Research Article

Driver's Fatigue Detection Based on Yawning Extraction

Nawal Alioua,[1] Aouatif Amine,[1,2] and Mohammed Rziza[1]

[1] LRIT Associated Unit to CNRST (URAC 29), Faculty of Sciences, University of Mohammed V-Agdal,
 4 Avenue Ibn Battouta, B.P. 1014, Rabat, Morocco
[2] LGS, ENSA, Ibn Tofail University, B.P 241, Kenitra, Morocco

Correspondence should be addressed to Nawal Alioua; nawal.alioua@yahoo.fr

The increasing number of traffic accidents is principally caused by fatigue. In fact, the fatigue presents a real danger on road since it reduces driver capacity to react and analyze information. In this paper we propose an efficient and nonintrusive system for monitoring driver fatigue using yawning extraction. The proposed scheme uses face extraction based support vector machine (SVM) and a new approach for mouth detection, based on circular Hough transform (CHT), applied on mouth extracted regions. Our system does not require any training data at any step or special cameras. Some experimental results showing system performance are reported. These experiments are applied over real video sequences acquired by low cost web camera and recorded in various lighting conditions.

1. Introduction

The increasing number of traffic accidents due to a diminished driver's vigilance level has become a serious problem for society. Statistics show that 20% of all the traffic accidents are due to drivers with a diminished vigilance level [1]. Furthermore, accidents related to driver hypovigilance are more serious than other types of accidents, since hypovigilant drivers do not take correct action prior to a collision. The active safety research focuses on studying the prevention of such accidents by developing systems for monitoring vigilance level and alerting the driver when he is not paying attention to the road.

Hypovigilance can be generally identified by sensing physiological characteristics, driver operations, or vehicle responses or monitoring driver's responses. Among these methods, the techniques based on human physiological phenomena are the most accurate. These techniques are implemented in two ways: measuring changes in physiological signals, such as brain waves, heart rate, and eye blinking and measuring physical changes such as the driver's head pose and the state of the eyes or the mouth. The first technique, while being most accurate, is not realistic since sensing electrodes would be attached directly on the driver's body and hence

would be annoying and distracting. In addition, long time driving would result in perspiration on the sensors, diminishing their ability to monitor accurately. The techniques based on measuring physical changes are nonintrusive and more suitable for real world driving conditions since they use video cameras to detect changes. Eye state analysis [2–6], head pose estimation [7, 8], and mouth state analysis [9, 10] are the most relevant physical changes allowing the detection of driver's hypovigilance. Driver operations and vehicle behaviors can be implemented by monitoring the steering wheel movement, accelerator or brake patterns, vehicle speed, lateral acceleration, and lateral displacement. These are also nonintrusive ways of detecting vigilance states, but they are limited to vehicle type and road condition. The final technique is monitoring the driver's response by requesting periodically from the driver to send a response in order to identify its alertness level. However, this technique will eventually become tiresome and annoying to the driver.

In this work, we focus our attention on detecting driver's fatigue from yawning, which is a good fatigue indicator. We propose a system aimed at identifying yawning by measuring physical changes occurring in driver's mouth using circular Hough transform (CHT). To do this, a normal video camera as a vision sensor is used to acquire face

and then extract the mouth if the environment is bright enough (daylight driving, e.g.). The remainder of this paper is organized as follows. In Section 2, we present some related works detecting driver's hypovigilance from physical changes. Section 3 explains the different steps of the proposed driver's yawning detection system. In Section 4, experimental results are exposed. Finally, conclusion is presented.

2. Related Works

Driver monitoring system based on facial feature analysis methods are divided into three categories: template based methods, appearance based methods, and feature based methods.

In template based methods, generic models of facial features are designed and template matching is used to find these models in the image. While these methods can provide good facial features detection, they are affected by image contrast and model initialization. The expensive time cost also prevents their wide application. Matched regions do not have always the same size or orientation as model; for this reason deformable templates [11] are used to enhance template matching approaches.

Appearance based methods detect facial features based on their photometric appearance. These methods need a large amount of data, representing features of different subjects in different conditions for training classifiers such as neural network (NN) or the support vector machine (SVM). Active contours model (Snake) [12] and active appearance models (AAM) [13] are the most used appearance based techniques. Zhu et al. [14] and Haro et al. [15] propose a real-time eye tracker based on bright pupil effect, eye appearance, and motion. In [16], an approach for locating eyes based on AdaBoost and fast radial symmetry is proposed. The trained eye-region detector is applied to segment eye regions using special gray distribution in this region. After getting eye region, fast radial symmetry operator utilizing eye blob information and eye neighbourhood information is used to precisely locate eye center. In [17], authors locate and track driver mouth movement using a dashboard-mounted CCD camera. They determine mouth region of interest by detecting face using color analysis and segmenting skin and lip pixels by Fisher classifier. After that, they detect mouth and extract lip features by connected component analysis. Tracking driver mouth via Kalman filtering in real time is then computed. Mouth region geometric features are taken to construct eigenvector as input of backpropagation neural network. Three different mouth states that represent normal, yawning, or talking state, respectively, are the outputs of this network.

Feature based methods identify some distinctive features of each facial characteristic. For example, nose is distinguished by nostril while eyes are characterized by color distribution of pupil, iris, and sclera. These kinds of methods use facial structure knowledge, such as Hough transform [18]. In [19] Zhou and Geng define a generalized projection function (GPF) for eye detection. Timm and Barth describe in [20] a feature based approach for eye centre localisation that can locate and track eye centres in low resolution images. For this, they introduce an approach for eye centre localisation, which defines the center of a semicircular pattern as the location where most of image gradients intersect. Therefore, they derive a mathematical function that reaches its maximum at the centre of circular pattern. By using this mathematical formulation a fast iterative scheme can be derived. After that, they incorporate prior knowledge about eye appearance and increase robustness. Then, postprocessing techniques are applied to reduce problems that arise in presence of glasses, reflections inside glasses, or prominent eyebrows. Fan et al. [21] locate and track a driver mouth movement to monitor yawning. They detect face using gravity-center template. Then, they locate left and right mouth corners by grey projection and extract texture features of mouth corners using Gabor wavelets. Finally, linear discriminant analysis (LDA) is applied to classify feature vectors to detect yawning.

3. Proposed Yawning Analysis Method

The aim of this study is to develop an algorithm for yawning detection to monitor driver fatigue level. The proposed system performs some steps before determining driver's state. The face is first extracted from the video frames. Then, the mouth localization is performed to isolate this component. Finally, the proposed method for fatigue detection is applied.

3.1. Face Extraction. The face is first extracted from the image frame to reduce the search region and therefore reduce the computational cost required for the subsequent steps. We have used an existing method, based on support vector machine (SVM) technique [22] developed by Romdhani et al. [23] and optimised by Franz et al. [24]. The basis of the method is to run an observation window at all positions, scales, and orientation in the image. A nonlinear SVM is applied to decide if a face is contained in the observation window. The nonlinear SVM operates by comparing the input patch to a set of support vectors (SV) which can be considered as face and antiface models. The SV is scored by a nonlinear function against the observation window. A face is detected if the resulting sum exceeds threshold. Since the search space is voluminous, a set of reduced vectors (RVs) is calculated from the SV to speed up the SVM. In fact, only a subset of the RVs is needed to eliminate nonface objects in order to reduce the computation cost.

The optimization introduced in [24] concerns the set of SVs replaced by a reduced set of synthesized input space points. Unlike the existing method that reduces the set via unconstrained optimization, a structural constraint is imposed on the synthetic points such that the resulting approximations can be evaluated via separable filters. To this end, the user-defined rank (i.e., the number of separable filters into which the RSVs are decomposed) provides a mechanism to control the tradeoff between accuracy and speed of the resulting approximation. The experiments provided by authors [24] show that the use of rank deficient RSVs leads to a speedup without losing accuracy. At run-time, rank
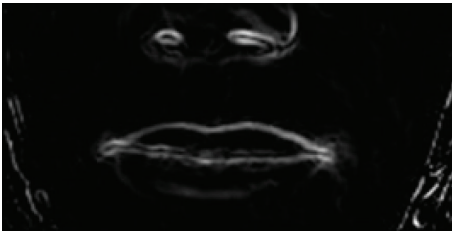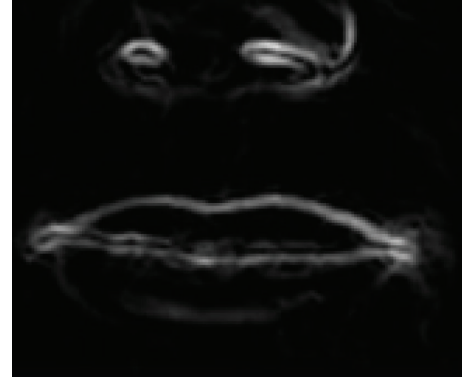
Figure 1: Face gradient.

Figure 3: Left and right mouth boundaries.
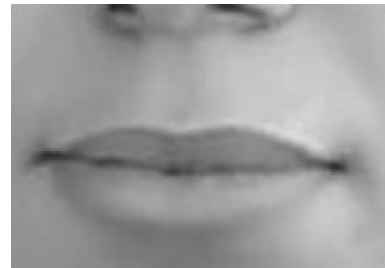

Figure 2: Mouth search region.

Figure 4: Localized mouth region.

deficient RSVs can be used together with unconstrained RSVs or SVs using the same canonical image representation.

3.2. Mouth Localization. We localise the mouth to eliminate the possibility of confusing it with other facial features, such as the eyes or the nose. To locate the mouth region, we perform the following steps.

(i) Detect facial edge using gradient edge detector (see Figure 1).

(ii) Compute vertical projection on the half lower face edge (see Figure 2) by $V_{\text{proj}}(j) = \sum_i \text{grad}(i, j)$, where grad symbolises the image gradient and i and j represent the ith row and jth column, respectively. This step aims to detect the right and the left mouth region boundaries (see Figure 3).

(iii) Compute horizontal projection to resulting gradient region according to $H_{\text{proj}}(i) = \sum_j (\text{grad}(i, j))$ to obtain the upper and lower limits of the mouth and then the mouth localized region (see Figure 4).

3.3. Circular Hough Transform. The Hough transform [25] is a transformation of a point from Cartesian space to the parameter space defined according to the shape of the object of interest. In the case of circular forms, the circle equation $r^2 = (x - a)^2 + (y - b)^2$ is considered for the transformation. Note that r represents the radius and a and b refer, respectively, to the abscissa and the ordinate of the circle center. The process of finding circles in an image

uses a modified Hough transform called circular Hough transform. The first step of this process is to find all edges in the image by an edge detection technique. At each edge point, we draw a circle having center in this point with the desired radius. This circle is drawn in the parameter space such that the x-axis is the a value and the y-axis is the b value while the z-axis represents the radius. To simplify the circle parametric representation, the radius can be fixed. We increment the value in the accumulator matrix at the coordinates belonging to the perimeter of the drawn circle. The accumulator has the same size of the parameter space. In this way we sweep over every edge point in the input image drawing circles with the desired radii and incrementing the values in the accumulator. When every edge point and every desired radius are considered, the accumulator contains the number of circles passing through the individual coordinates. Thus, the highest numbers correspond to the center of the circles in the image. Figure 5 illustrates the CHT from the Cartesian space to the parameter space.

3.4. Driver's Fatigue Analysis. The role of this step is crucial in our system, since it detects yawning in real time and issues immediately an alarm if the yawning frequency is high. To detect yawning, we apply CHT on mouth region images to identify wide open mouths. We consider that the driver is yawning if we find a significant number of consecutive frames where the mouth is wide open. As can be seen in previous section, CHT extracts circles from edge images. So, the obtained results depend on applied edge detector. Some well-known edge detectors such as Sobel, Prewitt, Roberts,
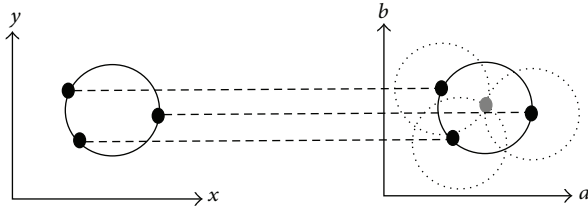
FIGURE 5: CHT from the Cartesian space to the parameter space.

	Closed mouth	Slightly open mouth	Widely open mouth
Original image			
Prewitt			
Sobel			
LoG			
Canny			
Our edge detector			

FIGURE 6: Comparison of mouth edge detectors.

Laplacian of Gaussian (LoG), and Canny are used to extract wide open mouth edges. The obtained edges by these filters did not provide the desired form, that is, a kind of circular form referring to the openness degree of the mouth. In order to solve this problem, we propose a new wide open mouth edge detector more suitable to the mouth morphology.

3.4.1. Wide Open Mouth Edge Detector. If we observe a wide open mouth, we see three or four main components. These are skin, lips, a dark region corresponding to the openness degree, and sometimes teeth. This distinguished structure enables us to extract the edge of the wide open mouth from significant intensity variations between the dark region and the lips or the teeth.

Our edge detector considers pixels x with a gray scale intensity lower than an optimal threshold in order to handle only with the pixels that can belong to the dark region. This threshold is computed from the mean intensity of the mouth

image. For each x, a neighborhood containing n pixels at the top and the bottom of x is specified. The n value is proportional to the number of image column. The intensity differences between x and its n top and bottom neighbors are then computed.

    (i) *Top (resp., bottom) edge*: if at least $n - 1$ top (resp., bottom) neighbors provide a high difference and also if at least $n - 1$ bottom (resp., top) neighbors are close to x, we deduct that x is a top (resp., bottom) edge pixel and we put it at 1.

    (ii) *Interpretation*: when x appertains to the top edge, its top (resp., bottom) neighbors pixel's intensity is very higher (resp., similar). Inversely, when x belongs to the bottom edge, the bottom (resp., top) neighbors pixel's intensity is very different (resp., similar).

Figure 6 shows some examples of mouth edge detection obtained by the proposed method compared to different classic edge detectors results. The classic edge detectors can not provide good wide open mouth edge detection. For example, some edge components having circular form are detected in closed or slightly opened mouths by classic edge detectors, while our edge detector did not identify such component. An interesting property of our edge detector is its ability to detect only wide open mouths. This property allows us to identify yawning without confusing it with talking, laughing, or singing, which correspond to small mouth opening.

3.4.2. Wide Open Mouth Detection by CHT. Once the appropriate edge detector is found, we can apply the CHT on this edge to obtain the radius of the wide open mouth from which we decide if the mouth is wide open or not. In the following, we present the CHT algorithm steps. At each iteration, three edge pixels are randomly chosen. If these pixels are not collinear and if the distance between each two pixels coordinates is higher than a fixed threshold, we compute the radius and center coordinates of the candidate circle defined by these three pixels. If the candidate circle parameters are between two thresholds, they are assigned to an accumulator. After that, we compute the distance between the center and all edge pixels. If this distance is low, we increment the counter of the candidate circle pixels. When the number of pixels is important, we consider that the candidate circle can represent a wide open mouth, we keep the other pixels as a new edge, and we repeat the previous steps. The algorithm stops when the edge contains few pixels (less than 30) or when maximal iterations are reached. Since we need to detect the circle representing the wide open mouth, we select the circle having the highest radius after the end of the algorithm. Figure 7 illustrates the CHT steps.

3.4.3. Fatigue Detection. Driver's fatigue is characterized by a high yawning frequency. If the mouth is widely open, the counter of consecutive wide open mouths is incremented. When the wide opening lasts more than 2 seconds, we consider that we have a yawning and the corresponding counter is incremented. Once this last counter is important,
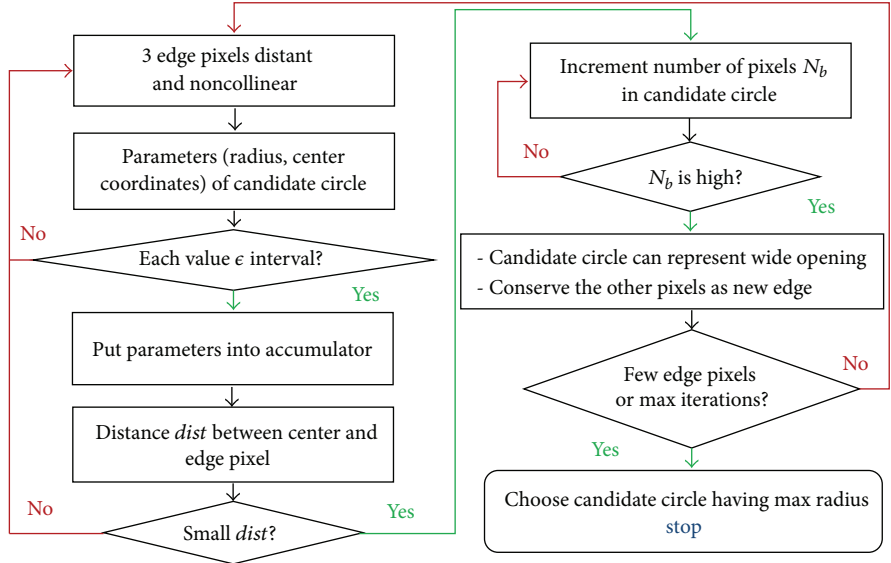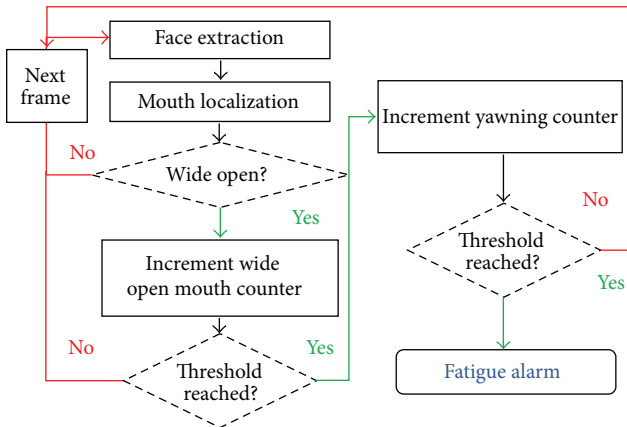
FIGURE 7: CHT steps.

FIGURE 8: Driver's fatigue detection system.

FIGURE 9: Acquisition system.

TABLE 1: Kappa statistic interpretation.

Kappa statistic	Interpretation
$\kappa > 0.81$	Almost perfect agreement
$0.61 < \kappa < 0.8$	Strong agreement
$0.21 < \kappa < 0.6$	Moderate agreement
$0.0 < \kappa < 0.2$	Poor agreement
$\kappa < 0$	Disagreement

the system indicated that the driver is suffering from fatigue by issuing a warning signal. Figure 8 illustrates our proposed system for driver's fatigue detection.

4. Experimental Results

The experimental results will be presented by confusion matrix composed of true positive (TP), false positive (FP), true negative (TN), false negative (FN), and total number of samples (T). We also use correct classification rate (CCR) and kappa statistic κ. κ is a measure of nonrandom agreement degree between observations of the same categorical variable interpreted by Table 1.

To validate our fatigue detection system, we acquire and annotate manually six video sequences representing real driving conditions. In each sequence (except in the last one), the driver is asked to simulate yawning. All sequences are taken with the same low cost web camera at 30 frames per second (fps) providing images of resolution 640×480. The web camera is fixed to the dashboard in order to avoid its movement and it is connected to a laptop with Intel Core 2 Duo Processor. Figure 9 shows the acquisition system. The experiment led us to reduce the considered number of fps from 30 to 3 to meet the real-time constraints. The automatic detections of faces and mouths are integrated and are taken into account in assessing the runtime system. In Figure 10, 6 video sequences are considered to expose the samples of TP, TN, FN, and FP. Table 2 presents the statistical measures. In this table, the two last columns are added to represent time

TABLE 2: Statistical measures of driver's fatigue detection.

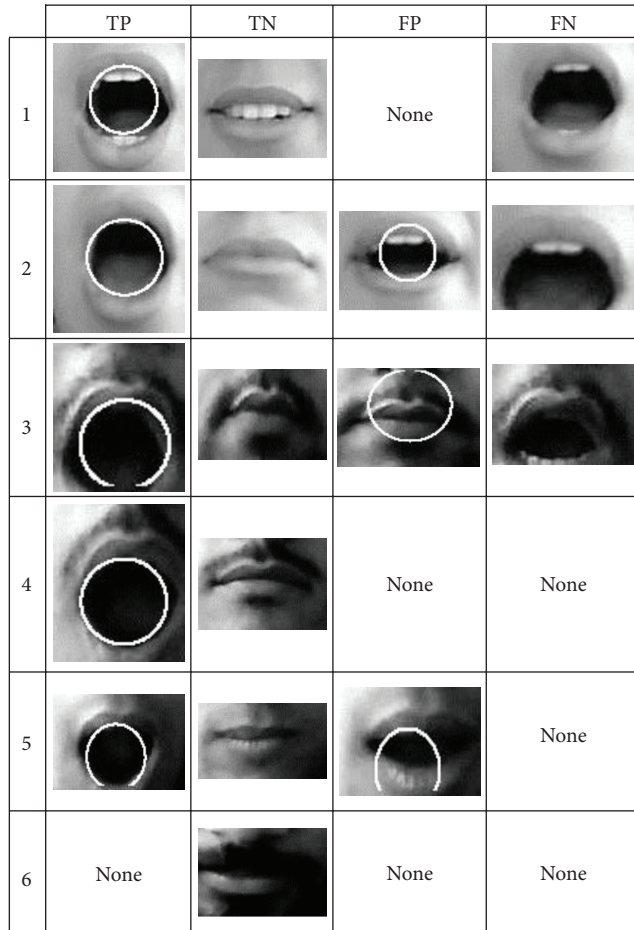| Video	TP	TN	FP	FN	T	CCR	Kappa	VidD	ExecT
1	46	132	0	2	180	0.98	0.97	61	63
2	56	155	1	5	217	0.97	0.93	72	75
3	14	108	1	1	124	0.98	0.92	41	46
4	46	96	0	0	142	1.00	1.00	47	52
5	28	90	1	0	119	0.99	0.97	40	44
6	0	132	0	0	132	1.00	1.00	44	47
					Average	0.98	0.96		

FIGURE 10: Results of driver's fatigue detection.

in seconds. The first one *VidD* refers to video duration and the second one *ExecT* refers to execution time of the whole system.

According to Table 2, the average of CCR is 98% and the average of κ is 96%. From Table 1 and this average, the wide open mouth detection method applied to fatigue detection procures an almost perfect agreement between the observers. This means that the method permits assignation of the right classes in most cases.

After comparing the two last columns, we deduce that the system respects the real-time constraints since execution time and video duration are almost the same. Thus, we deduct that

the proposed system can be used to have a good and real-time estimation of the driver's fatigue state.

We expose some results of existing systems permitting driver's hypovigilance detection. The system depicted in [5] is based on CHT to detect irises. This system uses 173 images of ORL database for tests and provides a success rate of 90.3%. The second system presented in [26] uses 70 images taken with an infrared camera for tests and obtains a success rate of 90%, while the third system [3], which is based on adaptive learning method to detect driver's eyes, uses real frames for tests and finds an accuracy of about 95%. The last system [10] detects fatigue using yawning analysis and tracking. The tests are effectuated on real images providing a CCR of 81% for yawning detection. We deduce that our proposed system provides a high success rate compared to the mentioned existing systems.

5. Conclusion

This paper presents a new approach for yawning detection algorithm for fatigue assessment based on CHT. The whole driver's fatigue detection system uses three steps: face detection using the SVM detector for face extraction, mouth region localization, and wide open mouth detection. In the last step, we apply the CHT on the results of the proposed wide open mouth edge detector. With 98% accuracy, it is obvious that our driver's fatigue detection system is robust. For future work an adequate attention will be given to the development of a system that combines more indicators including face expressions and head tracking to monitor driver hypovigilance.
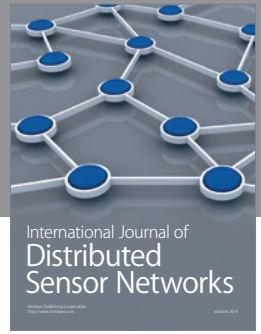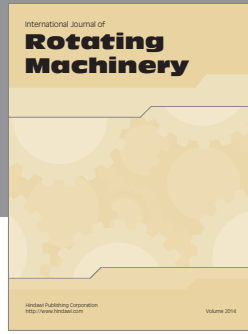
Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] L. Bergasa, J. Nuevo, M. Sotelo, and M. Vazquez, "Real-time system for monitoring driver vigilance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 63–77, 2006.

[2] N. P. Papanikolopoulos and M. Eriksson, "Driver fatigue: a vision-based approach to automatic diagnosis," *Transportation Research C: Emerging Technologies*, vol. 9, no. 6, pp. 399–413, 2001.

[3] G. Zhang, B. Cheng, R. Feng, and X. Zhang, "A real-time adaptive learning method for driver eye detection," in *Digital Image Computing: Techniques and Applications*, pp. 300–304, 2008.

[4] R. Grace, V. Byrne, D. Bierman et al., "A drowsy driver detection system for heavy vehicles," in *Proceedings of the 17th Digital Avionics Systems Conference*, vol. 2, pp. 136/1–136/8, 2001.

[5] D. Tripathi and N. Rath, "A novel approach to solve drowsy driver problem by using eye-localization technique using CHT," *International Journal of Recent Trends in Engineering*, vol. 2, no. 2, pp. 139–145, 2009.

[6] T. D'Orazio, M. Leo, P. Spagnolo, and C. Guaragnella, "A neural system for eye detection in a driver vigilance application," in *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems (ITSC '04)*, pp. 320–325, October 2004.

[7] P. Smith, M. Shah, and N. da Vitoria Lobo, "Monitoring head/eye motion for driver alertness with one camera," in *Proceedings of the 15th International Conference on Pattern Recognition (ICPR '00)*, vol. 4, pp. 636–642, Barcelona, Spain, 2000.

[8] T. Wang and P. Shi, "Yawning detection for determining driver drowsiness," in *Proceedings of the IEEE International Workshop on VLSI Design and Video Technology*, pp. 373–376, Suzhou, China, May 2005.

[9] M. Mohanty, A. Mishra, and A. Routray, "A non-rigid motion estimation algorithm for yawn detection in human drivers," *International Journal of Computational Vision and Robotics*, vol. 1, no. 1, pp. 89–109, 2009.

[10] M. Saradadevi and P. Bajaj, "Driver fatigue detection using Mouth and Yawning analysis," *International Journal of Computer Science and Network Security*, vol. 8, no. 6, 2008.

[11] A. L. Yuille, P. W. Hallinan, and D. S. Cohen, "Feature extraction from faces using deformable templates," *International Journal of Computer Vision*, vol. 8, no. 2, pp. 99–111, 1992.

[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] T. F. Coates, G. J. Edwards, and C. J. Taylor, "Active appearance model," in *European Conference on Computer Vision*, pp. 484–498, 1998.

[14] Z. Zhu, K. Fujimura, and Q. Ji, "Real-time eye detection and tracking under various light conditions," in *Proceedings of ETRA: Eye Tracking Research & Applications Symposium*, pp. 139–144, ACM Press, New York, NY, USA, 2002.

[15] A. Haro, M. Flickner, and I. Essa, "Detecting and tracking eyes by using their physiological properties, dynamics, and appearance," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '00)*, vol. 1, pp. 163–168, Hilton Head Island, SC , USA, June 2000.

[16] W. Zhang, H. Chen, P. Yao, B. Li, and Z. Zhuang, "Precise eye localization with AdaBoost and fast radial symmetry," in *Proceedings of the International Conference on Computational Intelligence and Security (ICCIAS '06)*, pp. 725–730, October 2006.

[17] W. Rongben, G. Lie, T. Bingliang, and J. Lisheng, "Monitoring mouth movement for driver fatigue or distraction with one camera," in *Proceedings of the 7th IEEE International Conference on Intelligent Transportation Systems*, pp. 314–319, October 2004.

[18] T. Kawaguchi, D. Hidaka, and M. Rizon, "Detection of eyes from human faces by Hough transform and separability filter," in *Proceedings of the International Conference on Image Processing (ICIP '00)*, pp. 49–52, Vancouver, Canada, September 2000.

[19] Z. Zhou and X. Geng, "Projection functions for eye detection," *Pattern Recognition*, vol. 37, no. 5, pp. 1049–1056, 2004.

[20] F. Timm and E. Barth, "Accurate eye centre localisation by means of gradients," in *Proceedings of the International Conference on Computer Vision Theory and Application (VISAPP '11)*, pp. 125–130, INSTICC, Algarve, Portugal, March 2011.

[21] X. Fan, B.-C. Yin, and Y.-F. Sun, "Yawning detection for monitoring driver fatigue," in *Proceedings of the 6th International Conference on Machine Learning and Cybernetics (ICMLC '07)*, vol. 2, pp. 664–668, Hong Kong, China, August 2007.

[22] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[23] S. Romdhani, P. Torr, B. Schölkopf, and A. Blake, "Computationally efficient face detection," in *Proceedings of the 8th International Conference on Computer Vision*, vol. 2, pp. 695–700, July 2001.

[24] M. Franz, W. Kienzle, G. Bakir, and B. Scholkopf, "Face detection-efficient and rank deficient," *Advances in Neural Information Processing Systems*, vol. 17, pp. 673–680, 2005.

[25] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, 1972.

[26] B. Hrishikesh, S. Mahajan, A. Bhagwat et al., "Design of drodeasys (drowsy detection and alarming system)," *Advances in Computational Algorithms and Data Analysis*, pp. 75–79, 2009.