

Video surveillance for monitoring driver's fatigue and distraction

R. Jiménez-Moreno ^{a,b}, S. A. Orjuela ^{b,d}, P. Van Hese ^d, F. A. Prieto ^c, V. H. Grisales ^c and W. Philips ^d

^a Universidad Militar Nueva Granada, Bogotá, Colombia;

^b Universidad Antonio Nariño, Bogotá, Colombia;

^c Universidad Nacional, Bogotá, Colombia;

^d Department of Telecommunications and Information Processing (TELIN), IPI, IBBT, Ghent University, Ghent, Belgium.

ABSTRACT

Fatigue and distraction effects in drivers represent a great risk for road safety. For both types of driver behavior problems, image analysis of eyes, mouth and head movements gives valuable information. We present in this paper a system for monitoring fatigue and distraction in drivers by evaluating their performance using image processing. We extract visual features related to nod, yawn, eye closure and opening, and mouth movements to detect fatigue as well as to identify diversion of attention from the road. We achieve an average of 98.3% and 98.8% in terms of sensitivity and specificity for detection of driver's fatigue, and 97.3% and 99.2% for detection of driver's distraction when evaluating four video sequences with different drivers.

Keywords: Driving surveillance, driver fatigue, driver distraction, driver drowsiness, driver inattention, driver natural impairment

1. INTRODUCTION

Every day, thousands of people die, are killed or get injured around the world due to traffic accidents.¹ Therefore, considerable research in search of tools for safety driving is constantly conducted.² Most motor vehicle crashes are caused by driver error (e.g. speeding) or poor operating practices including lack of seat belt use, distractions, fatigue, aggressive driving, and alcohol or drug use while driving.³ All these cases evidence problematic human behaviour i.e. due to personalities, social factors or deterrence.⁴ Therefore, human behaviour factors affecting driver performance are a main topic to consider when implementing reliable systems to assure safety driving.⁵

Fatigue and distraction signals can be detected by monitoring visual information related to the driver performance such as facial expressions.⁶⁻¹¹ Fatigue can be considered an internal distractor because it diminishes the alertness of the driver.¹² Fatigue can be monitored by measuring facial expressions such as percent eye closure, eye closure duration, blink frequency, nodding frequency, face pose, fixed gaze and yawning.^{8,13} Similarly, distraction from the visual focus can be identified by tracking the eyes movement.¹⁴⁻¹⁶ Additionally, an important driver action to monitor concerning to distraction is the movement of driver's head.^{17,18} Monitoring systems for distraction have been designed evaluating together head and eyes movements.¹⁹ Therefore, the recognition of facial expressions and head movements using image features can be used to automate preventive alarms.

Further author information: (Send correspondence to R. Jiménez-Moreno)

Universidad Militar Nueva Granada, Carrera 11 No. 101-80, Bogotá, Colombia: E-mail: robinson.jimenez@unimilitar.edu.co, telephone +5716343200

Fatigue and distraction alert systems require real time algorithms.^{24,25} Research has been conducted using information independently related to eyes²⁰ or mouth behaviour^{13,21} as well as a combination of both.²² Research reports that eyes and mouth movements have a correlation of more than 50%.²³

This paper presents a monitoring system for fatigue and distraction, which are considered problematic natural behaviours that produce driving impairment.⁹ On the one hand, fatigue reduces the capacity to pay attention to the driving tasks.²⁶ On the other hand, driving performance can be degraded by cognitive or visual distraction, where visual distraction poses the major risk.²⁷ The dimensions of fatal crashes due to these effects can be compared with other driver behaviour related problems including drinking and speeding.²⁸⁻³⁰

We present a system, in which visual features related to nod, yawn, eye closure and opening, and mouth movements are used to detect fatigue as well as to identify diversion of attention from the road. Mouth movement is used to identify effects of fatigue on speech. The combination of these features gives robustness to the system in preventing car accidents due to human factors. The set of visual features used in this approach are extracted by using histogram analysis and thresholding on a set of features based on a boosted cascade of simple features.³¹ We use these features to measure eye and mouth opening, magnitude of nod and time of diversion from the road. To extract these features we first identify the zone of the head. Then we use a window around it, processing only this region to speed up the algorithm. We convert images between different color spaces to enhance the visual characteristics, which facilitate the extraction of the features.

An automatic alert is activated with a multilayer neural network, which uses the features as input. The use of a neural network permits to adapt the system to the driver by conducting a training step before driving to recognize facial expressions related to fatigue and distraction of the driver. This allows the system to adapt to any driver. The degree of the alert depends of the intensity of the signals given by the features. The system is designed using open source to allow adapting it according to the needs of users.

The paper is organized as follows. In Section 2 we discuss the details regarding the image processing. In Section 3 we discuss how to identify patterns related to fatigue and distraction. In Section 4 we report the results and discuss the findings. Finally, in Sections 5 conclusions are drawn.

2. IMAGE PROCESSING

Drivers fatigue and distraction are determined by analysing image features extracted from the frames of a video. Image analysis on small areas of the original scene allows fast processing. Therefore, we detect the regions corresponding to areas surrounding the eyes and mouth of the driver. For this, we use an algorithm based on a boosted cascade of simple features to identify the region related to the face of the driver.³¹ This algorithm involves the sums of image pixels within rectangular areas, which bear some similarity to Haar basis functions. Features are extracted based on strong intensity changes in the image detected with the use of the rectangles. An iterative process based on sequentially evaluating similar features on the frames of the video yield to differentiate objects from the background scene. Objects are finally identified by the use of a threshold.

The algorithm of Viola and Jones is first used to detect the bounding box of the face of the driver. Because the camera is located behind the steering wheel, the vision scene covers part of the rear seat. Therefore, other faces apart from the face of the driver can be detected using the algorithm of Viola and Jones. However, the face of the driver is identified as the one with the largest size. An example of this is shown in Figure 1. Once the face of the driver is recognized, the algorithm of Viola and Jones is again applied, but now only on the bounding box of the face, to detect bounding boxes for the eyes and mouth of the driver. The three centres of gravity of the three rectangular areas are connected and describe a triangle, which is used to detect head movements. The rectangle areas related to eyes and mouth as well as the connecting triangle are shown in Figure 2a.

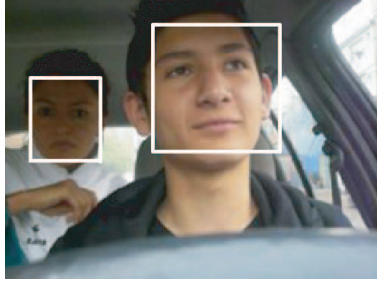


Figure 1: Detection of faces within bounding boxes using the algorithm of Viola and Jones.³¹ In case of detecting both, the driver and the person sitting in the back seat, the bounding box with the largest size is selected.

The next step consists of detecting the contours of eyes and mouth to monitor their opening and closure. Contours in the eyes can be extracted using the RGB color space.³² However, this model is not appropriate for recognizing the contour of lips, which are better recognized using the Cr component of the YCbCr colour representation.^{33,34} Therefore, the rectangular areas corresponding to eyes and mouth are processed using different color representations, RGB for eyes and YCbCr for lips. Figure 2b shows the results of the contours detection of lips and eyes of the driver.

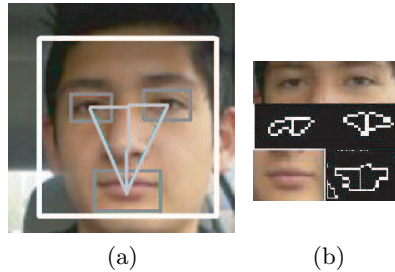


Figure 2: Detection of driver's eyes and mouth. (a) The algorithm of Viola and Jones is used to detect bounding boxes, corresponding to eyes and mouth, within the area of the driver's face. The three centres of gravity of the three rectangular areas are connected, describing a triangle. (b) The contours of eyes and lips are detected within the bounding boxes using different color representation spaces (RGB for eyes and YCbCr for mouth).

The middle point, denoted by $p^H(x, y)$, between the centre of gravity of both eyes defines the position of the head. The initial position, denoted by $p_i^H(x, y)$, is computed as the average value among the first 30 frames. When the driver moves the head, the final position is denoted by $p_f^H(x, y)$. The movement of the head is quantified as the vector, of magnitude A and angle θ , representing the distance between $p_i^H(x, y)$ and $p_f^H(x, y)$. The displacement vector is graphically illustrated in Figure 3. Both points are drawn above their real positions to facilitate the visualization of the vector while drawing also the bounding boxes and their connections.

We detect two points in the rectangular frame of each eye to quantify the level of opening/closure of eyes. We denote x_s the central-higher point on the contour of an eye and x_i the central-lower point on the contour of the eye. The location of these points are graphically illustrated in Figure 4b and Figure 4a.

The eye opening is quantified as the absolute value of the difference between these points.¹⁰ The signal, related to the Eyes Opening (EO), used for the alarm is computed as the average between this measurement computed on both, Right (R) and Left (L), eyes as follows:

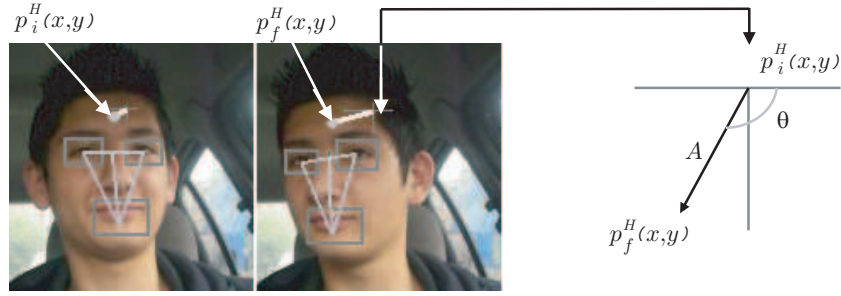


Figure 3: Measure of vector displacement to quantify head movement. The initial position of the head is denoted by $p_i^H(x, y)$ and the final position by $p_f^H(x, y)$. The movement of the head is characterized with a vector representing the distance between these two points.

$$EO = \frac{|x_s^R - x_i^R| + |x_s^L - x_i^L|}{2}, \quad (1)$$

with x_{Rs} , x_{Ri} , x_{Ls} and x_{Li} correspondingly being the points for the right and left eyes. Points on right and left eyes are identified with superscripts.

The gestures related to the mouth are similarly quantified with two measures, one in the horizontal axis and the other one in the vertical axis. For this, four points, instead of two, are extracted from the rectangle corresponding to the mouth surrounding. The two measures are computed as a set, denoted by MO , as follows:

$$MO = \{|x_a - x_b|, |y_a - y_b|\}, \quad (2)$$

where x_a and x_b are the two points representing the maximum distance of the lips contours in the horizontal axis, and y_a and y_b are the two points representing the maximum distance of the lips contours in the vertical axis.

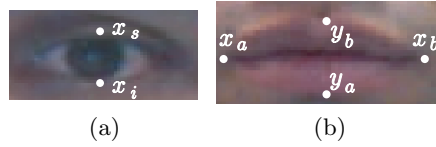


Figure 4: Extracting measures from driver's eyes and mouth. (a) Two points are calculated using the eyes contours. (b) Four points are calculated using the mouth contours.

3. PATTERN ANALYSIS: IDENTIFICATION OF FATIGUE AND/OR DISTRACTION

We design a system that identifies signals of driver's fatigue or distraction using a neural network based on multilayer perceptrons. The designed system gives three types of signals in the output, namely normal, warning and critical. The system was developed using four video sequences, each with a different driver. In the following we discuss how the system was designed.

Driver's fatigue is detected by monitoring a driver's nodding, yawing and closure of the eyes. These actions are identified by using the measures discussed in Section 2. The combination of parameters $\{EO, MO, A, \theta\}$ associated to recognize these actions vary from individual to individual.³⁴ E.g. in the case of yawing, the

separation between the contours of the lips in the horizontal axis may reduce more than 60% compared to the case when the mouth is closed. Similarly, in the vertical axis the separation of the lips' contours may increase more than 70%. Therefore, we propose to use a neural network, which is capable of adapting to the facial expressions of a driver to recognize driver's nodding, yawing and closure of the eyes.

The multilayer perceptron is composed of four neurones in the input layer (eye closure, mouth opening, and magnitude and angle of head movement) and three neurons in the output layer (normal, warning and critical). The design of the system based on neural networks is performed in three stages namely training, validation and testing.³⁵ Therefore each video sequence is divided into three parts with major proportion for the data used in the training stage. To train the neural network, we assume that the distance from the driver's head and the camera is constant while driving. A ground truth is manually created based on Table 1 using 60% of the video sequence.

Table 1: Scores used by humans to train the neural network for detecting fatigue based on nodding, yawing and blinking of the driver.

INPUTS				OUTPUT
EYES CLOSURE	YAWING	NODDING	FATIGUE	
NO	NO	NO	NO	
NO	NO	YES	NO	
NO	YES	NO	NO	
NO	YES	YES	WARNING	
YES	NO	NO	WARNING	
YES	NO	YES	CRITICAL	
YES	YES	NO	CRITICAL	
YES	YES	YES	CRITICAL	

A maximum number of iterations is first determined by using the training and validation sets, and one thousand of neurons in the hidden layer. The maximum number of iterations is identified when the error remains constant for the training set while increasing for the validation set. We found this at iteration 200. Then, a similar process is performed but now varying the number of neurons with increments of ten each time and keeping the number of iterations constant at 200. This give a number of 60 hidden neurons for our case.

To monitor driver's distraction, a neural network is trained based on Table 2 following the same procedures discussed in the case of the fatigue detection. We obtain a number of 92 iterations and 78 hidden neurons for the case of distraction detection. Figure 5 shows three examples of cases of warnings for distraction.

Table 2: Scores used by humans to train the neural network for determining levels of distraction.

INPUTS				OUTPUT
AMPLITUDE EJE X	MOUTH EJE Y	HEAD MOVEMENT		
		MAGNITUDE	ANGLE	
SMALL	SMALL	SMALL	SMALL	NO
SMALL	SMALL	MODERATE	SMALL	WARNING
MODERATE	SMALL	MODERATE	SMALL	WARNING
MODERATE	SMALL	MODERATE	MODERATE	WARNING
BIG	MODERATE	SMALL	SMALL	NO
BIG	MODERATE	MODERATE	MODERATE	WARNING
SMALL	MODERATE	MODERATE	MODERATE	CRITICAL
MODERATE	MODERATE	MODERATE	MODERATE	CRITICAL
BIG	SMALL	BIG	MODERATE	WARNING

In both cases (fatigue and distraction) a prediction of the status of the driver is estimated based on the last ten frames. For this, we denote $NN(t)$ as the response of the Neural Network (NN) in the actual frame. We denote Normal: $NN(t) = 0.0$, Warning: $NN(t) = 0.5$ and Critical: $NN(t) = 1.0$. The prediction is estimated as follows:

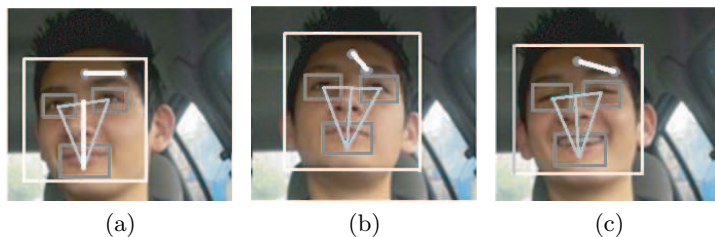


Figure 5: Examples of cases of detected warnings of distraction in the validation step.

$$Pred = \frac{1}{10} \sum_{i=1}^{10} (NN(t-i+1)/i), Decision = \begin{cases} \text{Normal,} & 0.00 < Pred < 0.33 \\ \text{Warning,} & 0.33 < Pred < 0.66 \\ \text{Critical,} & 0.66 < Pred < 1.00 \end{cases} \quad (3)$$

with $i = 1, \dots, 10$, the index used to identify the last ten frames.

4. RESULTS

Table 3: Detection of fatigue and distraction results using four video sequences of five minutes each.

Video	1	2	3	4	Video	1	2	3	4
TP	534	533	535	536	TP	412	351	602	516
FP	58	65	53	49	FP	35	35	44	46
TN	4315	4423	5702	4989	TN	4458	4634	5643	5001
FN	12	8	10	6	FN	14	9	11	17
Sensitivity	97,8	98,5	98,2	98,9	Sensitivity	96,7	97,5	98,2	96,8
Specificity	98,7	98,6	99,1	99,0	Specificity	99,2	99,3	99,2	99,1

The system was implemented using openCV 2.1 on a computer with 1.5 GB of RAM and a 1.7 Ghz processor. We use a database of four videos, each of five minutes, which are captured with a rate of 20 frames per second. We respectively use 60%, 20% and 20% of the database for training, validation and test stages. A ground truth was manually created.

The results show an average of 98.3% and 98.8% in terms of sensitivity and specificity for detection of driver’s fatigue, and 97.3% and 99.2% for detection of driver’s distraction. These findings are based on the true and positive results, which are listed in Table 3 for monitoring of both, fatigue and distraction, status on the test data sets corresponding to the four videos.

We found that many of the false positives were due to sun direct exposure on the driver’s face. Also contour detection was affected by shadows, given by objects on the road, i.e, buildings, trucks passing close or trees, and environmental illumination conditions. Also false positive were given when the driver looked to the driving mirrors, i.e, rear-view mirror, since this factor were not taken into account when designing the system. These factors must be considered for further developments.

5. CONCLUSION

We proposed an algorithm for real time detection of driver’s fatigue and/or distraction. We achieve an average of 98.3% and 98.8% in terms of sensitivity and specificity for detection of driver’s fatigue, and 97.3% and 99.2% for detection of driver’s distraction. Despite this, the algorithm is not robust enough since its performance varies with environmental conditions, i.e, illumination and obstacles on the road. Therefore, further research must be conducted to achieve a robust system.

REFERENCES

- [1] M. Pede. World report on road traffic injury prevention. Technical report, World Health Organization, 2004.
- [2] J. D. Lee. Fifty years of driving safety research. *The Journal of the Human Factors and Ergonomics Society*, 50(3):521–528, June 2008.
- [3] Utah. Dept. of Administrative Services. Utah defensive driving guide. Technical report, Risk Management; Administrative Services, 2009.
- [4] N. Jiang, Y. Chen, K. Shi, Y. Liu, Y. Xiao, B. Watson, J. Fleiter, M. King, and C. Chen. Big-five personality factors affecting driving behaviors. In *2d IEEE International Conference on Emergency Management and Management Sciences (ICEMMS)*, 2011.
- [5] E. Petridou and M. Moustaki. Human factors in the causation of road traffic crashes. *European Journal of Epidemiology*, 16(9):818–826, 2000.
- [6] N. Wijesuriya, Y. Tran, and A. Craig. The psychophysiological determinants of fatigue. *International Journal of Psychophysiology*, 63(1):77–86, 2007.
- [7] X. Fan, Y. Sun, and B. Yin. Multi-scale dynamic human fatigue detection with feature level fusion. In *8th IEEE International Conference on Automatic Face & Gesture Recognition*, 2008.
- [8] L. M. Bergasa, J. Nuevo, M. A. Sotelo, R. Barea, and E. López. Visual monitoring of driver inattention. In *Computational Intelligence in Automotive Applications*, 2008.
- [9] L. C. Barr, C. Y. D. Yang, R. J. Hanowski, and R. Olson. An assessment of driver drowsiness, distraction, and performance in a naturalistic setting. Technical report, Volpe National Transportation Systems Center, Federal Motor Carrier Safety Administration, 2011.
- [10] R. Jimenez, F. Prieto, and V. H. Grisales. Detection of the tiredness level of drivers using machine vision techniques. In *IEEE Electronics, Robotics and Automotive Mechanics Conference (CERMA)*, 2011.
- [11] H. Sun, Q. Zhu, W. Shen, Z. Leng, and H. Wu. Driver fatigue detection system based on the feature of face. *American Journal of Engineering and Technology Research*, 12(1):56–63, 2012.
- [12] A. Williamson. Fatigue and coping with driver distraction. In *Proc. of International Conference on the Distractions in Driving*, 2005.
- [13] X. Fan, B. Yin, and Y. Sun. Yawing detection for monitoring driver fatigue. In *Proc. of Sixth International Conference on Machine Learning and Cybernetics*, 2007.
- [14] and K. Matsunaga T. Hayami, Shidoji K., and Y. Matsuki. Detecting drowsiness while driving by measuring eye movement - a pilot study. In *The IEEE 5th International Conference on Intelligent Transportation Systems, 2002. Proceedings*, 2002.
- [15] R. St.Clair, M. Huff, and A. E. Seiffert. Conflicting motion information impairs multiple object tracking. *Journal of Vision*, 10(4):1–13, 2010.
- [16] H. M. Fehd and A. E. Seiffert. Looking at the center of the targets helps multiple object tracking. *Journal of Vision*, 10(4):1–13, 2010.
- [17] E. Murphy-Chutorian and M. M. Trivedi. Hyhope: Hybrid head orientation and position estimation for vision-based driver head tracking. In *2008 IEEE Intelligent Vehicles Symposium*, 2008.
- [18] B. Metz and H. P. Krueger. Measuring visual distraction in driving: The potential of head movement analysis. *IET Intelligent Transport Systems*, 4(4):289–297, 2010.
- [19] A. Doshi and M. M. Trivedi. On the roles of eye gaze and head dynamics in predicting driver’s intent to change lanes. *IEEE Transactions on Intelligent Transportation Systems*, 10(3):453–462, 2009.
- [20] J. Jo, S. J. Lee, H. G. Jung, K. R. Park, and J. Kim. Vision-based method for detecting driver drowsiness and distraction in driver monitoring system. *Optical Engineering*, 50(12):127202, 2011.

- [21] W. Rongben, G. Lie, T. Bingliang, and J. Lisheng. Monitoring mouth movement for driver fatigue or distraction with one camera. In *The 7th International IEEE Conference on Intelligent Transportation Systems, 2004. Proceedings*, 2004.
- [22] A. Azman, Q. Meng, E. A. Edirisinghe, and H. Azman. Non-intrusive physiological measurement for driver cognitive distraction detection: Eye and mouth movements. *International Journal of Advanced Computer Science*, 1(3):92–99, 2011.
- [23] A. Azman, Q. Meng, and E. Edirisinghe. Correlation between eye movements and mouth movements to detect driver cognitive distraction. In *Conference on Brain-Inspired Cognitive Systems*, 2010.
- [24] C. J. Chen, H. Y. Peng, B. F. Wu, and Y. H. Chen. A real-time driving assistance and surveillance system. *Journal of Information Science and Engineering*, 25(5):1501–1523, 2009.
- [25] F. Zhai, Z. Yang, and Y. Song. A detection model for driver’s unsafe states based on real-time face-vision. In *International Conference Image Analysis and Signal Processing (IASP)*, 2010.
- [26] K. A. Brookhuis, D. de Waard, J. H. Kraaij, and E. Bekiaris. *Human Factors and Ergonomics Society Europe ChapterBooks*, chapter How important is driver fatigue, and what can we do about it, pages 193–207. Shaker Publishing, 2003.
- [27] T. Titchener and I. Wong. Driver distractions: Characteristics underlying drivers’ risk perceptions. *Journal of Risk Research*, 13(6):771–780, 2010.
- [28] M. Rakauskas and N. Ward. Behavioral effects of driver distraction and alcohol impairment. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting September*, volume 49, pages 1912–1916, 2005.
- [29] D. L. Strayer, F. A. Drews, and D. J. Crouch. A comparison of the cell phone driver and the drunk driver. *The Journal of the Human Factors and Ergonomics Society*, 48(2):381–391, 2006.
- [30] A. Eskandarian and R. A. Sayed. Analysis of driver impairment, fatigue, and drowsiness and an unobtrusive vehicle-based detection scheme. In *Proc. of First International Conference on Traffic Accidents*, 2010.
- [31] Pand M. Jones P. Viola. Rapid object detection using a boosted cascade of simple features. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001.
- [32] D. D. Sidibe, P. Montesinos, and S. Janaqi. A simple and efficient eye detection method in color images. In *International Conference Image and Vision Computing New Zealand*, 2006.
- [33] J. Kadlec. Lip detection in low resolution images. In *10th Conference and Competition*, 2004.
- [34] J. B. Gómez, F. Prieto, and T. Redarce. Mouth segmentation in images: A review. *Revista Facultad Ingeniería Universidad Antioquia*, 47:103–116, 2009.
- [35] C. G. Looney. *Pattern Recognition Using Neural Networks: Theory and Algorithms for Engineers and Scientists*. Oxford University Press, 1997.