Driver's drowsiness detection using an enhanced image processing technique inspired by the human visual system

Hedyeh A. Kholerdi, Nima TaheriNejad, Reza Ghaderi & Yaser Baleghi

Published online: 22 Feb 2016.

Submit your article to this journal

Article views: 266

View related articles

View Crossmark data

Citing articles: 4 View citing articles

Taylor & Francis
Taylor & Francis Group

Driver's drowsiness detection using an enhanced image processing technique inspired by the human visual system

Hedyeh A. Kholerdi[a], Nima TaheriNejad[b], Reza Ghaderi[a][†] and Yaser Baleghi[a]

[a]Department of Electrical and Computer Engineering, Babol Nushirvani University of Technology, Babol, Iran; [b]Institute of Computer Technology, TU Wien, Gusshausstrasse 27-29, Vienna, Austria

ABSTRACT

Unfit drivers are the cause of tens of thousands of incidents on the roads which lead to injuries and deaths. Therefore, it is very important to take preventive measures against such incidents. One of the unfit driving conditions is driving while being drowsy. Using image processing techniques, drowsiness of the driver could be detected and hence such incidents could be prevented. In this work, inspired by how images are processed by the human visual system, an enhancement for driver's drowsiness detection is suggested. Furthermore, to improve the robustness of the drowsiness detection system, the mechanism for using energy levels in frames is changed. Lastly, a better decision making process is proposed. To measure the merit of the system, it is applied to a set of drivers' data. Test results show that using the proposed system, success rate of the drowsiness detection system is 90%.

1. Introduction

Road incidents are the cause of hundreds of thousands of injuries and deaths every year. Irresponsibly driving in an unfit condition, such as driving in a state of drowsiness, is one of the major causes of these incidents. According to a survey published in 2003 by the National Highway Traffic Safety Administration, data shows that drowsiness is the cause of more than 56,000 car crashes a year in the USA; more than 40,000 non-fatal and 15,500 fatal incidents (Royal, 2003). According to European Transport Safety Council (2001) in England, drivers' drowsiness accounts for approximately 20% of crashes each year. The Royal Society for the Prevention of Accidents (2001), Birmingham-England, also published a review of car crashes and their respective causes in USA, England, Australia, New Zealand, Germany and many other countries. This review states that from 5% to 25% of car accidents happen due to drowsiness. In consequence, driver drowsiness detection has recently attracted further attention of researchers.

CONTACT Hedyeh A. Kholerdi ✉ hedyeh.kholerdi@ymail.com
†Currently with the School of Cognitive Sciences Institute for Research in Fundamental Sciences, Niavaran Sq., Tehran 19395-5746, Iran

In this paper, we present an image processing technique used to detect the state of eye, mouth and head. These techniques are based on the model of retina and visual cortex V1. We then introduce a new feature extraction method to estimate the state of eye, mouth and head in a sequence of images. This technique uses the energy of static information in an image. To improve this process, a sharpening filter is added to the system. Finally, a robust decision algorithm is proposed. This algorithm detects the driver's drowsiness based on various possible combinations of extracted features.

The rest of this paper is organised as follows: Section 2 provides a brief overview of the literature and an outline of the driver's drowsiness detection system. Section 3 includes the fundamental information on the pre-processing. In Section 4, details of feature extractions and the process of drowsiness detection (decision making process) are discussed. Section 5 provides the description of enhancements and changes in the proposed system. The experimental set-up and results are described in Section 6. The rest of that section includes an extensive discussion on the performance of the suggested system compared to other recent works on driver's drowsiness detection. Finally, the paper is concluded by Section 7.

2. Literature overview

One of the first methods of measuring the level of driver's consciousness and fatigue, is using driver's Electro-Encephalo-Graphy (EEG) and Electro-Cardio-Graphy (ECG) signals (Picot, Charbonnier, Caplier, and Vu, 2012). Picot, Charbonnier, and Caplier (2012) propose an algorithm to categorise the EEG signal. They use Electro-Oculo-Graphy to capture blinking information and finally determine the level of fatigue. However, the inconvenience of wired sensors being connected to the driver's body makes such methods impractical. This gives way to image processing techniques as an alternative practical solution.

The flowchart shown in Figure 1 shows the three main stages of driver's drowsiness detection using image processing techniques. These stages are pre-processing, feature extraction and decision making. Not all methods go through all the stages and the ones which do, use various algorithms for them. In the proposed system, as we will show, we take advantage of the added-value of all these stages. In the following part of this section, some of these techniques will be briefly discussed.

The research which has been carried out using image processing is mostly based on feature extraction and the analysis of spatial characteristic of facial features (Delmas, Eveno, and Liévin, 2002; Jo, Lee, Park, Kim, and Kim, 2014; Tian, Kanade, and Cohn, 2000; Wu and Chen, 2008). The algorithm proposed by Wu and Chen (2008) binarises the image of eyes and then counts the number of black and white pixels. Algorithms such as lip segmentation (Delmas et al., 2002) or mouth shape estimation (Tian et al., 2000) are other examples for tracking the state of facial features. These methods are very sensitive to the appearance of facial features, skin colour and noise. Some other approaches use a combination of these features for fatigue detection. Nonetheless, closed-eyes remains the strongest feature of drowsiness detection. However, often the combination of yawning, closing eyes and dropping head can cast serious challenges upon these algorithms. Therefore, proper detection of this combination, as we will show in this paper, plays an important role in improving the robustness of the system.

One of the inspiring works in the field, namely (Benoit and Caplier, 2010), detects the state of the eyes, mouth and head based on the model of retina and visual cortex V1. The
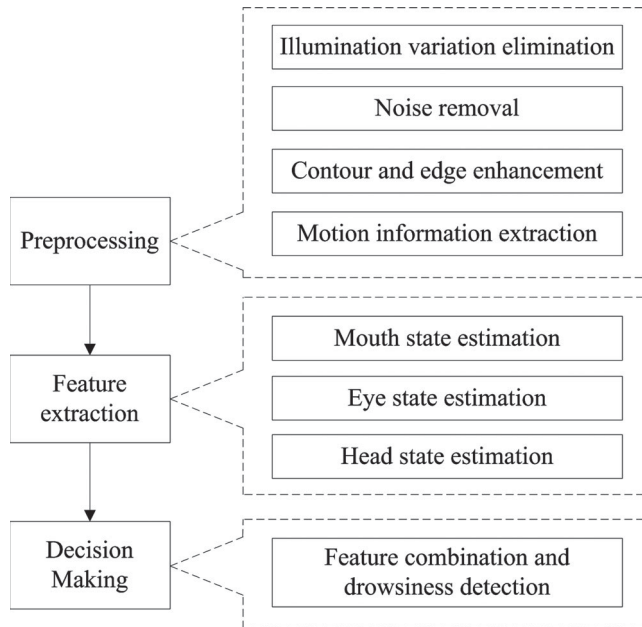
Figure 1. Detection of driver's drowsiness using image processing techniques; an overview.

multi-scale Retinex filter proposed by Rahman, Jobson, Woodell, and Hines (2005) uses the combination of dynamic range compression, colour consistency, and colour /lightness tonal rendition to enhance a digital image. A similar process is done in the retina and in the cortex for this purpose. An image-coding scheme presented in Senane, Saadane, and Barba (2001) is another model inspired by the human visual system (HVS). In their work, visual information are analysed and quantised based on the image decomposition and low pass filtering, similar to the visual cortex. However, this method does not pre-process the images, as retina does in the HVS.

Pre-processing is the key missing element in other related works as well. One of the most interesting works in literature in terms of performance is Azim, Jaffar, and Mirza (2014). They report a 100% correct drowsiness detection for their proposed system. Their system uses a classifier to estimate the state of mouth and eyes, based on spatial features of the image. Even though the system has a performance of 100% on their test video, it has not been tested in a variety of conditions. Sensitivity of a classifier to various conditions, such as facial appearances, could potentially degrade the performance of their algorithm. This effect could be more pronounced since there seems to be no pre-processing to eliminate some of such effects. Therefore, further experiments seem to be warranted for a comprehensive evaluation of their system.

Next to that is the system proposed by Jo et al. (2014). A camera set on the dashboard with a near InfraRed (NIR) filter and two NIR illuminating light emitting diodes installed on each side of the camera capture the images for their system. They only analyse the state of eyes and use support vector machine classifier and 2D Gaussian model to detect fatigue and drowsiness. They tested the performance of their system during the day and night, with drivers wearing glasses and sunglasses as well. However, such a performance is achieved

only if the size of the processing window is long enough. Their system needs between 2 and 3 s and up to a maximum of 6 seconds for completing the analysis. This amount of time is independent of the processing power and merely based on strategies employed in their algorithm for drowsiness detection. We also believe that only one feature, namely the eye state, does not always provide a conclusive clue to drowsiness. This will be discussed further in the next sections. Adding another feature detection could also add to the processing time, which can deteriorate the overall processing time.

In this paper, based on studies performed by Beaudot (1994), Hérault and Durette (2007), Beaudot, Palagi, and Hérault (1993), and Benoit and Caplier (2010), a new motion analysis is presented. This analysis is performed in two stages; During the first stage pre-processing is performed. The second stage uses the fast Fourier transform and orientation estimation to extract motion information. This stage is inspired by how a similar process is performed in the primary visual cortex. Even though the algorithm proposed in Benoit and Caplier (2010) is a reliable approach to this application, the large number of thresholds results in a highly intricate system. Using image processing operators -which model the behaviour of the human visual cells more precisely- the robustness of system will be enhanced. To follow the example of the HVS[1] better, we also apply a Gabor filter which helps the system to extract the features of pictures better. Finally better extracted features and a strong new decision making algorithm leads to improved drowsiness detection.

3. Pre-processing

In this section various issues regarding the pre-processing of images in the system will be presented. Pre-processing prepares the bases for the next step; feature extraction. Feature extraction is sensitive to many factors such as illumination variations, movement of subject, and noises. In the pre-processing stage we try to improve the resilience of the system against these variations.

The images received by camera have illumination variation. This means that some parts are darker than the others. Also, images might be partially or fully dark due to poor light condition. Therefore, there is a need to pre-process the images in order to extract necessary details from the dark parts of an image as well. To model this phenomenon, luminance variation elimination was proposed by Michaelis–Menten and used by others too, see for example, Benoit et al. (2010, more specifically Equations (1) and (2)).

The images captured by the camera have spatio-temporal noise too. This noise is generated due to reasons such as temperature variation and car movements. This noise is imposed on all pixels of each frame in an image sequence. To remove this noise, a spatio-temporal low pass filter is needed. This filter is implemented using two low pass spatio-temporal filters with different spatial and temporal frequencies (Beaudot, 1994, Equation (2.7); Hérault and Durette, 2007, Equation (2)). The same filter is used as the local luminance enhancement filter as well.

To enhance the static contours and edges, following the HVS, the outputs of two spatio-temporal low-pass filters described in Benoit et al. (2010, Equation (4)), are combined. Figure 2 shows the effect of this filter on a sample image. Similar to the corresponding counterpart in the HVS, this output is called Parvo and will be used in feature extraction.

Next stage of the pre-processing is motion information extraction. To extract motion, high-pass temporal filters are used, which are modelled by a first order filter (Benoit
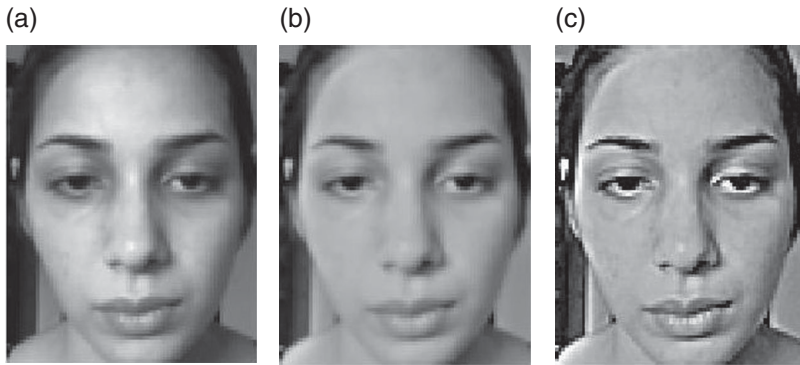
Figure 2. Parvo Channel (a) input, (b) output and (c) output of sharpening module.

et al., 2010, Equation (4)). Similar to the corresponding counterpart in the HVS, the output of the motion information extractor is called Magno. Output of the Magno channel is a sequence of images with a black background, and contours perpendicular to the motion direction. The output energy of this channel is used in the pre-processing stage, mainly to trigger initiation of further processes on the images. They are also used for head-state evaluation since it involves rather larger movements.

Pre-processed image sequences pass through the Magno and Parvo channels and enter to visual cortex V1. Visual cortex V1 processes the input signal in terms of frequency and orientation. Therefore, a bank of Gabor filters has been used in the algorithm to model the duties of V1 (Grigorescu, Petkov, and Kruizinga, 2002; Guyader, Massot, Hérault, and Chauvin, 2006; Le Meur, Le Callet, Barba, and Thoreau, 2006)

The Gabor filters transfer data from the Cartesian domain to the Frequency-Orientation domain. This transformation contributes to a decrease in the size of data to be processed. By selecting different frequencies distributed in different directions and by computing the energy of each filter, the energy spectrum versus frequency and orientation is obtained for an image. To improve the accuracy of the system, number of tune-on frequencies and orientations can be increased, but that would decrease speed of computation. Therefore, a compromise shall be made between speed and accuracy of Gabor filters.

The output of Magno channel is constituted of contours perpendicular to the direction of motion. After extracting moving contours, the bank of Gabor filters is used and the sum of energy in each direction is calculated. The maximum energy will be in the direction of the extracted contours. As a result, to achieve the motion direction, the angle calculated by Gabor filters is subtracted from $90°$.

By analysing the energy spectrum of each frame, frames in which a motion has happened can be recognised. When a motion happens, the contours and edges perpendicular to motion direction are extracted. However, when no motion happens, no edge and contour exists in the output of Magno channel. The higher the number of contours in an image, the higher the energy. To calculate the energy, Parseval's equation is used (Hazewinkel, 2001, Equation (1)).

Figure 3 shows an example of a driver closing her eyes for prolonged time-spans and the respective energy calculation. Figure 3(b) shows the motion energy associated with images

(a)

(b)

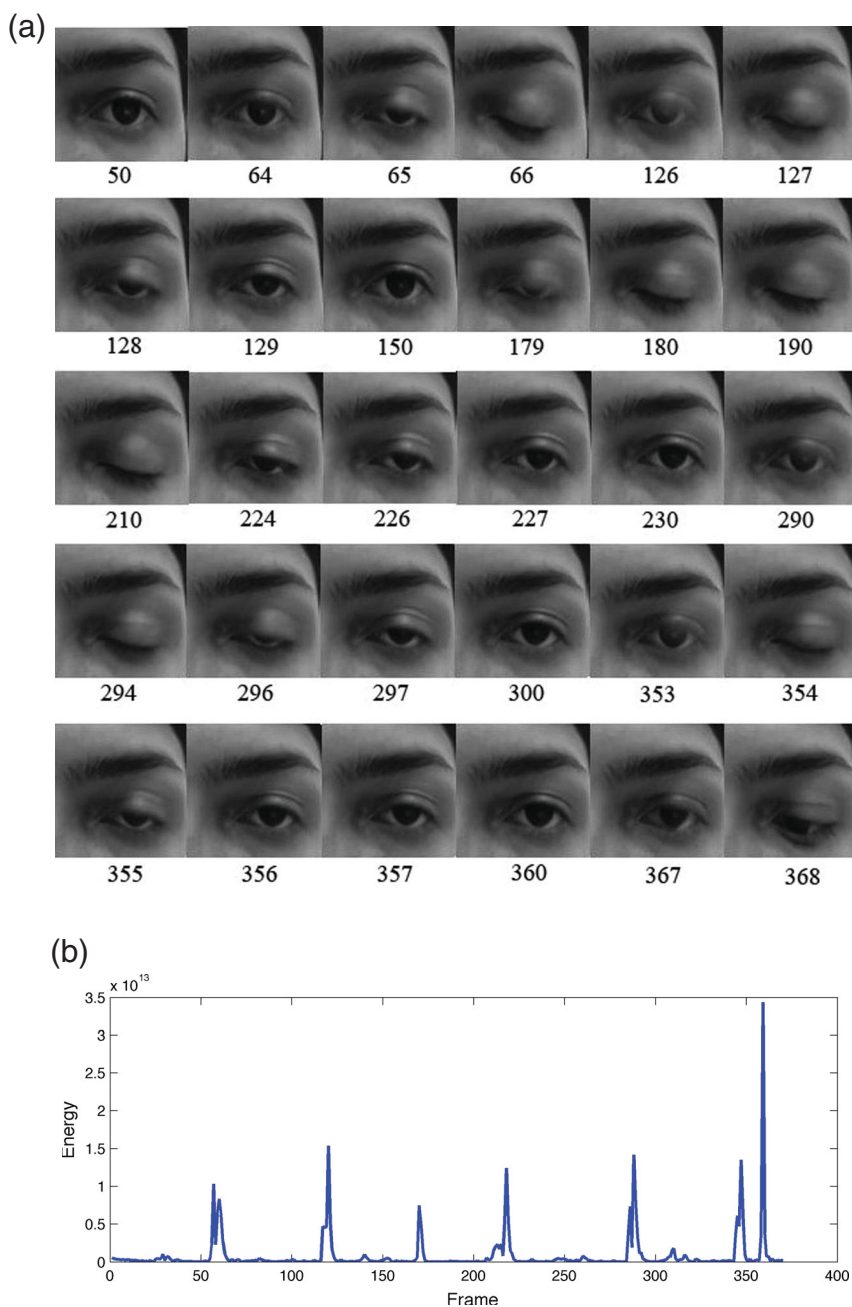**Figure 3.** Motion detection using energy spectrum; (a) image sequences of prolong closures of eyes, (b) energy spectrum versus frame.

in Figure 3(a) during frames 1–400. Every impulse observed in the latter figure is denoted as a prolonged closure of eye.

This example assumes that motions are obtained by capturing images from a fixed camera. However, it should be noted that several other movements can affect the eye motion

detection in the image sequence. Some examples of such movements could be head move-
ment, flickering of the scene or observed object, and other similar movements. Therefore, to
avoid false detection, it is important to properly set the parameters. This also improves the
processing speed by avoiding the unnecessary processing due to unnecessary stimulation
of the motion analyser.

4. Feature extraction and decision making

To detect drowsiness, the next step is extracting facial features and their status. For this
purpose, first the area of the face, eyes and the mouth should be found, which could be
done using the Viola-Jones algorithm (Viola and Jones, 2004). To find the eyes and mouth
in a face, a fast eye-tracking algorithm should be applied to the image of the face .[2] Even
though these algorithms are written in C++ and are fast, they are not particularly robust
in detecting the facial features under every condition. Conditions like head rotation and a
blocked face or other facial features could lead to failure in a proper detection of a bounding
box for that feature. Performance of these algorithms consequently affects the performance
of the fatigue detection system as well.

   To detect the motion event, the approach presented in Benoit and Caplier (2010) uses
the energy of Magno in bounding boxes around the eyes and the mouth. If a motion hap-
pens, the system analyses the Parvo energy by obtaining the energy of the first frame after
the motion and comparing it with the Parvo energy of the frame before the motion. If the
result shows an increase, the eye or mouth is opening and if it shows a decrease, it denotes
a closing motion.

   Even though the Viola-Jones algorithm is a good algorithm, it may not work as desired
at all times. For example, a wrong bounding box detection may impose a wrong motion
detection on the system or cause the system to miss a motion in the next processes. This
will cause a Miss or a False detection of fatigue. Another problem is that the one-frame
distance before and after the motion frame may not provide enough time for the Parvo
energy to undergo large enough changes. Therefore, it may not reflect the state change,
which in turn makes the approach not reliable enough. As we will see, to address these
issues, in this work a new strategy for using Parvo and Magno energy will be proposed.

   Head state estimation is also a complex process because the head may have numerous
motions imposed by the situation of the road, car and the driver. To estimate the state of
the head, Benoit and Caplier (2010) consider two factors: Orientation of the motion and
direction of the motion. Orientation of the motion is detected by a Gabor filter bank and
direction of the motion by the velocity analyser (Torralba and Hérault, 1997). The velocity
analyser extracts the velocity vector for every pixel in the box of the face. If these two algo-
rithms, Gabor filter and velocity analyser, show the same result while the head is having a
dropping motion, the system detects a dropping motion on that frame.

   In Benoit and Caplier (2010) a drowsiness event is detected if one of these three events
– closed eyes, dropped head or yawning – has been detected for one second. However, it
should be noted that a stand-alone yawn is not a conclusive sign for drowsiness. Since such
a false alarm could disturb and distract the driver, in the proposed decision tree we will
suggest a more comprehensive algorithm which uses a more robust combination of these
three features to reach a more reliable decision.

5. Proposed system

In the proposed algorithm, first, the calculated energy of Magno and Parvo video data is analysed using a new approach. In this approach, the calculations are enhanced with a sharpening module. Second, a new interpretation of these energies and combination of the results will be presented for every extracted bounding box. The new interpretation will lead to a more clear and better distinction of the states of features. Finally, we will present a new decision making algorithm which will result into a more robust drowsiness detection and less false alarms. We describe the details of our approach in the rest of this section.

5.1. Sharpening module

Ganglion cells, at the end of the Parvo channel in the HVS, have duties such as compression and contour enhancement. However, modelling the latter responsibility has been shown to be not particularly advantageous (Benoit and Caplier, 2010). Therefore, instead we add a sharpening module at the end of the Parvo channel to improve the static information. Figure 2(c) shows the effect of this module on the output of the Parvo channel. For this filter, a 3×3 mask contrast enhancement filter (Equation (1)), used in Jain, Kasturi, and Schunck (1995) and Park and Jeon (2014), was chosen as it offers high processing speed and effectiveness at the same time. In Equation (1), α is a control parameter in the range of 0 to 1, which determines the level of sharpening.

$$F_{\text{sharpening}} = \frac{1}{\alpha + 1} \begin{bmatrix} -\alpha & \alpha - 1 & -\alpha \\ \alpha - 1 & \alpha + 5 & \alpha - 1 \\ \alpha & \alpha - 1 & \alpha \end{bmatrix}. \tag{1}$$

Figure 4 illustrates the Parvo energy before and after applying the sharpening module. As it can be seen, the difference between the energy level in the closed and open states of the eye is considerably more distinct. This demonstrates the effectiveness of this module and its importance in decreasing error rate in detection of the status of facial features.

5.2. Reduction of variation dependencies

Magno energy is used to detect any motion event. To avoid unnecessary computational loads, higher levels of processing will start only when a motion is detected. A proper threshold on the energy level is hence crucial. However, the energy of the Magno channel depends also on the frame rate, that is, number of frames per second, image dimension, τ_A parameter and the appearance of the face (e.g. wrinkles and glasses lead to more contours and hence more energy). To alleviate this dependency, the proposed algorithm computes the relative Magno energy of each frame over the sum of the energy of the first second.

Similarly for Parvo energy also relative energy is calculated and used. Calculating relative energy makes the energy of motion independent from illumination variation as well as appearance of the driver's facial features and the frame rate. The assumption is that while the system starts up, the driver does not have significant motions of head and/or facial features during this first one second. In other words, the driver should allow a time span of one second for the system to set up its parameters. In return, this makes the system more robust.
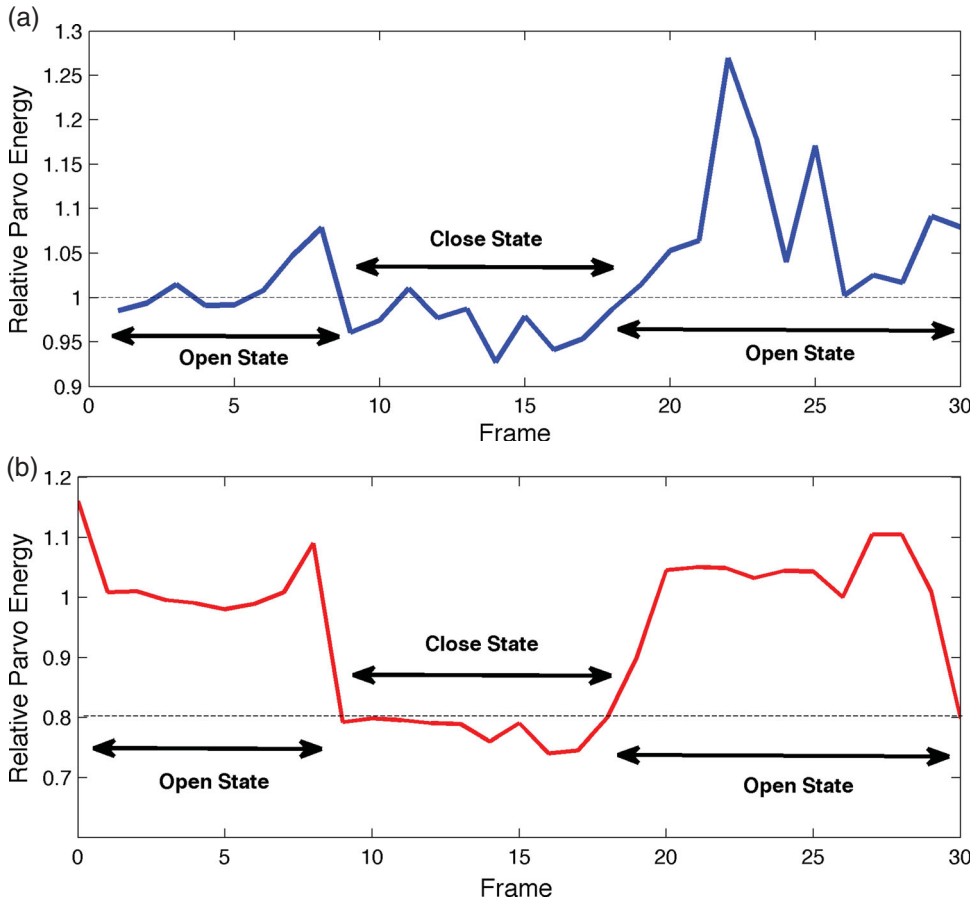
Figure 4. Effect of sharpening module on Parvo energy (a) Parvo energy of an eye sequence without a sharpening filter, (b) Parvo energy of the same sequence with a sharpening filter.

It should be noted that receiving the set-up information of the driver at the beginning of system start-up is a common practice. In other researches, for example (Lal, Craig, Boord, Kirkup, and Nguyen, 2003), the EEG and ECG information of individuals are collected first and then used by them for later processing. Nevertheless, we do consider a compromise; number of frames used to calculate the average energy needs to be high enough to provide reliable information but it should not be too high to take too much time and cause inconvenience for the driver. Therefore, we believe considering our experimental results, the short time span of one second is a proper compromise.

Although the proposed algorithm makes the system more robust, the possibility of faulty motion detection still exists. The system might find the facial features in two frames in such a way that the location of the eye or mouth in the bounding box are slightly different, without any considerable motion in reality. This will lead to detecting a wrong motion event, but our decision making algorithm prevents a wrong fatigue detection. Nonetheless, this error exerts an extra processing burden on the system. Therefore, a more robust and reliable algorithm for casting proper bounding boxes on facial features could be one future avenue for further improvement of processing speed of the system.
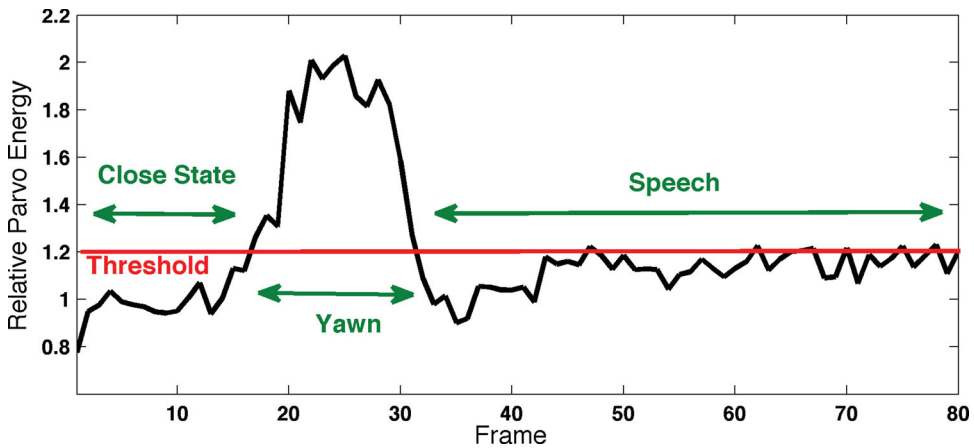
Figure 5. Parvo energy obtained in an image sequence of yawn and speech.

5.3. Mouth state estimation

Yawning could be a sign of potential drowsiness and hence its detection is very important. However, it causes a vertical motion of the head, which might be detected as a dropping state of the head. In order to prevent this false detection, before analysing the head's status, state of the mouth is estimated. Two states of yawning and speaking are considered for the mouth and thus analysed within the proposed approach.

Once the Magno channel motion detector recognises a motion, first the system analyses the state of the mouth. Therefore, the bounding box around the mouth is extracted and its relative Parvo energy is calculated. Figure 5 shows the Parvo energy spectrum of an 80-frame video where a significant difference in energy levels between yawning frames and other frames can be observed. In the closed state between frames 1 and 17, the level of energy is lower than that of the speaking state. Speaking happened during the frames 35–60, in which the level of energy in some frames is more than the threshold. By analysing the frequency of passing the threshold, the speaking span can be distinguished from other frames. This work focuses on yawn detection only and does not interpret the frequency of Parvo energy in speaking spans.

5.4. Eye state detection

Prolonged closure of eyes is a reliable characteristic for fatigue detection. The system suggested in this paper uses the analysis of relative Parvo energy to detect the state of eyes, which is calculated by Parseval's equation. The sharpening filter is applied afterwards, therefore, the distinction of Parvo energy is highly improved. Finally, a threshold separates the open state of the eyes from the closed one.

Figure 6 shows the calculated Parvo energy spectrum of an image sequence of an eye in open and closed states. As it can be seen, the level of relative energy is lower than the threshold in a few frames of the open state. On the other hand, the relative energy is beyond the threshold in a few frames of the closed-eye state as well. This irregularity can be due to several reasons: for example the movement of pupils, inaccurate detection of
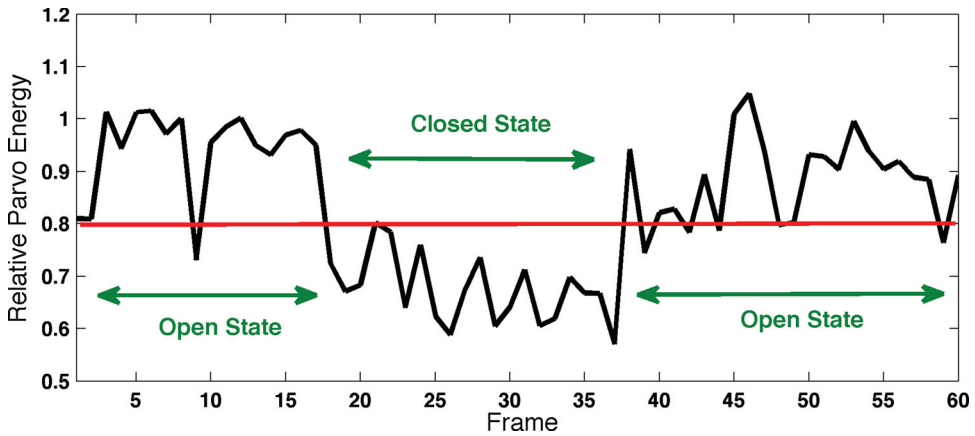
Figure 6. Parvo energy and separation of eye states using a threshold.

facial bounding boxes, and increase or decrease in the number of contours in some frames could lead to such irregularity. Therefore, to avoid a false detection, state of the eyes is considered as open in a span of 2 s, if more than 70% of frames during that time span have higher energy than the threshold. The same method is used for detecting the closed state of the eye.

5.5. Head state detection

The output of the Magno channel has appropriate information for the head state detection. The reason is that the motion of the head is strong enough to be detected by the motion indicators described. In the head state analysis, dropping the head to see something lower than the camera may cause the system to detect it as fatigue and a yawn may also be detected as a dropping state of head. Therefore, first state of the mouth is checked; the dropping state of head is ignored, if head dropping is detected in addition to yawning.

The algorithm starts with calculating the Magno energy of the box of face. The state of the head is analysed, if the relative Magno energy surpasses a certain threshold, that is, if a motion has happened. Then, the bank of Gabor filters is applied on the image of the face to estimate the dominant direction. The θ of Gabor filters is set to $18°$ to achieve a good trade-off between precision and speed. It should be kept in mind that the head does not necessarily drop only with an exact angle of $0°$; It may happen with $18°$, $162°$ or $180°$. Therefore, the head state is detected as dropping, if the dominant angle is either 0, 18, 162 or $180°$.

Having all that in mind, it should be noted that the mere detection of a head dropping cannot be considered as fatigue or drowsiness. To detect fatigue in a sequence, two factors need to be considered. The first factor is the detection of head dropping in a time span. Such detection is regarded as drowsiness, if the head has been dropping in all frames during 1.5s, or 80% of frames during 3s.

The second factor, which increases the precision of the system, is the state of the eyes. The eyes are normally closed when the head is dropping due to fatigue. When the head starts dropping, the eyes are most often closed, in other words, often eyes get closed first

and then the head moves down. Consequently, another condition added to the system obtained by experiment is whether the eyes have been closed within 0.66 s before the head starts dropping or not.

5.6. *Decision making algorithm*

As mentioned before, the proposed system uses three features to estimate the level of consciousness: head dropping, yawning and closed eyes. In the proposed decision algorithm, the interpretation is performed as following:

- If the eyes are closed for equal or more than 70% of a two second time span, the fatigue event is set to 1 and driver is recognised as being drowsy.
- As mentioned before, a stand-alone yawn does not necessarily imply drowsiness. However, combined with the other variables, a more conclusive decision on drowsiness may be made. Combined decision making helps the algorithm to become more robust. Hence, this event will be saved in the memory for further reference in decision making process. For example, if driver's head drops and a yawn had been detected once or more during the last 15s, the driver is recognised as being drowsy instantly.
- If the head is detected as dropping but eyes are not identified as closed or no yawn has happened during the last 15s, the head dropping is ignored. In such cases dropping the head is most likely due to reasons other than drowsiness. For example, driver may be momentarily distracted with looking at something lower than the dashboard.
- If the head is dropped or turned away from the road in 80% of the last three seconds, the fatigue variable is set to 1 and driver is recognised as being drowsy. Even though the head may not be necessarily falling because of fatigue, but not facing the road for 3 seconds or more is definitely dangerous. Therefore, the driver needs to be warned by the alarm system.
- If the closed eyes event or yawning have been saved in the memory of the system during 15 seconds before the head dropping, head dropping is regarded as a fatigue sign and driver is recognised as being drowsy.

To provide a better perspective and a summarised overview of the decision making process, the flow chart diagram of the algorithm is presented in Figure 7.

It should be noted that once a decision, either Drowsy or Sober, has been made, the system stops processing the current data. The system will afterwards move to the next (new) frame for processing.

6. Experiments

After elaborating on the functionality and details of the proposed algorithm, in this section experimental results will be presented. The results demonstrate the effectiveness and merit of the suggested algorithm over the majority of the other state-of-the-art algorithms in literature.

We first compare our system with the most similar method, proposed by Benoit and Caplier (2010). Afterwards, the result of the algorithm is compared with results reported for various other algorithms of driver's drowsiness detection, which do not necessarily have the
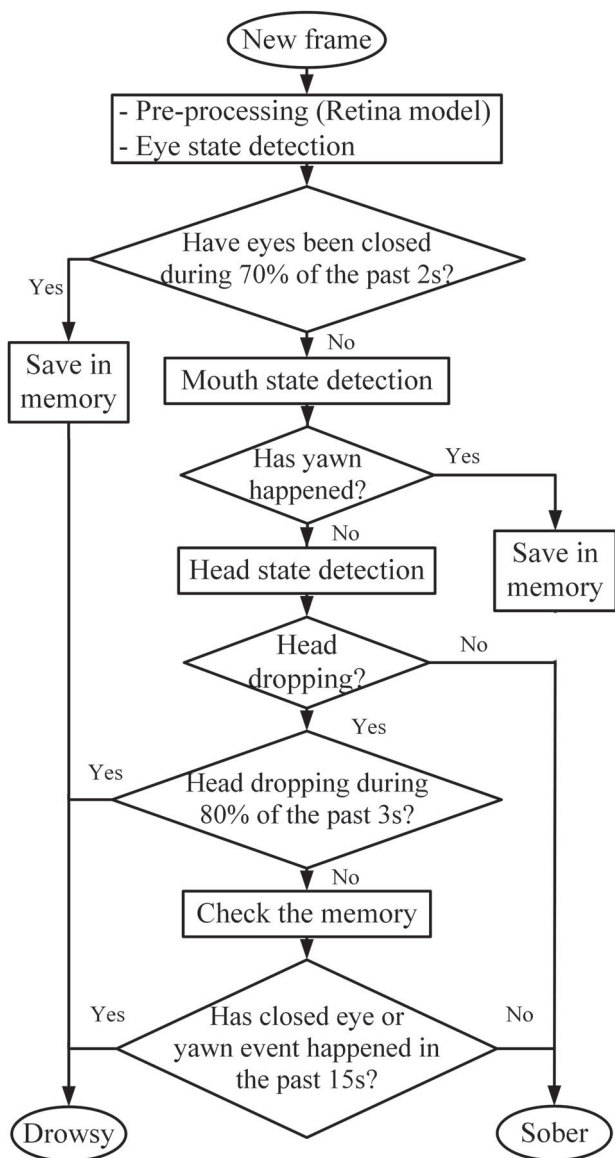
Figure 7. Flow chart diagram of decision making algorithm.

same approach as suggested here, that is, an HVS inspired image processing system. To provide a better perspective on different aspects of their methods compared to the proposed one, their approaches and respective advantages and disadvantages are discussed.

6.1. Set-up

The experiments and processes were run on a personal computer, namely a MacBook laptop, with an " Intel Core 2 Duo" 2.26 GHz processor, and 2 GB of RAM. Image sequences used in this work have been captured by the web-cam of the laptop. The facial feature

Table 1. Set-up values for experiments.

Unit	Parameter	Value	Parameter	Value
Illum. Var. Elimination[a]	V_0	0.5	V_{max}	255
Local Luminance[b]	α, β, τ	1		
Contour Enhancement[c]	$\alpha_h, \beta_h, \tau_h$	1	α_{ph}, β_{ph}	1
Magno[d]	τ_A	1	Δt	1
Sharpening Module (Equation (1))	α	0.2		
Gabor Filter[e]	# of Freq.	9	Range	[0.07,0.4]
	Angle Step	18°	Range	$[-\pi, \pi]$
	σ	10	ϕ	0
Thresholds	Parvo-Mouth	0.5	Parvo-Eye	0.8
	Magno	0.5		

[a] Benoit et al. (2010, Equations (1) and (2)).
[b] Beaudot (1994, Equation (2.7)).
[c] Benoit et al. (2010, Equation (3)).
[d] Benoit et al. (2010, Equation (4)).
[e] Grigorescu et al. (2002, Equation (2)).

Table 2. Results of experiments proving effectiveness of addition of the sharpening module.

	Parameter	Without	With
Sharpening module	Success (%)	81	**89**
	False (%)	15	**4**
	Miss (%)	**4**	7

Bold values indicate the best result between "Without" and "With" experiments.

location detection system receives the sequence from the web-cam and draws a bounding box around each facial feature. Due to lack of enough processing power in the laptop, the output sequence is reduced to a 4 frame per second (fps) image sequence, and each frame is 160 × 213 pixels in size. The system receives the images in RGB and then transfers them to gray scale. The experiments have been performed on a total of 55 min of video, in which 130 drowsiness events have occurred. Three subjects with various facial appearance (e.g. with or without glasses) were asked to participate in our experiments. The illumination of the recording set-up is different in each video to cover a rather thorough range of situations. However, complete darkness, that is, night situation were not simulated. The results described in the next sections are the average of all tests run on different subjects and under various light conditions. For simpler presentation, all result values are rounded to the closest integer number. The parameters used for the experiments are outlined in Table 1.

6.2. Results and preliminary comparison

Here, we present our results and compare it to the most similar algorithm, proposed by Benoit and Caplier (2010). To have the fairest comparison, the experiments have been performed under the same conditions. Once based on the method presented in Benoit and Caplier (2010) and another time based on the presented decision algorithm.

Table 2 shows the results of experiments before and after adding the sharpening filter to the system. In both experiments, the state of the eyes is estimated based on applying a threshold to the relative Parvo energy. The result shows that, as expected, success[3] rate

Table 3. Results of experiments and comparison with the best similar algorithm in the literature.

Experiment	Parameter	Benoit and Caplier (2010)	Proposed
Mouth state	Success (%)	83	**91**
	False (%)	**6**	6
	Miss (%)	11	**3**
Eye state	Success (%)	85	**89**
	False (%)	7	**4**
	Miss (%)	8	**7**
Head state	Success (%)	86	**88**
	False (%)	**4**	8
	Miss (%)	10	**4**
Overall performance	Success (%)	83	**90**
	False (%)	10	**4**
	Miss (%)	7	**6**

Bold values indicate the best results between "Proposed" and "Benoit & Caplier" algorithms.

has improved as well as false[4] detection rate. This shows the algorithm with the sharpening module to be more reliable. Miss[5] rate has slightly increased in the proposed algorithm. The reason is the smaller threshold selected for the Parvo energy. It should be noted that a higher threshold could decrease the miss rate however it would increase the false rate and degrade the success rate. Therefore, in a trade-off between all these parameters the threshold of 0.8 was experimentally obtained.

Table 3 shows the considerable increase in the success rate of yawn detection. Many extra contours are visible in the bounding box of the mouth while yawning, contours which do not exist in the closed mouth state. This increases the difference in the level of the Parvo energy between those two states. Miss rate also has a considerable improvement while false rate is remained the same. It should be noted that the false detection is mostly caused by large movements. In this situation, contours which do not belong to the mouth, appear in the bounding box and raise the level of energy. This consequently causes a false detection.

The result of the eye state detection algorithm shows the unanimous superiority of the proposed algorithm compared to the one proposed in Benoit and Caplier (2010). Specifically because adding the sharpening filter module to the system has led to a Parvo energy extraction with a higher distinction between different states.

The head motion is more complex than the two other features because it is possible that this movement could be triggered due to other events. Also, when the head is dropping, the face detection system is not able to detect the location of the face in the image. Therefore, the system cannot process the images to detect the dropping state of the head. Despite all these complications, as Table 3 shows, the proposed system has a high success rate of 88% which is higher than the system proposed by Benoit and Caplier (2010). Even though the False rate has a slight increase compared to their system, the Miss rate has a larger decrease in comparison to their Miss rate which shows a bigger overall improvement. Therefore, due to the aforementioned complications, the overall improvement in head-state detection, is smaller compared to the other proposed method.

Finally, the overall performance of the suggested system compared to the system proposed by Benoit and Caplier (2010), as shown in Table 3, has unanimously improved. This leads to a successful detection of more drowsiness events. The algorithm presented here,

in comparison to the system proposed by Benoit and Caplier (2010), employs a better combination of the three extracted features for an improved decision making. As the results show, the algorithm is able to successfully detect 90% of drowsiness events with only 4% of false detection and missing only 6% of fatigue events.

6.3. Further comparison and discussion

To have a deeper understanding about effectiveness, advantages and disadvantages of the proposed algorithm, we have compared it to other methods of driver's drowsiness detection. In contrast to the proposed method and the system proposed by Benoit and Caplier (2010), these works are not necessarily inspired by the HVS or do not necessarily follow a similar approach. Summary of these comparisons are compiled in Table 4. It can be observed that the proposed algorithm outperforms the majority of the other algorithms.

There are a few points which need to be considered in comparison: the features used in each method, light conditions, facial appearance, and required (used) equipment. A summary of these information is presented in Table 5.

All but one method in the comparison table have used facial features and image processing techniques to detect fatigue. Exception is the Picot et al. (2012) which used additional data such as EEG. As mentioned before, capturing these signals needs sensors, wired to the driver's body, which is inconvenient and impractical. The proposed algorithm shows a better performance in comparison to Zhang et al. (2012), Akrout and Mahdi (2013), Hemadri

Table 4. Comparison of different driver's drowsiness detection algorithms in the literature.

Algorithm	Miss (%)	False (%)	Success (%)
This work	6	4	90
Benoit and Caplier (2010)	7	10	83
Zhang, Cheng, and Lin (2012)	9	5	86
Picot et al. (2012)	14	5	81
Akrout and Mahdi (2013)	N/A	N/A	86
Hemadri and Kulkarni (2013)	N/A	N/A	80
Azim et al. (2014)	0	0	100
Jo et al. (2014)	0	1	99
Seeing-Machines-Technology (2015)	N/A	N/A	70

Table 5. Test set-up of various driver 's drowsiness detection algorithms in the literature.

Algorithm	Light	Facial appearance	Special gadget
This Work	All	w/ & w/o glasses	–
Benoit and Caplier (2010)	All	w/ & w/o glasses	–
Zhang et al. (2012)	All	N/A	–
Picot et al. (2012)	N/A	N/A	EEG sensor
Akrout and Mahdi (2013)	All	N/A	–
Hemadri and Kulkarni (2013)	All	N/A	–
Azim et al. (2014)	All	w/o glasses	NIR camera
Jo et al. (2014)	All	w/ & w/o glasses	NIR camera
Seeing-Machines-Technology (2015)	All	w/ & w/o glasses	–

Table 6. The ranking of the top three designs with regard to each parameters.

Algorithm	Special requirements	Agility	Success rate	Tests thoroughness
This work	**1**	**1**	3	**1**
Azim et al. (2014)	2	N/A	**1**	3
Jo et al. (2014)	2	2	2	**1**

Bold values indicate the best value of each parameter among mentioned algorithms.

and Kulkarni (2013), and Seeing-Machines-Technology (2015). Among the methods in comparison table, Azim et al. (2014) and Jo et al. (2014) achieved a better result compared to our method. In the following we briefly compare these two works with our proposed method.

- The approach by Azim et al. (2014) extracts the state of the mouth and the eyes and then translates it to the level of driver's drowsiness. They tested their approach and achieved a success rate of 100% when drivers do not wear glasses. Hence, even though the system has a performance of 100% on their test video, it is limited to a smaller set of conditions. For example, it is not clear how it would perform when the driver wears glasses or drops his head. The eye state is also detected based on infrared illumination, which has a lower performance during the day.
- Jo et al. (2014) also obtained a high success rate. The advantage of their system is fatigue detection under various light and facial appearances. However, based on their technique, the time they need to trigger a drowsiness alarm is normally 3 s and at best 2 s. Their system may have an overall delay, up to a maximum of 6 s before it detects a fatigue and drowsiness event. It should be mentioned that this delay is not due to processing power and time but rather the strategic reasons and decision making procedure used by their system. Since this long delay is more than enough to lead to an accident, it could certainly impose serious dangers to the driver's safety. This short-coming can render this algorithm impractical for application in real scenarios. Our system on the other hand, is able to raise an alarm flag normally in 2 s and maximally in 3 s. This makes it a more suitable and reliable option for real world scenarios.

Therefore, even though the proposed system does not match the reported performances of Azim et al. (2014) and Jo et al. (2014), considering its lower cost (no special gadget is required), faster response (maximum 3 seconds delay) and wider range of experimental tests performed on it, it can be still considered more favourable and reliable in comparison to the other methods. To provide an overview of the comparison between these top three works, we have summarised the main critical comparison parameters and respective ranking of each work regarding that parameter in Table 6.

7. Conclusion

Driving in an unfit condition leads to hundreds of thousands of deaths and injuries every year around the world. More specifically, drowsy drivers are one of the major causes of these incidents. Hence, it is very important to develop robust algorithms to detect such cases and prevent any incidents by providing an in-time alert.

In this paper, we first reviewed the concept of driver's drowsiness detection algorithms and the state-of-the-art literature. Afterwards, a new robust approach inspired by the HVS

was presented. In the proposed algorithm, new methods to estimate the state of the mouth, eyes and head, have been introduced to help the detection of driver's drowsiness. After extracting these three features from every frame, a new decision algorithm based on the extracted features determines whether the driver is drowsy or not.

To verify the functionality and performance of the proposed system, a series of experiments were run. These experiments proved the effectiveness and robustness of the suggested approach with 90% success rate in detecting drowsiness. The experiments show that the new proposed algorithm is able to reach to a higher success rate in different light conditions as well as in analysing people with different appearances.

To ensure merit of the suggested system in comparison to the similar state-of-the-art algorithms, both the proposed algorithm and the system suggested by Benoit and Caplier (2010) were simulated and run under the same condition. Results of experiments confirmed the superiority of the proposed system over the previous work in all three figures of merits; Success rate, false detection rate and rate of missing a detection.

Finally, the proposed work was compared to eight other methods of driver's drowsiness detection appeared in the literature during recent years, including a commercial product. A discussion about the respective advantages and disadvantages, showed that the suggested algorithm is very reliable and overall, favourable.

Notes

1. For the sake of brevity and simplicity, the reader is referred to Kandel, Schwartz, and Jessell (2000) and Benoit, Caplier, Durette, and Herault (2010) for more details on the physiology of the HVS. How their models and functionality are implemented in the proposed system however, will be discussed within this paper.
2. One of these algorithms is written by Aldrian and Meier (2009) and another example can be found on OpenCV (2014) which provides some other samples of face detecting algorithms as well.
3. Proper detection of the closed-eye state.
4. Detecting a closed-eye whereas it has not actually happened.
5. Not detecting a closed eye whereas it has actually happened.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Akrout B., & Mahdi W. (2013). Vision based approach for driver drowsiness detection based on 3D head orientation. In J. J. J. H. Park, J. K.-Y. Ng, H.-Y. Jeong, & B. Waluyo (Eds.), *Multimedia and ubiquitous engineering* (Vol. 240, pp. 43–50). Netherlands: Springer.

Aldrian P., P A., & Meier. U. (2009). *Extract feature points from faces to track eye's movement* [Computer software manual]. Retrieved from http://www.mathworks.com/matlabcentral/fileexchange/25056-fast-eyetracking.

Azim T., Jaffar M. A., & Mirza A. M. (2014). Fully automated real time fatigue detection of drivers through fuzzy expert systems. *Applied Soft Computing*, *18*, 25–38.

Beaudot W. (1994). *The neural information processing in the vertebrate retina: A melting pot of ideas for artificial vision* (Unpublished doctoral dissertation). INPG, France.

Beaudot W., Palagi P., & Hérault J. (1993). Realistic simulation tool for early visual processing including space, time and colour data. In J. Mira, J. Cabestany, & A. Prieto (Eds.), *New trends in neural computation* (Vol. 686, pp. 370–375). Berlin, Heidelberg: Springer.

Benoit A., & Caplier A. (2010). Fusing bio-inspired vision data for simplified high level scene inter-pretation: Application to face motion analysis. *Computer Vision and Image Understanding*, *114*(7), 774–789.

Benoit A., Caplier A., Durette B., & Herault J. (2010). Using human visual system modeling for bio-inspired low level image processing. *Computer Vision and Image Understanding*, *114*(7), 758–773.

Delmas, P., Eveno, N., & Lievin, M. (2002). Towards robust lip tracking. In Pattern Recognition, 2002. Proceedings. 16th International Conference on (Vol. 2, pp. 528–531). Quebec City, QC: IEEE.

European Transport Safety Council (2001). *The role of driver fatigue in commercial road transport crashes*. European Transport Safety Council. Retrieved from http://books.google.ca/books?id = xlr7OgAACAAJ.

Grigorescu S. E., Petkov N., & Kruizinga P. (2002). Comparison of texture features based on Gabor filters. *IEEE Transactions on Image Processing*, *11*(10), 1160–1167.

Guyader, N., Chauvin, A., Massot, C., Hérault, J., & Marendaz, C. (2006). A biological model of low-level vision suitable for image analysis and cognitive visual perception. *Perception* (ECVP Abstract), *35*, 1–252, doi:10.1177/03010066060350S101.

Hazewinkel M. (2001). *Parseval equality*. Encyclopedia of Mathematics, Springer, ISBN 978-1-55608-010-4.

Hemadri V., & Kulkarni U. (2013). Detection of drowsiness using fusion of yawning and eyelid move-ments. In S. Unnikrishnan, S. Surve, & D. Bhoir (Eds.), *Advances in computing, communication, and control* (Vol. 361, pp. 583–594). Berlin, Heidelberg: Springer.

Hérault J., & Durette B. (2007). Modeling visual perception for image processing. In F. Sandoval, A. Pri-eto, J. Cabestany, & M. Graça (Eds.), *Computational and ambient intelligence* (Vol 4507, pp. 662–675). Berlin, Heidelberg: Springer.

Jain R., Kasturi R., & Schunck B. G. (1995). *Machine vision*. New York, NY: McGraw-Hill.

Jo J., Lee S. J., Park K. R., Kim I.-J., & Kim J. (2014). Detecting driver drowsiness using feature-level fusion and user-specific classification. *Expert Systems with Applications*, *41*(4, Part 1), 1139–1152.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (Eds.). (2000). *Principles of neural science* (Vol. 4, pp. 1227–1246). New York: McGraw-Hill.

Lal S. K., Craig A., Boord P., Kirkup L., & Nguyen H. (2003). Development of an algorithm for an EEG-based driver fatigue countermeasure. *Journal of Safety Research*, *34*(3), 321–328.

Le Meur O., Le Callet P., Barba D., & Thoreau D. (2006). A coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(5), 802–817.

OpenCV (2014, August). Retrieved from http://opencv.org/opencv-3-0-alpha.html.

Park H. D., & Jeon J. W. (2014, February). *FPGA design and implementation of edge enhancement by using 3 × 3 mask filter*. IEEE International Conference on Industrial Technology (ICIT), 2014, pp. 630–635.

Picot A., Charbonnier S., & Caplier A. (2012, May). On-line detection of drowsiness using brain and visual information. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, *42*(3), 764–775. doi:10.1109/TSMCA.2011.2164242.

Picot A., Charbonnier S., Caplier A., & Vu N.-S. (2012). Using retina modelling to characterize blinking: Comparison between EOG and video analysis. *Machine Vision and Applications*, *23*(6), 1195–1208.

Rahman, Z. U., Jobson, D. J., Woodell, G. A., & Hines, G. D. (2005, August). Image enhancement, image quality, and noise, photonic devices and algorithms for computing VII, *Proc. SPIE.* (Vol. 5907, pp. 164–178). San Diego, CA:Society of Photo Optical. doi:10.1117/12.619460

Royal D. (2003, 04). *Volume I: Findings, national survey of distracted and drowsy driving attitudes and behaviors: 2002* (Tech. Rep.). U.S. Department of Transportation, National Highway Traffic Safety Administration (NHTSA).

Seeing-Machines-Technology (2015, February). Retrieved from http://www.seeingmachines.com/technology/.

Senane H., Saadane A., & Barba D. (2001). Design and evaluation of an entirely psychovisual-based coding scheme. *Journal of Visual Communication and Image Representation*, *12*(4), 401–421.

The Royal Society for the Prevention of Accidents (2001, 2002). *Driver fatigue and road accidents, a literature review and position paper* (Tech. Rep.). The Royal Society for the Prevention of Accidents.

Tian Y.-L., Kanade T., & Cohn J. (2000, January). *Robust lip tracking by combining shape, color and motion*. Proceedings of the 4th Asian conference on computer vision (ACCV'00).

Torralba, A., & Hérault, J. (1997). From retinal circuits to motion processing: a neuromorphic approach to velocity estimation. In *European symposium on artificial neural networks* (pp. 47–54). Bruges: D-Facto public, ISBN 2-9600049-7-3, April 1997.

Viola P., & Jones M. J. (2004, May). Robust real-time face detection. *International Journal of Computer Vision*, *57*(2), 137–154.

Wu J.-D., & Chen T.-R. (2008, February). Development of a drowsiness warning system based on the fuzzy logic images analysis. *Expert Systems With Applications*, *34*(2), 1556–1561.

Zhang W., Cheng B., & Lin Y. (2012, June). Driver drowsiness recognition based on computer vision technology. *Tsinghua Science and Technology*, *17*(3), 354–362. doi:10.1109/TST.2012.6216768.