MECHATRONICS ENGINEERING

Driver distraction detection using machine vision techniques

INGENIERIA MECATRÓNICA

Detección de distracción en conductores mediante técnicas de visión de máquina

Robinson Jiménez Moreno*§, Oscar Avilés Sánchez*, Darío Amaya Hurtado*

**GAV research group, Mechatronics engineering, Universidad Militar Nueva Granada, Bogotá, Colombia*
§Robinson.jimenez, Oscar.aviles, Dario.amaya@unimilitar.edu.co

Recibido: Noviembre 26 de 2012- Aceptado: Noviembre 1 de 2013

Resumen

En este artículo se presenta un sistema para la detección de estados de distracción en conductores de vehículos en horas diurnas mediante técnicas de visión de máquina, el cual se basa en la segmentación de la imagen respecto a los ojos y la boca de una persona, vista de frente por una cámara. De dicha segmentación se establece los estados de movimiento que de la boca y de la cabeza, permiten inferir un estado de distracción. Las imágenes se extraen de videos de corta duración y con una resolución de 640x480 píxeles, sobre las cuales se emplean técnicas de procesamiento de imagen como transformación de espacios de color y análisis de histograma. La decisión del estado es el resultado de una combinación de las características extraídas ingresadas a una red neuronal del tipo perceptrón multicapa. El desempeño logrado en la detección en un ambiente controlado de pruebas es del 90% y del 86% en ambiente real, con un tiempo de respuesta promedio de 30 ms.

Palabras clave: *Distracción, procesamiento de imagen, red neuronal, visión de máquina.*

Abstract

This article presents a system for detecting states of distraction in drivers during daylight hours using machine vision techniques, which is based on the image segmentation of the eyes and mouth of a person, with a front-face-view camera. From said segmentation states of motion of the mouth and head are established, thus allowing to infer the corresponding state of distraction. Images are extracted from short videos with a resolution of 640x480 pixels and image processing techniques such as color space transformation and histogram analysis are applied. A decision concerning the state of the driver is the result from a multilayer perceptron-type neural network with all extracted features as inputs. Achieved performance is 90% for a controlled environment screening test and 86% in real environment, with an average response time of 30 ms.

Keywords: *Distraction, image processing, neural network, machine vision.*

1. Introduction

Drivers' distraction states are the source of a great number of accidents and dangerous situations, which expose the lives of drivers, passengers and pedestrians, according to a study from the National Highway Traffic Safety Administration (2009). Psychology has been performing several studies analyzing and establishing possible theoretical solutions to this problem, based in driver behavior and using invasive methods (Muhrer & Vollrath, 2011; Crundall & Underwood, 2011; Xue, 2012).

Through the use of image processing techniques, it is possible to develop an autonomous system, oriented towards providing the driver feedback concerning his concentration state, allowing detection of states which could create a dangerous situation.

Previous research had focused on the detection of dangerous driving conditions through the identification of certain states of the eyes (Coetzery & Hancke, 2009; Bajaj *et al*., 2010). Concerning distraction states, Zhai & Yang (2010) presented a study that relates the mouth movements when the driver is talking with someone, directly with the state of distraction. Efficiency in the aforementioned work is limited by factors such as occlusion of the eyes (e.g. If the driver wear glasses) and mouth movement, which should not be substantial when the driver is alone.

The present work proposes an automatic detection system of a driver's distraction states using computer vision techniques; this system takes into consideration at least three decisive factors in the distraction state determination, such as eye movement, head movement and mouth movement. Several parameters were introduced in order to increase system reliability and discrimination robustness of the distraction state when one of said decisive factors is not detected. As a differential factor opposed to the state of the art, this work presents the inclusion of head movement and training of an artificial neural network, that allow independence from the driver´s anthropometry features and increased robustness when movements of eyes or mouth are not detected. Controller validation is performed in both a test environment and a real environment for different users.

This document has the following structure: Section 2 presents characteristics obtained from the driver's face through the use of image processing techniques. At section 3 training of the decision system (neural network) is described. Section 4 presents results obtained from this research. At the end, achieved conclusions are socialized.

2. Feature extraction

Detection of the distraction state in video sequences is performed in two stages, the first one devoted to the detection of features which allow the system to identify the distraction regarding the driver's attention on the road, which is characterized by constant frontal sight of the driver; from here on said process will be named coarse segmentation. The second one discriminates quantifiable parameters in the features that were obtained, in order to derive information of interest from each sequence; this step will be named from here on out fine segmentation.

2.1 Coarse segmentation

Identification of the distraction state is related to eye, head and mouth movements, thus it is necessary to segment these three elements from the driver's image. A typical method for face segmentation is the use of Haar classifier techniques (Viola & Jones ,2001).

Haar classifiers operate in function of rectangular descriptors, which are related to the intensity of an image region. Viola & Jones (2001) presented the general algorithm to implement a classifier of this kind; here this classifier is used to detect faces in an image. Jiménez *et al*. (2011), used Haar classifiers to perform a driver's eye and mouth segmentation in order to establish his level of fatigue, this work is a extension and variation of the basic algorithm used in the first reference.

The system proposed by Viola & Jones (2001) uses a machine learning algorithm called Adaboost in order to create the classifier. Adaboost creates several weak classifiers (h_j), each of these evaluates a Haar characteristic (j) over an image (x_j), and through the comparison between the obtained value from the evaluation and a threshold (θ_j), it decides if this characteristic represents effectively the interest object. A weak classifier is defined by Eq. (1).

$$h_j(x) = \begin{cases} 1; f_j(x) < \theta_j \\ 0; f_j(x) \geq \theta_j \end{cases} \quad (1)$$

Adaboost will find the best threshold and the best classifier through linear searches and a reweighting of the examples with the highest classification error (ε_j), thus maximizing the margin between a positive and negative set of examples (x_j, y_j), being $y_j = 1$ or $y_j = 0$ for positives and negatives examples respectively. This classification error is defined by Eq. (2).

$$\varepsilon_j = \sum^i w_i \, | \, h_j(x_j) - y_i \, | \quad (2)$$

In Equation (2) the term (w_i) represents the weight given to the samples after each classification; as w_i increases for those misclassified samples, this will allow future iterations to pay more attention to these examples. Through this process Adaboost will use the best classifier to create a combination with better discrimination accuracy; this combination is called strong classifier (h) and is defined by Eq. (3).

$$h(x) = \begin{cases} 1 \sum_{t=1}^{T} \alpha h_j(x) \geq \sum_{t=1}^{T} \alpha \\ 0 \sum_{t=1}^{T} \alpha h_j(x) < \sum_{t=1}^{T} \alpha \end{cases} \quad (3)$$

For images used in the training of the classifier two object of interest had been marked: eyes and mouth. Saying so, the output of the equation 1 should be 1 if an eye or a mouth is detected.

Figure 1 presents the output of the Haar classifier trained to detect eyes. In order to validate the performance of this classifier a confusion matrix is used, which considers a true positive value as the proper detection of eyes and a false positive value as the incorrect detection of those. Table 1 presents performance of the eye detection algorithm with the confusion matrix, analyzed as a function of 20 videos of 1 minute length at a resolution of 640 x 480 pixels, running at a speed of 30 frames per second.

Figure 1. *Detection algorithm's output.*

Table 1*. Eye detection confusion matrix*

Eye detection		***PREDICTED***	
		POSITIVE	***NEGATIVE***
REAL	***FALSE***	*31*	*46*
	TRUE	*15830*	*215*
Sensibility		*99,71*	
Specificity		*87,40*	
Precision		*99,80*	

Figure 2 shows the output of the Haar classifier trained to detect mouths and Table 2 presents the confusion matrix of this classifier's performance using the same 20 videos.

Figure 2*. Detection algorithm's output*

Table 2. *Mouth detection confusion matrix*

Mouth detection		***PREDICTED***	
		POSITIVE	***NEGATIVE***
REAL	***FALSE***	*15*	*68*
	TRUE	*15930*	*109*
Sensibility		*99,57*	
Specificity		*87,90*	
Precision		*99,91*	

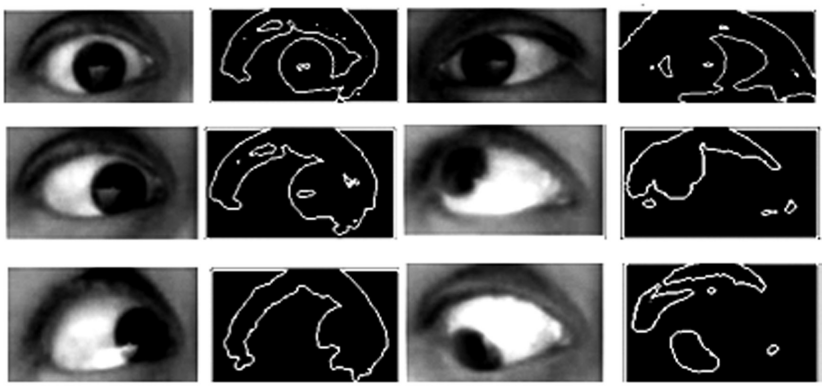

Figure 3. *Output of the edge detection algorithm over eye images*

In general terms, the performance achieved by each one of the classifiers falls within efficient qualification, allowing an accurate identification of both parameters in different individuals. The classifiers are able to obtain the position of the mouth and eye regions, which later will be subject to the fine segmentation process through image processing techniques. This mode of operation also increases the system's processing speed.

2.2. Fine segmentation

Once the location of the eye is obtained by the Haar classifier, the image is transformed to gray scale, in order to apply a threshold operation which returns a binary image. Over this binary image a Canny-type edge detection algorithm is applied (Canny, 1986), as seen on Figure 3.

Using a Hough circle detection algorithm (Chan & Siu, 1989) the location of the eye's iris is determined; this can be observed in Figure 4.

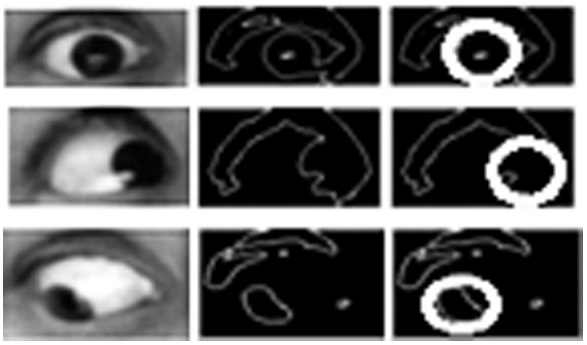

Figure 4. *On thick white, iris detection through the Hough transform*

Dividing the image into four parts through the use of perpendicular lines, five characteristic points are obtained and numbered from 0 to 5, which correspond to the center of each new part and the points where the two lines cross. Gaze direction is established regarding the center of the circle obtained through the Hough transform and its proximity to each of the points described above (Figure 5).

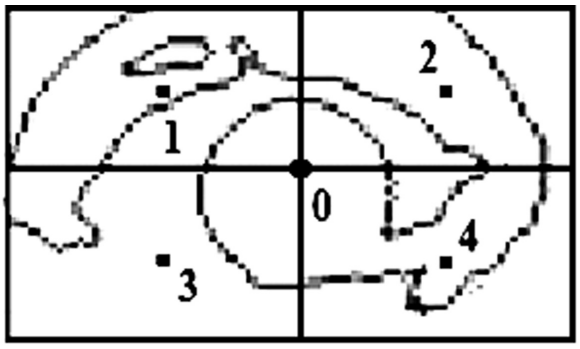

Figure 5. *Gaze direction*

In Figure 5, distance from Point 0 to any of the other points is always the same and is denoted as **dpc**; a reference distance **dr** is obtained from this and is given by **dr = dpc/2**. Therefore, if the distance from the center of the iris to Point 0 (**dpo**) is greater or equal to dr, a distraction state is present (Eq. (4)).

$$distraction = \begin{cases} 1, \forall dco \geq dr \\ 1, \forall dco < dr \end{cases} \quad (4)$$

Mouth fine segmentation is performed through a conversion from the RGB space to the YCbCr

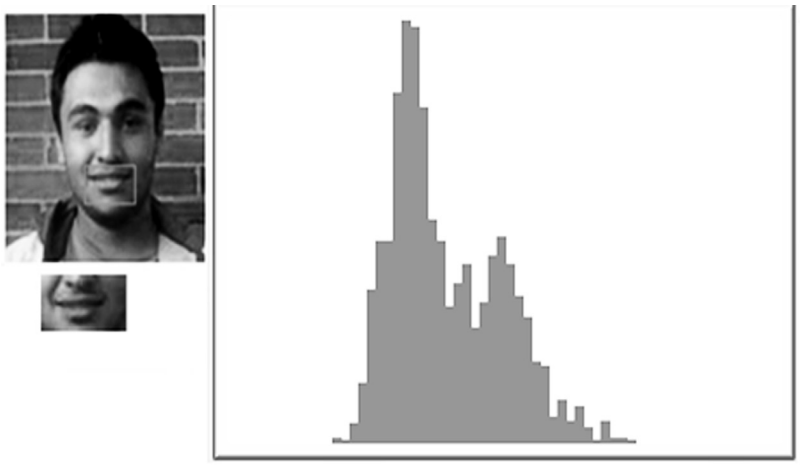

Figure 6. *Mouth region histogram.*

space (Eq. (5)) and a later histogram thresholding over the Cr component, thus obtaining the pixels of the lips, which were generally found in the second lobe of the histogram (Figure 6).

$$Cr = 0,4998 * R - 0,4185 * G - 0,0812B \quad (5)$$

Figure 7 shows an inherent problem on mouth detection, which is the difference between a closed mouth versus an open mouth, regarding the intensity from the teeth region.

Figure 7. *Lips fine segmentation*

This problem was solved through the use of sweep processing through the contours of the detected lip, thus obtaining a measurement for the mouth's openness. Figure 8 presents fine segmentation for several mouth movements.

Determination of head movement was implemented through the use of the method proposed by Jiménez *et al*. (2012). It adds robustness to the system in cases where the eye's movement cannot be determined (e.g. driver wearing sunglasses). Figure 9 presents the scheme proposed by Jiménez *et al*. (2012) which uses a movement vector with initial Pi (the medium distance between the eyes with frontal gaze). Then this point varies its position when the head moves, thus converting in the final point of the vector Pf. The angle between Pf and the X along with the magnitude given by the Eq. (6), establish the movement vector.

$$d(Pf,Pi) = \sqrt{(x_f - x_i)^2 + (y_f - y_i)^2} \quad (6)$$

Figure 10 shows the output of the system blending coarse and fine segmentation, using the video database, it can be noted the detection of mouth and eyes as well as the direction vector of the head.

3. Distraction state determination.

Once the characteristics associated with the distraction state are quantified, it is observed that changes in movement vector are spread along a wide range of values of magnitude and angle, as well as the mouth movements. This is why an artificial neural network is used to establish the level of distraction. This network uses movements and combinations of head, mouth and eyes position as its training set.

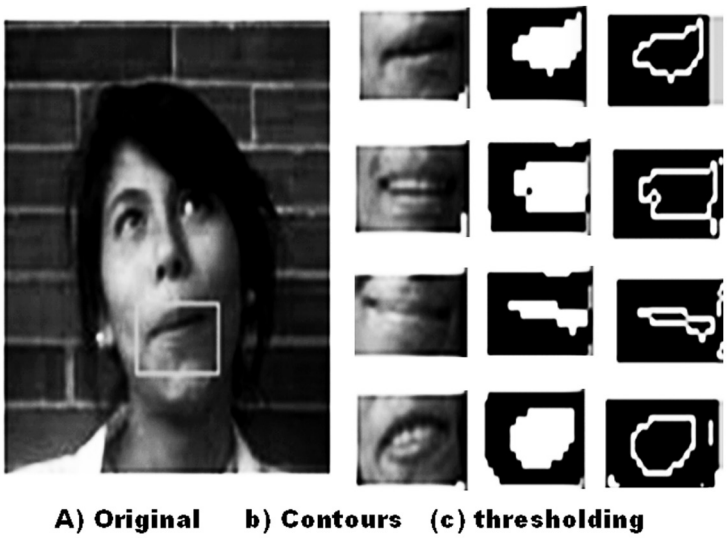

Figure 8. *Mouth state determination*

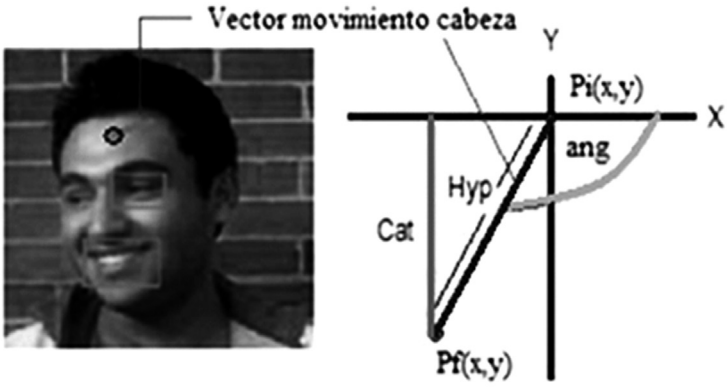

Figure 9. *Head movement detection*

Figure 10. *Output of the detection system*

Within the structure of the neural network the number of neurons in the input layer is determined by the variables established in the fine segmentation: the angle and the magnitude of the head's movement, the amplitude of the mouth's openness (both horizontally and vertically), the existence of eye distraction and feedback from the output of the network on instants t-1 and t-2. These last variables are due to the implicit temporal relation in a movement that indicates distraction, which should have minimal persistence.

Only one hidden layer is used in order to obtain a prediction time as low as possible without altering precision; the number of neurons on this layer is established using the iterative method described in Nilsson (2001 pp.46-49) and Looney (1997 pp. 307-309). Through this method a total of 65 neurons for 110 epochs were obtained. The output layer consists of two neurons in order to obtain a codified output of 4 states (Table 3).

Table 3. *Codified output from the network*

Neuron 1	*Neuron 2*	*Meaning*
0	*0*	*No distraction*
0	*1*	*Warning*
1	*0*	*Distraction*
1	*1*	*No assigned*

For validation of artificial neural network training, data set used was distributed like show Table 4, with about 10000 magnitudes and angles of the features extracted from 20 users.

Table 4. *RNA data set distribution*

Data set	*Percentage*
Training	*15*
Validation	*40*
Test	*45*

4. Result and discussion

Initially, several tests were conducted in a controlled environment with homogenous conditions, where 20 mechatronics engineering students from Nueva Granada Military University (UMNG) served as study subjects; they simulated typical driver distraction movements, which were the foundation in order to proceed with development of the algorithms and final tests of the neural network's prediction of the distraction state. Said results are presented at Table 5.

The system detected fatigue with a precision of 90 % in the test environment. This value was obtained from 962 images extracted from the 20 videos. False positive cases mainly affect mouth detection and are probably caused by shadows, which cause disturbance in the binarization process.

Table 5. *Detection distraction confusion matrix*

Mouth detection		*PREDICTED*	
		POSITIVE	*NEGATIVE*
REAL	*FALSE*	*34*	*29*
	TRUE	*314*	*585*
Sensibility		*91,5451895*	
Specificity		*94,50726979*	
Precision		*90,22988506*	

Finally, four videos in the real environment were analyzed and divided in two cases: with or without a partner alongside the driver. In the first case, Figure 11 shows a driver taking with a passenger and the state of distraction related to this action. Figure 12 shows a driver with an open mouth and looking between the rearview mirror, where the amount of exceeded

Figure 11. *Output of the detection system in real environment with partner*

Figure 12. *Output of the detection system in real environment without partner*

Table 6. *Detection distraction confusion matrix in real environment*

Mouth detection		***PREDICTED***	
		POSITIVE	***NEGATIVE***
REAL	***FALSE***	*48*	*30*
	TRUE	*300*	*584*
Sensibility		*90,9090*	
Specificity		*92,40506*	
Precision		*86,206885*	

time determines the state of distraction. Table 6 presents the final results in the real environment, where precision was of 86%. For these cases, the number of false positives increases due to major incidence of shadows, variations in light and distance between the driver and the camera; error associated to this last state was observed only in the real environment due to driver movements while driving a desired route, given different possibilities concerning driving requirements.

Ten percent of the real distraction states were not detected, this error is derivate at the same conditions of false positives. Is a low percent but is necessary tuning the algorithms for reduce this value.

Tests were performed using test subjects driving while wearing glasses in order to create eye, thus testing the resilience of the system; in this case precision drops to 80%.

Comparing with the state of art, the results obtained show that the different variables used and Neural Network training, permit to obtain a robust generic prototype in relation of use of this for many users. For example (Zhai, 2010) use some similar machine vision techniques but not presents the evaluation for different users. (Jiang,2011) and (Jia,2012) use other machine vision techniques but neither presents the evaluation for different users.

5. Conclusions

The proposed system is able to identify driver distraction states using movement and orientation of eyes, mouth and head as identification parameters. Global detection achieved a precision of 99% under the following controlled conditions as: maximum rotation of the head of 120 °, soft movements and a distance between driver and camera of no more than 70cm.

Relying on several factors to perform the discrimination of the distraction state allows the system to increase the robustness against similar work found in the literature under controlled conditions. Under non-controlled conditions, system performance drops in a lower amount compared to other work. Furthermore, global performance achieves a precision of 86% using a solution which had not been previously considered.

Mouth binarization through histogram analysis did not fully resolve problems caused by variations in light; as future work, robustness through machine vision algorithms against changes in light condition is required.

6. References

Bajaj Preeti, Narendra Narole & Mandalapu Sarada Devi , (2010). *Research on Driver's Fatigue Detection.* In: IEEE eNews SMC, Issue 31.p.1-9.

Canny, J. (1986). A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8(6), 679-698.

Chan, R., & Siu, W.C. (1989). *A new approach for efficient Hough transform for circles*. In: Conference Proceeding on Communications, Computers and Signal Processing. Conference Proceeding., IEEE Pacific Rim Conference on , vol., no., pp.99,102, 1-2 June 1989

Coetzer, R. C. & Hancke, G.P., (2009) "Driver fatigue detection : A survey," AFRICON, 2009.. , vol., no., p.1-6.

Crundall, D., & Underwood, G. (2011). Visual Attention While Driving: Measures of Eye Movements Used in Driving Research, In: Bryan E. Porter, Editor(s), *Handbook of Traffic Psychology.* Academic Press, (p. 137-148)

Jia Mingxing; Xu Hengyuan; Wang Fei, (2012). "Research on driver's face detection and position method based on image processing," Control and Decision Conference (CCDC), 2012 24th Chinese, vol., no., pp.1954,1959, 23-25 May 2012. doi: 10.1109/CCDC.2012.6244315.

Jiang Yuying; Wu Yazhen; Xu Haitao, "A surveillance method for driver's fatigue and distraction based on machine vision," Transportation, Mechanical, and Electrical Engineering (TMEE), 2011 International Conference on , vol., no., pp.727,730, 16-18 Dec. 2011. Changchun, China .

Jiménez R., Prieto F., Grisales V. (2011). *Detection of the tiredness level of drivers using machine vision techniques.* In: IEEE CERMA. México.

Jiménez R., Orjuela S. A; Van Hese P., Prieto F. A., Grisales V. H. & Philips W. (2012). *Video surveillance for monitoring driver's fatigue and distraction.* Proc. SPIE 8436 84360T; http://dx.doi.org/10.1117/12.922085.

Looney, C. (1997). *Patter Recognition Using Neural Networks: Theory and Algorithms for Engineers and Scientists.* Oxford University Press. New York.

Muhrer, E., Vollrath, M.. (2011). The effect of visual and cognitive distraction on driver's anticipation in a simulated car following scenario. *Transportation Research Part F: Traffic Psychology and Behaviour* 14(6), 555-566.

NHTSA (2009). *Traffic safety facts*. Research note: An examination of driver distraction as recorded in NHTSA databases

Nilsson, N. (2001). Inteligencia Artificial: Una nueva síntesis. McGraw Hill. España

Viola, P.; Jones, M.,(2001). "Rapid object detection using a boosted cascade of simple features," Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on , vol.1, no., pp.I-511,I-518 vol.1, 2001.

Xue Qing; Xiaoming Ren; Changwei Zheng; Yonghong Liu, (2012). "Research of Driver's Visual Perception Modeling and Simulation," Computer Science and Electronics Engineering (ICCSEE), 2012 International Conference on , vol.1, no., pp.484,488, 23-25 March 2012. Hangzhou, China.

Fangwen Zhai; Zehong Yang; Yixu Song; Hongbiao Ma, (2010). "A detection model for driver's unsafe states based on real-time face-vision," Image Analysis and Signal Processing (IASP), 2010 International Conference on , vol., no., pp.142,145, 9-11 April 2010. Xiamen, China.