed eyes, even for a second, can be dangerous.

a) Face identification in the input video stream

b) Locating facial features to extract eye regions

c) Alarm status

Fig. 10. Stages of sleepiness detection.

The results showed that the sleep detector can identify when the driver is at risk of falling asleep and emits an audible alarm to attract his attention. The drowsiness detector can operate in various conditions, including direct sunlight or low artificial lighting (see Fig. 10 c)).

4 Conclusions

The proposed system is an experimental model based on the use of a video camera for face detection mounted on the top of the dashboard and a Raspberry Pi Zero W development board. The experimental model has a low price and can be used to emit a sleep signal in case of detection of the driver's fatigue.

Algorithms commonly used for eye detection include the Hough transform, matching predefined patterns, principal component analysis (PCA), and the Adaboost algorithm. In this paper, the eye was detected using facial cues to locate important regions of the face. Drowsiness is a relevant and real problem that is the cause of many road accidents. The number of victims of these accidents is high and has been constant recently, with an associated economic impact. All these determine a strong motivation for developing a measure to solve this problem. The implementation of this system on autonomous vehicles will lead to increased traffic safety for both drivers and pedestrians.

References

1. Joly, A., Zheng, R., Kaizuka, T., Nakano, K.: Efect of drowsiness on mechanical arm admittance and driving performances. IET Intell. Transp. Syst. **3**(12), 220–226 (2018)
2. Tripathi, A., Kumar, T.V., Dhansetty, T., Kumar, J.: Real time object detection using CNN. Int. J. Eng. Technol. (UAE). **7**, 33–36 (2018). https://doi.org/10.14419/ijet.v7i2.24.11994
3. Shehab, M.A., Al-Gizi, A., Swadi, S.M.: Efficient real-time object detection based on convolutional neural network. In: International Conference on Applied and Theoretical Electricity (ICATE) 2021, pp. 1–5 (2021). https://doi.org/10.1109/ICATE49685.2021.9465015
4. Hashemi, M., Mirrashid, A., Beheshti Shirazi, A.: Driver safety development: real-time driver drowsiness detection system based on convolutional neural network. SN Comput. Sci. **1**(5), 1 (2020). https://doi.org/10.1007/s42979-020-00306-9
5. Vural, E.: Video based detection of driver fatigue, Graduate School of Engineering and Natural Sciences, Sabanci University, Spring (2009)

6. Dong, Y., Hu, Z., Uchimura, K., Murayama, N.: Driver inattention monitoring system for intelligent vehicles: a review. IEEE Transp. Intell. Transp. Syst. **12**, 596–614 (2011)
7. Meireles, T., Dantas, F.: A low-cost prototype for driver fatigue detection. Multimodal Technol. Interact. **3**(5), 1–11 (2019)
8. https://www.ebay.com/itm/Mercedes-ME9-7-ME-9-7-ECU-ECM-Engine-Computer-Programming-Cloning-Unlocking-x2-/262689159375
9. Barea, R., Boquete, L., Mazo, M., Lopez, E.: System for assisted mobility using eye movements based on electrooculography. IEEE Trans. Neural Syst. Rehabil. Eng. **10**(4), 209–218 (2002)
10. Cech, J., Soukupova, T.: Real-time eye blink detection using facial landmarks. In: Cehovin, L., Mandeljc, R., Struc, V. (eds.) 21st Computer Vision Winter Workshop Luka, Rimske Toplice, Slovenia, 3–5 February 2016
11. Chen, M.-C., Chen, J.-L., Chang, T.-W.: Android/OSGi-based vehicular network management system. Comput. Commun. **34**(2), 169–183 (2011)
12. Tzimiropoulos, G.,: Project-out cascaded regression with an application to face alignment. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, June 2015
13. Fischer, I., Hennecke, F., Bannes, C., Zell, A.: Java neural network simulator web site (2001). http://www.ra.cs.uni-tuebingen.de/software/JavaNNS/welcome.html
14. Mouser Electronics. https://pt.mouser.com
15. Leon, F.: Artificial Intelligence: Cars with Support Vectors. Tehnopress, Iasi (2014)