Yawning detection by the analysis of variational descriptor for monitoring driver drowsiness

Belhassen AKROUT*, Walid Mahdi†,
*MIRACL Laboratory, Higher Institute of Management Gabes, Sfax university - Gabes university, Tunisia.
Email: bakrout@gmail.com
†MIRACL Laboratory, College of Computers and Information Technology, Taif University, KSA.
Email: walid.mahdi@isimsf.rnu.tn

***Abstract*—The road safety is a problem which was approached by several countries following a big raise of the number of accidents. The drowsiness represents one among the causes of the road accidents. The accidents related to the drowsiness often occur on the highways, but also on the main roads, even inside the localities. Today, it is possible to detect the state of tiredness of the driver with the development of the technology of the computer vision. The results of research in physiology show that the first level of lack of vigilance appears by an increase in the frequency of the yawn.**

In this work, we propose a novel approach for yawning detection for monitoring driver fatigue.In fact, our approach rests on the study of the spatio-temporal descriptors of a non-stationary and non-linear signal. This approach is evaluated by both *YawDD* [1] and our *MiraclHB* [2] databases. The evaluation shows many promising results and shows the effectiveness of the suggested approach.

Keywords*—Intelligent vehicle, Yawning detection, , Human safety, Driver drowsiness detection.

I. Introduction

These last years, various useful measurements were taken to avoid the increase in the road accidents. However, the causes related to the human being continue to be one among the major factors of these accidents. The reduction in the vigilance of the driver is regarded as the main reason for the road accidents.
The objective of this paper consists in identifying the drowsiness state by the detection of the driver yawn. This identification rests on a paramount stage of localization of the mouth interest zone. Several approaches were proposed for the detection of the yawn states. These approaches make it possible to describe mainly the image content according to many characteristics based on colours, forms and textures.
Indeed, in order to determine the opening degree of the mouth, Teishing Wang *et al* [3] propose the distance D_{OO} between the two lines of the box delimiting the mouth (equation 1).

$$D_{OO} = h\ \frac{cos(\theta)}{w} \tag{1}$$

Lu Yufeng *et al* [4] calculated the distance D_y of the two points y_1 and y_2 (equation 2). These two points respectively represent the chin and the nostrils. Lu Yufeng *et al* [4] estimated the states of yawn by comparing the distance D_y with an empirical threshold H_y.

$$D_y = |y2 - y1| \tag{2}$$

Nawal *et al* [5], as for them, have proposed a system which consists, initially, on the extraction of the face by the *SVM* [6] (Support Vector Machins). They detect, then, the circle which surrounds the mouth by the Circular Hough Transform (CHT). This detection is guided by a stage of localization of the mouth area. With this intention, Nawel *et al* detect the contour of the face, first of all, by using a gradient detector. Thereafter, they calculate the vertical projection (V_{proj}) of the lower half of the face to detect the borders left and right-hand side of the mouth. The tiredness of the driver is characterised by a high frequency of yawn. The driver, in this work, is considered a drowsiness state if he exceeds the two seconds with a broad opening of the mouth. The experiments are applied to real video sequences acquired by a camera to weak resolution and recorded under various conditions of lighting.
Shabnam *et al* [7] exploit in their method the strong difference between the colour of the lips and the colour of the face. They note that the red colour is the strongest component while the blue component is weakest in the area of the lips. The area of the localised mouth requires then a stages of post-processing such as conversion into black and white, erosion and dilation in order to locate the lips of the mouth. The state of yawn, according to Shabnam, is detected if the number of the pixels lips of the localised mouth exceeds a threshold S_X defined experimentally.
Lingling Li *et al.* [8] classified the mouth states in three categories. These categories are: closed, expression and yawn. Initially, the mouth is closed in the normal case, on the other hand, it is considered in a state of the yawn if it is largely open. Thus, the authors characterize the yawn according to the degree of the mouth opening by calculating three descriptors such as H_{max} which represents the maximum height of the mouth, W_{max} the maximum width of the mouth and H_m the height between the upper lip and the lower lip. The three descriptors, calculated previously, are then used like three entries for a classifier with the artificial neural network of three layers, in order to classify the state of the mouth on three levels.
The study of the descriptors used in the literature for the

978-1-5090-1645-7/16/$31.00

drowsiness detection reveals the presence of gaps on the level of various work which we have just presented.
Indeed, the work of Teishing Wang *et al* [3] and Yufeng [4] require a good localization of the mouth zone in order to calculate the distances D_{OO} and D_y. This work shows a failure when the driver turns its head. This rotation even if it is tiny affects the localization of the mouth.
Nawal *et al* [5] use in their work the Transform of Hough Circulaire to detect the external contour of the mouth. This method remains very sensitive to the intrinsic movements of a not-rigid object.
We propose, in the following, our new approach for the detection of the driver yawn states. This approach is based on the analysis of spatiotemporal descriptors following the detection of the lips edge. A paramount stage for localising the interest area of mouth region is necessary. With this intention, we adopt the method of Viola and Jones [9].

II. Proposed approach

The goal of our approach is to carry out the localization and the segmentation of the interest zone in order to extract a signal who describes various states of the mouth and consequently to identify the states of driver yawn. The figure 1 represents the stages of tiredness detection states based on the analysis of the spatio-temporal descriptors.

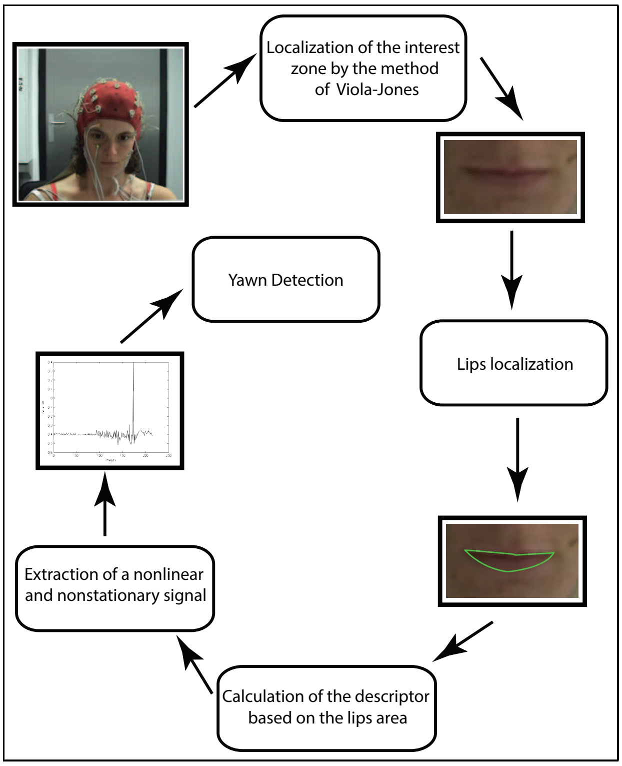

Fig. 1. Description of the stages of yawn detection.

A. Localization of the face and the mouth

As already mentioned previously, we propose to exploit the technique of Viola and Jones [9] for the stage of the localization of the face and the mouth. It is a largely exploited classifier who based on the Haar descriptors. This technique has the advantage of being very quick and the results which it reaches are satisfactory [9]. The figure 2 illustrates an example of face and mouth detection.

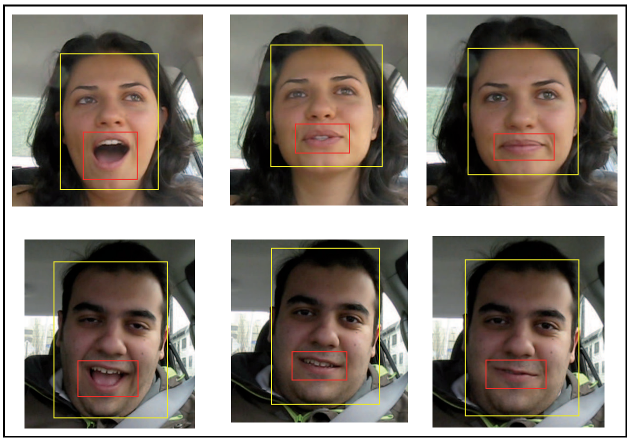

Fig. 2. Result of the face and the mouth detection by the technique of Viola and Jones.

In fact, the technique of Viola and Jones is based on a fast stage extraction of certain descriptors while making use of an integral image and a classifier based on Adaboost implemented according to a structure of cascade for a training supervised through several examples of objects analysed and classified in advance. Thereafter, the analysis and the classification of these objects are made by boosting in order to separate the positive examples from the negative examples according to the principle cascades of a decision. The figure 3 presents a diagram describing the various stages of this technique.

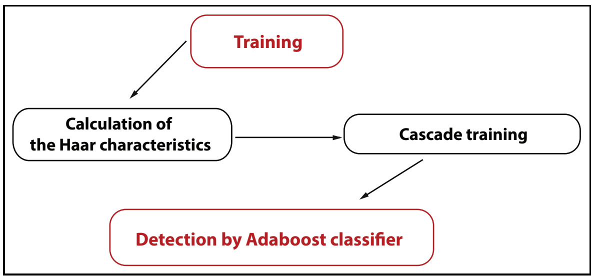

Fig. 3. The various stages of the Viola and Jones technique.

B. Detection of the external lips contours

Once the interest zone is located by the method of Viola and Jones [9], comes the stage of the descriptors lips extraction. This stage is essential to manage to detect the yawning states. In fact, the goal of this stage is to detect the external contour of the lips. The approaches known as contour making it possible to segment the lips are primarily based on deformable models

of contours. This method represents a robustness on the level of the segmentation and the follow-up of the nonrigid objects in real-time [10] [11], in our case the lips of the driver.
This type of method consists in choosing a mathematical model so that it adapts to contours of the interest object. The deformation of the initial model is guided by the minimization of an energy function. This function is made up of an energy internal term, describing the geometrical properties of contour and an energy external term, calculated starting from the properties of the image. This function represents the principle of active contours.
The principle of active contours was introduced by Kass *et al* [12] in 1988. Active contour (Snake) is a parametric curve which tries to move in a position where its energy is minimised as the equation 3.

$$E^*_{snake} = \int_0^1 E_{int}(V(s)) + E_{img}(V(s)) + E_{con}(V(s))ds \quad (3)$$

Active contours is divided into three terms. The first term E_{int} represents internal energy. The second term E_{img} represents the force of the image. The last term E_{idiot} constitutes the constraints of external force. Internal energy $Snake$ is written as follow

$$E_{int} = \frac{1}{2}(\alpha(s)||V_s(s)||^2 + \beta(s)||V_{ss}(s)||^2) \quad (4)$$

Such as $||V_S(S)||^2$ represents the measurement of elasticity, while $||V_{SS}(S)||^2$ represents the measurement of curve. That means that in certain parts of the $Snake$ where the curve is stretched, the term of elasticity will have a raised value, whereas, in certain parts of the $Snake$ where the curve is folded, the term of the curve has a great value.
The influence on the total energy of the $Snake$ is controlled by the coefficients $\alpha(s)$ and $\beta(s)$, these two parameters make the $Snake$ increasingly elastic and less rigid. Thus the $Snake$ positions in certain parts of the image with high values of the gradient.
Two main categories exist for the segmentation with active contours containing energy: approaches based on the edges and approaches based on the regions.
The approaches of active contours based on the edges use the gradient of the image to identify the limits of the object [13] [14].
The models of contours active based on the edges do not hold counts the appearance of areas inside or outside contour; only the information located near contour is examined.
The active contour based on the region relies on the assumption that the areas of an image have a constant intensity [15]. More advanced techniques try to model the areas by known distributions, by intensity histograms and by the analysis of the texture [16]. If the object to be segmented cannot be easily distinguished, active contours based on regions can lead to an erroneous segmentation. The figure 4(b) shows this type of error.

In our case, as one is interested in segmenting the lips, the teeth and the tongue in only one object and not each component with share. We exploit the method of active contours based on the edges like shows the figure 5.

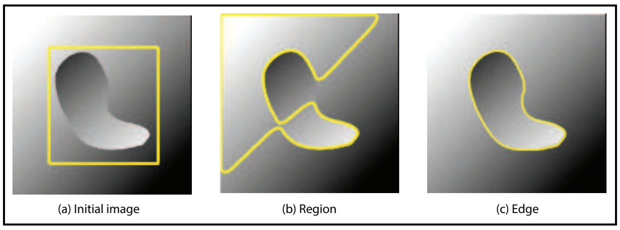

Fig. 4. A synthesised image with a heterogeneous intensity on a similar bottom. (a) Initial contour, (b) failed result of segmentation based-region, (c) successful result of segmentation based-edge.

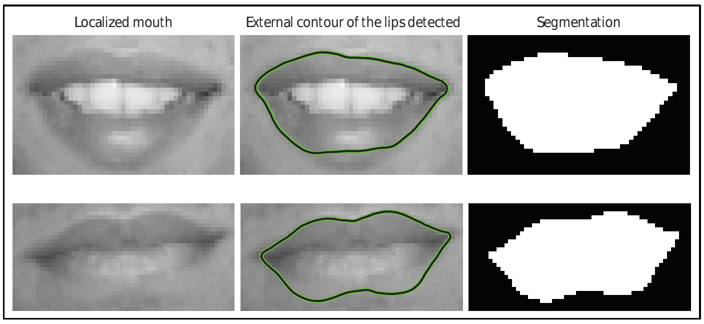

Fig. 5. Examples of detection of external lips contour with the method of active contours.

The figure 5 shows an example of results of mouth localization by the method of Viola and Jones [9], and the segmentation of the lips region by the technique of active contours. The goal of this stage does not consist in obtaining a precise segmentation of the lips, but rather to have a total idea of the evolution of the mouth shape (closed, moderately open, strongly open). Consequently, small errors of precise details in the localization of the lips region will be tolerable.

C. Extraction of the spatio-temporal descriptor

The extraction of the descriptors is a fundamental stage of the yawn detection process in our method suggested. In fact, after the localisation of the lips, we will be able to calculate a spatio-temporal descriptor based on the tracking of the lips surface evolution and the internal zone of the mouth (figure 6).

We propose the descriptor D_{mouth} which represents the sum of two surfaces. This surface represents the zone of the lips and the internal zone of the mouth (including tongue and teeth). Indeed, a tired driver can present states of yawn at driving. These states are characterised by the increase and the reduction in the internal zone of the mouth what influences on the value of the descriptor D_{mouth}. The descriptor D_{mouth} is calculated following the segmentation of the lips (figure 5 segmented image). It represents the sum of the white pixels P_n (equation 5) in the segmented image of the figure 5.

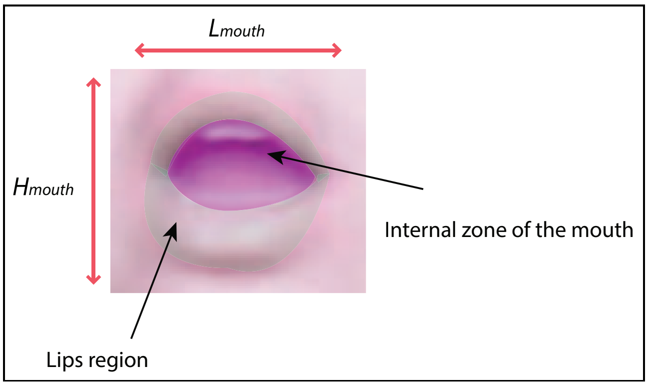

Fig. 6. Diagram describing the various components of the mouth.

$$D_{mouth} = \sum_{i=1}^{n} P_n \quad (5)$$

With an aim of having standardized data, we present the descriptor D_{mouth} according to K_{lh} (equation 6).

$$K_{lh} = L_{mouth} * H_{mouth} \quad (6)$$

The value of K_{lh} is calculated by the multiplication of two dimensions L_{mouth} and H_{mouth}. These dimensions respectively represent the height and the width of the mouth box (figure 6) localised by the Viola and Jones [9] method. This standardisation is represented by the equation 7.

$$D_{mouthNorm} = \frac{D_{mouth}}{K_{lh}} \quad (7)$$

The detection of the yawn states consists in thus evaluating the standardised variations of the descriptor D_{mouth}. This variation makes it possible to generate a spatio-temporal descriptor what results from it a signal $g(x_t)$ non-stationary and non-linear according to the equation 8.

$$g(x_t) = D^t_{mouthNorm} \quad (8)$$

The signal $g(x_t)$ is calculated by the extraction of the indices starting from a video sequence made up of three types of the mouth states: closed, a driver which is speaking and finally the yawn. The figure 7 shows an example of the signal $g(x_t)$ calculated with the possible variations of the mouth.

D. Yawning detection

The states of yawns appeared by high frequencies in the signal $g(x_t)$. In order to facilitate the detection of these states, a paramount stage is necessary. We thus propose to bring back the values of the signal between -1 and 1, in order to calculate the number of passages by zero. Each passage represents a state where the mouth is strongly open. With this intention, we propose the equation 9 which makes it possible to obtain a signal $g(x_t)$ in the interval [- 1 1].

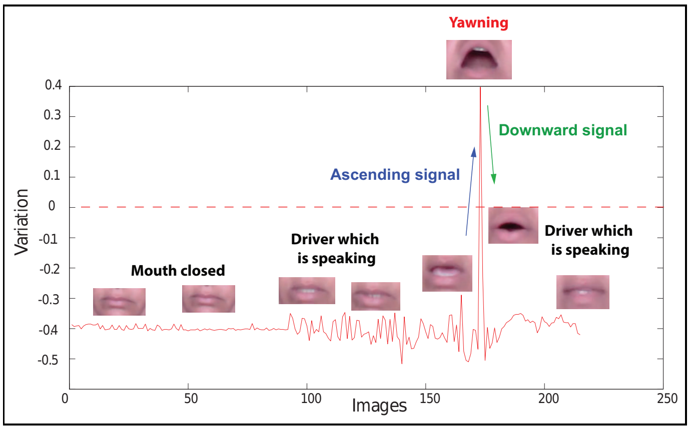

Fig. 7. Various states of the mouth.

$$z(x_t) = g(x_t) - (\frac{1}{2}\sum_{t=1}^{j} g(x_t)) \quad (9)$$

To locate the states of yawn thus amounts analysing the signal $z(x_t)$ for a window of 4 seconds, knowing that, according to the studies, the yawn lasts between 4 and 7 seconds [17]. The states of yawns are given if we detect two passages by zero, an ascending signal and another downward of the same window or an ascending passage and another downward of two successive windows of the signal $z(x_t)$. An example of a result of yawn detection between image 150 and image 200 is shown in the figure 7.

III. Experimentation and results

In this section, we present the results obtained by our approach of driver yawning detection. Indeed, in order to study the behaviour of our approach allowing to detect the states of the yawn, we carried out experiments by both *YawDD* [1] and our *MiraclHB* [2] databases. The opinion of an expert in this stage is essential to determine the real states of driver yawn. Indeed, B_r represents the opinion of an expert for the detection of the yawns real, B_a indicates correctly detected alarms and B_x indicates all the alarms generated by our approach. The rates of recall and accuracy are thus calculated according to the equation 10 and 11.

$$Recall = \frac{B_a}{B_r} \quad (10)$$

$$\text{Accuracy} = \frac{B_a}{B_x} \quad (11)$$

The table I shows a good performance of our method which reaches a rate an average of recall equal to 84 % and an average rate of accuracy equal to 85 %.

According to the exposed results, we note some failures. These failures are due primarily to the bad localizations of the mouth. Indeed, some drivers unconsciously hide their mouth by putting the hand in the mouth as is illustrated in the figure 8.

TABLE I. RECALL AND ACCURACY RATE OF YAWN DETECTION

Databases	Recall	Accuracy
MiraclHB	86%	87%
YawDD	82%	83%
Average	84%	85%

Fig. 8. An example where the mouth of the driver is invisible.

IV. CONCLUSION AND PERSPECTIVE

We presented in this paper a new approach to yawn detection for monitoring driver drowsiness. Our approach consists in locating the lips by active contours. The area of the localised lips is then represented by the variation of the descriptor D_{mouth} in the form of a non-stationary and non-linear signal $g(x_t)$. This signal is then analysed by the method of $Zero - Crossing$ in order to detect the yawn states of the driver.

We note that this method of the analysis of the spatio-temporal descriptors can present some gaps under certain conditions in occurrence the case of occlusion. The occlusions generate, in the case of the use of the active contours technique, a bad segmentation of the ROI (detection of the lips by active contours of the bearded drivers and/or men with a mustache). For this reason, and like perspective, we propose to carry out another approach of yawn detection based on the analysis of the lips movement and not to localise the lips. To improve moreover our approach, we need to produce a hybrid which uses the yawn analysis, study the states of the eyes [18] [19] and the *3D* pose estimation [20] to localise the driver drowsiness.

REFERENCES

[1] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, and B. Hariri, "Yawdd: A yawning detection dataset," in *Proceedings of the 5th ACM Multimedia Systems Conference*, ser. MMSys '14. New York, NY, USA: ACM, 2014, pp. 24–28. [Online]. Available: http://doi.acm.org/10.1145/2557642.2563678

[2] B. Akrout and W. Mahdi, "Spatio-temporal features for the automatic control of driver drowsiness state and lack of concentration," *Machine Vision and Applications*, vol. 26, no. 1, pp. 1–13, 2015. [Online]. Available: http://dx.doi.org/10.1007/s00138-014-0644-z

[3] W. Tiesheng and S. Pengfei, "Yawning detection for determining driver drowsiness," in *VLSI Design and Video Technology, 2005. Proceedings of 2005 IEEE International Workshop on*, May 2005, pp. 373–376.

[4] Y. Lu and Z. Wang, "Detecting driver yawning in successive images," in *Bioinformatics and Biomedical Engineering, 2007. ICBBE 2007. The 1st International Conference on*, July 2007, pp. 581–583.

[5] A. Nawal, A. Aouatif, and R. Mohammed, "Driver's fatigue detection based on yawning extraction," *International Journal of Vehicular Technology*, vol. 2014, no. 7, pp. 1 – 7, 2014.

[6] H. Tong, "A note on support vector machines with polynomial kernels," *Neural Computation*, vol. 28, no. 1, pp. 71–88, Jan 2016.

[7] A. Shabnam, H. Behnoosh, and S. Shervin, "Driver drowsiness monitoring based on yawning detection," in *IEEE Instrumentation and Measurement Technology Conference*, 2011, pp. 1–4.

[8] R. Wang, L. Guo, B. Tong, and L. Jin, "Monitoring mouth movement for driver fatigue or distraction with one camera," in *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on*, Oct 2004, pp. 314–319.

[9] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *CVPR (1)*, 2001, pp. 511–518.

[10] P. Lv, Q. Zhao, and D. Gu, "Contour tracking via on-line discriminative active contours," in *Image Processing (ICIP), 2014 IEEE International Conference on*, Oct 2014, pp. 481–485.

[11] F. Precioso, M. Barlaud, T. Blu, and M. Unser, "Robust real-time segmentation of images and videos using a smooth-spline snake-based algorithm," *Image Processing, IEEE Transactions on*, vol. 14, no. 7, pp. 910–924, July 2005.

[12] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal Of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[13] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *International Journal of Computer Vision*, vol. 22, no. 1, pp. 61–79, 1997. [Online]. Available: http://dx.doi.org/10.1023/A%3A1007979827043

[14] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. Yezzi, "Conformal curvature flows: From phase transitions to active vision," *Archive for Rational Mechanics and Analysis*, vol. 134, no. 3, pp. 275–301, 1996. [Online]. Available: http://dx.doi.org/10.1007/BF00379537

[15] T. Chan and L. Vese, "Active contours without edges," *Image Processing, IEEE Transactions on*, vol. 10, no. 2, pp. 266–277, Feb 2001.

[16] K. Junmo, W. John, Y. Anthony, C. Mjdat, and S. Alan, "A nonparametric statistical method for image segmentation using information theory and curve evolution," *IEEE Trans. Image Processing*, vol. 14, pp. 1486–1502, 2005.

[17] J. Massen, K. Dusch, O. Eldakar, and A. Gallup, "A thermal window for yawning in humans: Yawning as a brain cooling mechanism," *Physiology and Behavior*, vol. 130, no. 0, pp. 145 – 148, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031938414001784

[18] B. Akrout and W. Mahdi, "A visual based approach for drowsiness detection," in *Intelligent Vehicles Symposium (IV), 2013 IEEE*, June 2013, pp. 1324–1329.

[19] B. Akrout and W. MAhdi, "Hypovigilance detection based on eyelids behavior study," *International Journal of Recent Contributions from Engineering Science and IT (iJES)*, vol. 1, no. 1, pp. 39–45, 2013. [Online]. Available: http://dx.doi.org/10.3991/ijes.v1i1.2927

[20] B. Akrout and W. Mahdi, "Vision based approach for driver drowsiness detection based on 3d head orientation," in *The 7th FTRA International Conference on Multimedia and Ubiquitous Engineering (MUE 2013)*. Seoul, Korea: Springer Netherlands, 2013.