

FACIAL FEATURES TRACKING APPLIED TO DRIVERS DROWSINESS DETECTION

L. M. Bergasa, R. Barea, E. López, M. Escudero, L. Boquete, J. I. Pinedo
Departamento de Electrónica, Universidad de Alcalá
Campus universitario s/n. 28871 Alcalá de Henares (Madrid)
SPAIN

ABSTRACT

In this paper we describe a real time facial features tracker applied to the detection of some basic drowsiness behaviours in drivers, with a colour camera. It uses stochastic colour segmentation to robustly track a person's skin and some facial features (eyes and lips). The system recovers 3D face direction, classifies rotation in all viewing directions and detects the driver's state analysing eye blinking and face direction. We show some experimental results, using real sequences with several users, and some conclusions about the performance of the system.

KEY WORDS

Facial Features Tracking, Driver Drowsiness, Colour Segmentation, 3D Recover, Eye Blinking.

1. INTRODUCTION

A facial features tracking system could be useful in warning drivers when they fell asleep. We have taken into account that between the 20% and the 30% of vial accidents in Spain are caused because of tiredness, according to Spanish Traffic Police. Several researchers have worked on head and facial features tracking in order to estimate gaze direction. Some of ones use artificial environments (cameras mounted on a helmet, infrared lighting, marking on the face, etc) [1]. The discomfort and the restriction of the motion affect the person's behaviour, which therefore makes it difficult to measure his/her natural behaviour. To solve this problem, many research results have been reported to visually detect gaze direction. Most of these systems use monocular vision and recover the 3D pose from monocular image stream [2], [3], [10]. Some others use stereo vision and a 3D facial features model [4]. Stereo vision is robuster and more accurate than monocular vision but the last one is complexer and more expensive than the first one. All of seen systems only detect gaze direction and, except [3], work in inside environments. There are some previous works on driver alertness, [5] relies on LEDs and uses multiple cameras, [6] uses only a colour camera; both of them systems was checked with a stopped car. A moving vehicle presents new challenges like variable

lighting and changing backgrounds. The MIT Smart Car [7] uses a lot of sensors embedded in an automobile and measure drive stress using electrocardiogram, electromyogram, respiration, skin conductance and a camera which test face movements of the driver for detecting driver state with a moving vehicle. In this case the visual information is secondary and it is used for other sensors confirmation.

On the other hand, in last years cars companies have been working on driver's drowsiness warnings, using several techniques. However, nowadays it doesn't exist a commercialized prototype yet. There is an important national project called TCD (Tech Co Driver) performed by the INTA (Aerospace Technique National Institute) in collaboration with some public organisms and two private companies [8]. This system analyzes the vehicle trajectory as function of time. The problem with this method is that it takes it too much time to analyse user behaviours and thereby it doesn't work with fast distractions. In a couple of years this system will be commercialized. Another important project is the ASV (Advanced Safety Vehicle), performed by Toyota, Nissan and Honda [9]. This project includes a warning drivers based on the pulse monitoring, but this system is intrusive because the driver must wear a wristband.

We describe a framework for analyzing images of driving and determining when the driver is not paying adequate attention to the road. We use a single colour camera placed on the car dashboard. We have developed a robust real time face direction tracker using skin, eyes and mouth colour detection by Gaussian models. Automatic initialization of Gaussian model parameters is achieved by colour clustering in the image. In this manner we obtain initial facial features positions without manual processing in the initial image. With the 2D coordinates of eyes and mouth on the image we recover a 3D face model in order to obtain the 3D face direction from a single camera. Then, we have implemented a blink detection algorithm.

The system is able to automatically initialize when it miss-tracks head rotations. Using blink information and face direction we classify rotation in all viewing directions (front, left, right, above, below) monitoring driver alertness.

In Section 2 we describe our face tracking and driver alertness method in detail. Experimental results that show the performance of the system in real situation are presented in Section 3. Finally the conclusions and a discussion of the future work are given in Section 4.

2. SYSTEM ARCHITECTURE

The outline of the software configuration for face tracking and driver alertness is show in figure 1. It consists of three major parts, 1) Initialization, 2) Face tracking and 3) Driver Alertness. In the initialization stage, the system detects skin, eyes and mouth of the user in an automatic and unsupervised manner. After that, the system stars lips and eyes tracking, recovering 3D face direction from a single camera and estimating 2D positions of lip and eyes in the new frame. If face direction tracking is not successful, the system regards the face to be lost and it comes back to the initialization stage for finding facial features again. If face tracking is successful, the system warns to driver when it detects he falls asleep. This behaviour is detected analysing temporal evolution of head rotation and blinking. Finally, the algorithm jumps back to the face tracking stage in the next frame.

2.1. Initialization

This process is used to initialize head position at the beginning of the tracking and when eyes and lips tracking fail. It starts locating face skin using an stochastic model, called UASGM (Unsupervised and Adaptive Gaussian skin-colour model), developed by the authors and explained in reference [11].

The algorithm is able to locate user's skin on an colour image in an unsupervised way using clustering and detecting the cluster closer to a colour patron equal for everybody. It works properly with random backgrounds and light changes if the skin cluster has a significant size on the image. There is some differences in the skin colour as function on light. In order to reduce it we work on a normalized RG space ("rg") where the intensity influence is considerable reduced. In an initial clustering process image is divided in its main colours. Then, the skin colour cluster is located and it is modelled by a 2D Gaussian function. If the probability that a pixel belongs to the skin class is bigger than certain threshold, this will be classified as skin. A linear combination of the previous model parameters will be used to predict the new ones. The maximum likelihood criterion will be used to find the bet set of optimal coefficients for prediction. In figure 2 the results of the skin segmentator are shown.

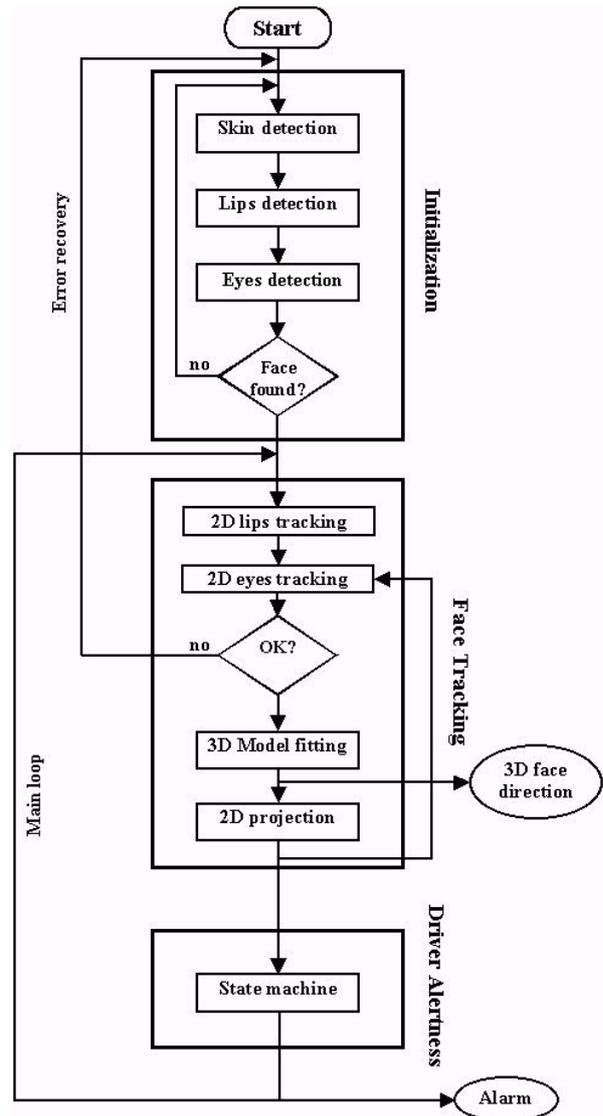


Fig 1. General architecture

On the skin blob and keeping in mind geometric characteristics of the face some search areas for lips and eyes are defined. On these, the same segmentation algorithm is applied, but in this case we search for lips and the sclera (white of eyes) colours. Lips colour has an "r" component very marked that allows to different it of the rest of the skin. For locating eye position we detect white of eyes because is a common colour for every body and it is different of skin, rainbow and brows colours. The segmentation algorithm models lips and eyes colours automatically by means of two Gaussian functions in an RG normalized colour space. Pixels will be segmented as lips or eyes as function of their probability to belong to its respective classes. Figure 2 shows lips and eyes colour distributions as well as the segmentation of these characteristics.

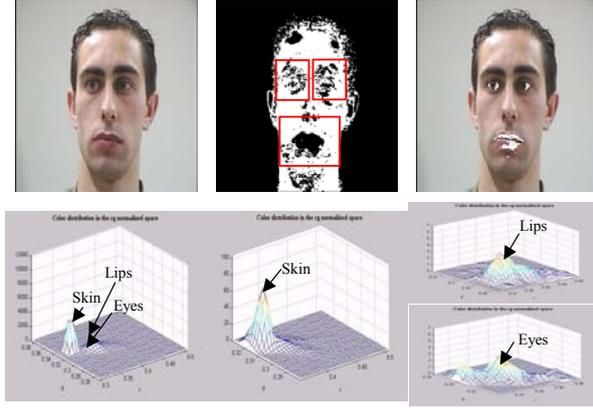


Fig 2. Colour distribution in the RG normalized space

In order to calculate 2D positions of the facial features an integral projection of the segmented pixels on the X and Y coordinate axes is done, obtaining some profiles for mouth and eye like those shown in figure 3. Lips vertical position is calculated by average the maximum values of the projection. In horizontal coordinate, lips edges correspond to the maximum derivative positive and negative of the projection. Horizontal value is obtained by average of these points. Calculation of eyes position is done of the same manner, in this case horizontal and vertical positions correspond to the maximum values of the projection.

In conclusion, in the initialization phase users eyes and mouth 2D positions are obtained (x_{er}, y_{er}) (x_{el}, y_{el}) (x_l, y_l) in an automatic way and for any head position.

2.2. Face Tracking

In this stage eyes and lips location are estimated in a frame taken in account 2D positions of them in previous frames. For improving the tracking a 3D facial model has been reconstructed from the information of a single camera. The reason we can do this is that we only need the direction of the face. For practical purposes, this could go on through the windshield. By making this assumption we eliminate the need to know the distance from head to the camera. Also, if we assume that this distance stays constant and that head size is relatively constant between people then we have all information we need.

As we show in Figure 4, with 2D eyes and lips location and estimating the distance to head rotation point (D) we can recover 3D face direction using basic projections. Distance to head rotation point can be well calculated by the average of the two eyes subtracted from its distance from the centre of the head in the first frame [6]. We consider vertical and horizontal rotations only (\forall, Ξ) because there are the most common in drivers.

Horizontal rotation (\forall) is calculated as a function of the horizontal projection of the distance between the face centre in the n iteration $(x_f(n), y_f(n))$ and its initial position

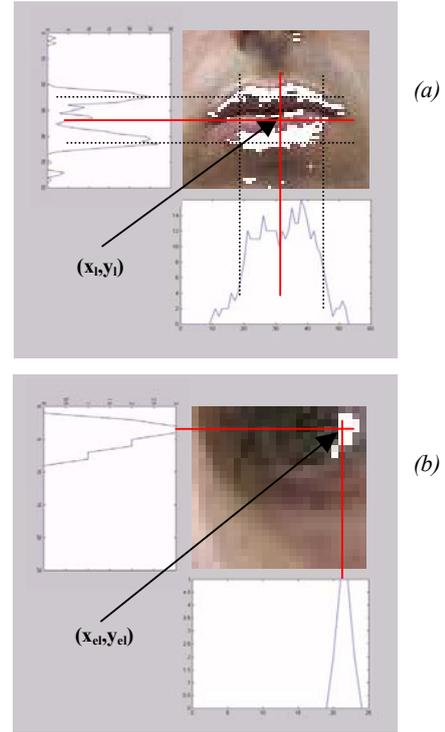


Fig 3. (a) Lips projection results (b) Eye projection results

$(x_f(0), y_f(0))$, when the user is looking at the front. Then, it is given by,

$$\alpha = \arcsin\left(\frac{\Delta x_f(n)}{D}\right) \quad (1)$$

Vertical rotation is calculated in the same manner using vertical projection, it is given by,

$$\beta = \arcsin\left(\frac{\Delta y_f(n)}{D}\right) \quad (2)$$

Utilizing as state vector, $\mathbf{X}(n)$, horizontal and vertical rotations, we can track robustness facial direction estimating the state vector by a zero order Kalman filter as we show in the following equations:

$$\hat{m}(n) = \frac{\sum_{i=1}^{N-1} X(n-i)}{N} \quad \text{when } X(n) = \begin{bmatrix} \alpha(n) \\ \beta(n) \end{bmatrix} \quad (3)$$

$$P_w(n) = \frac{1}{\lambda_w} (P_w(n-1) - K_w(n) \cdot P_w^T(n-1)) \quad (4)$$

$$\hat{X}(n) = \hat{X}(n-1) + K_w(n) \cdot (X(n) - \hat{m}(n)) \quad (5)$$

$$K_w(n) = \frac{P_w(n-1)}{(\lambda_w + P_w(n-1))} \quad (6)$$

where δ_w was obtained in an experimental form.

Projecting the 3D model for the estimated vector we find the estimated lips and eyes position for the next frame. In

these positions some search windows are located in order to find the facial features. Applying the 3D model we obtain the observed angles. If these are close to the estimated ones they are validated and the facial model is updated. If not, the model is updated with the estimation when the features segmentation are correct. If two eyes detection fails (but not lips detection) it can be given by a blinking. Then, the model is updated with observed lips position and eyes positions are estimated. In this case driver alertness stage must detect if this action is only a blink or the user has fallen asleep. When lips detection fails the algorithm comes back to the initialisation stage because this segmentation is very robust and if it fails this would mean that the tracking is lost.

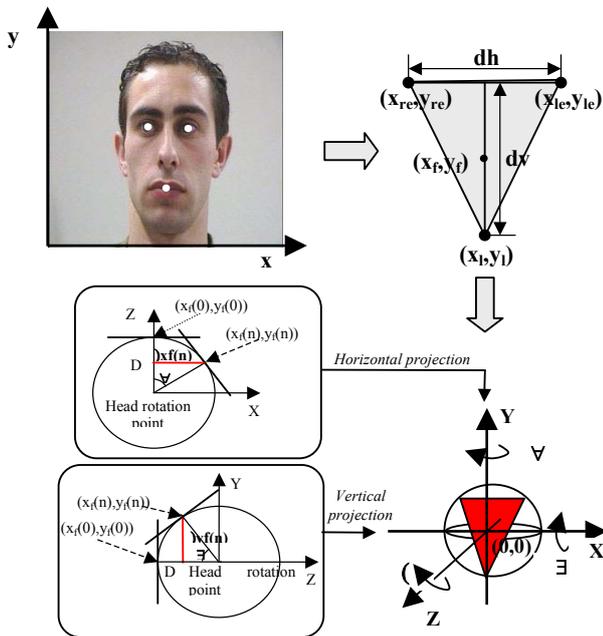


Fig 4. Recovering the 3D facial model

2.3. Driver Alertness

A rigorous analysis on the physiology of sleep behaviours would be necessary before accurately determining when a driver has fallen asleep. For our system however, we are able to determine a basic set of criteria to determine driver vigilance. When the driver's eyes are occluded (either from blinking or rotation occlusion), for too long we warn that the driver's alertness is too low. Then, when the blinking frequency is so big, then we warn that the driver's alertness is too low. We assume the driver has a low vigilance level if the eyes are closed for more than 2 seconds or if the blinking frequency is more than 0.5 Hz. Finally for general rotation, we will not give drive alertness warnings unless the rotation (30° for horizontal and 20° for vertical) is prolonged (more than 3 s). For controlling the different behaviours as a function of time we have used a state machine.

3. EXPERIMENTAL RESULTS

The system processes 15 images per second, with a resolution of 256×256 pixels, in its normal operation mode, using a not optimized processor (Pentium II to 400 MHz) and with a commercial "frame-grabber" Matrox Meteor II. The initialization stage has an average time process of 300 ms. In figure 5 an operation example of the face tracking system with a sequence recorded from a car is shown. On them, some messages that have been shown automatically appear, indicating estimated face direction. We have restricted these to the following groups: front, right, left, up and down.

Figure 6 shows some blinking detector results in a real situation sequence for other user. In this case some prolonged eyes closed behaviours are detected.

In figure 7 we can see another sequence, in this case for a girl. Figure 7.(a) shows some frames of the sequence. Figure 7.(b) and (c) show the estimated angles (\forall, \exists) for this sequence. In figure 7.(d) we see an example for turn right and left detection and in (e) the closed eyes detection. The algorithm has been tested using 5 sequences of 3 minutes each one for 5 different users. The results are shown in table 1.

As we can see the best results are obtained for "closed eyes" behaviour decreasing in the case of "high blinking frequency" one because the detection time is less. In rotation behaviours, the results are best for the horizontal rotation because there is more variation in the image than in the vertical rotation. The system performance decrease when light conditions change quickly, sudden head movements are done or the user wears glasses.

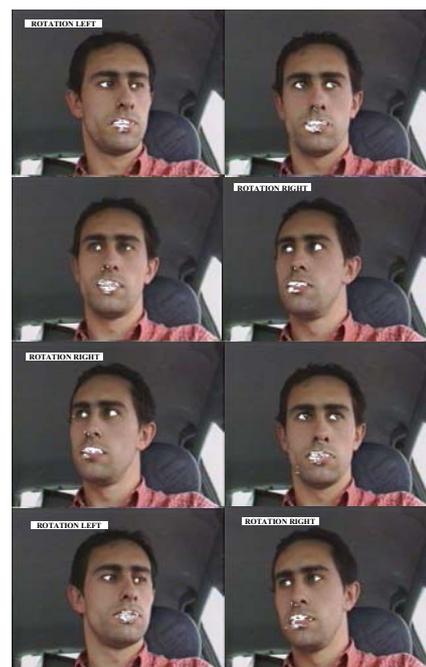


Fig 5. Result of face tracking



Fig 6. Result of blinks

Behaviours	Accuracy (%)
Prolonged eyes closed	95
High blinking frequency	70
Prolonged horizontal rotation	85
Prolonged vertical rotation	75

Table 1. Accuracy of the results

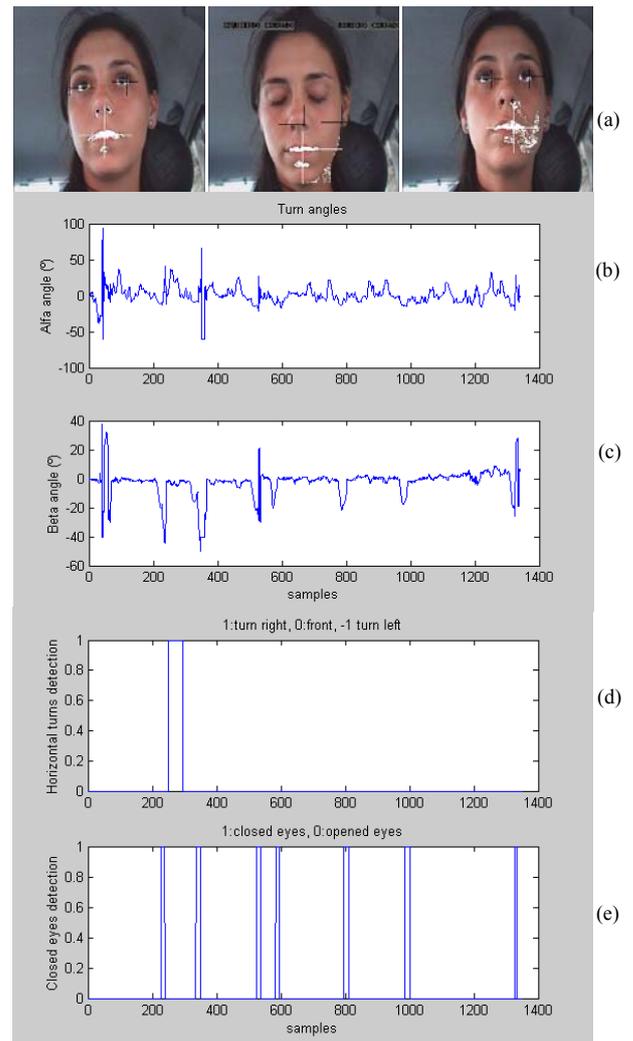


Fig 7. Sequence results

4. CONCLUSIONS AND FUTURE WORKS

We have presented a facial tracking method based on stochastic colour segmentation using adaptive Gaussian functions obtained in an unsupervised manner. A 3D facial model is recovered for estimating facial features positions and for classifying rotations in some fuzzy direction groups. It is able to detect some tiredness driver symptoms as: closed eyes, continuous blinking and extended head rotate. It has been tested by day under changing light conditions obtaining some acceptable results.

In the future we have the intention to improve driver alertness system with new behaviours as well as face direction estimation by the using of 2 cameras. Then, we would like to introduce an infrared camera for nocturnal working.

5. ACKNOWLEDGEMENTS

The authors wish to acknowledge the contribution of the *Comunidad de Madrid* for PIROGAUCE project (Ref:07T/0057/2000) financing and *Universidad de Alcalá* for DISISROM project (Ref: E017/2002) financing.

REFERENCES

- [1] J. J. Heuring and D. W. Murray. "Modeling and Copying Human Head Movements". IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION. Volume 15, Number 06, pp: 1080-1095. December 1999.
- [2] A.H.Gee and R. Cipolla. "Determining the Gaze of Faces in images ".Image and Vision Computing, 30:639-647, 1994.
- [3] Y. Matsumoto, J. Heinzmann and A. Zelinsky "The Essential Components of Human-Friendly Robot Systems". Procs. Of the International Conference on Field and Service Robotics (FSR'99). Pittsburgh, Pennsylvania, USA, 1999.
- [4] R. Stiefelhagen, J. Yang and Alex Waibel "Tracking Eyes and Monitoring Eye Gaze", Proc. of the Workshop on Perceptual User Interfaces, pp. 98-100, Banff, Canada, October 1997.
- [5] Quiang Ji and George Bebis "Visual Cues Extraction for Monitoring Driver's Vigilance". Procs. Honda Symposium, pp.48-55, 1999.
- [6] P. Smith, M. Shah and N. Da Vitoria Lobo. "Monitoring Head/Eye Motion for Driver Alertness with One Camera". 15 th International Conference on Pattern Recognition, Volumen 4, pp.636-642. Barcelona, Spain, 2000.
- [7] J. Healey and R. Picard, "SmartCar: Detecting Driver Stress". 15 th International Conference on Pattern Recognition, Volumen 4, pp.218-221. Barcelona, Spain, 2000.
- [8] Jesús Soria, "Llega el piloto tecnológico", Revista Tráfico, pp: 20-21, Marzo-Abril 2002
- [9] Richard Bishop, "Intelligent Vehicle applications worldwide", IEEE Intelligent System, Jan/Feb 2000, pp: 78-81.
- [10] T. Moriyama, T. Kanade, J. Cohn, J. Xiao, Z. Ambadar, J. Gao, and M. Imanura, "Automatic recognition of eye blinking in spontaneously occurring behaviour", Proceedings of the International Conference on Pattern Recognition (ICPR '2002), August, 2002
- [11] L. M. Bergasa, M. Mazo, A. Gardel, M.A. Sotelo, L. Boquete. "Unsupervised and Adaptive Gaussian skin-colour model". Image and Vision Computing (Elsevier Ed.), n° 18, pp:987-1003, September 2000.