2009 Ninth International Conference on Intelligent Systems Design and Applications

Computer Vision-Based Eyelid Closure Detection: a Comparison of MLP and SVM classifiers

D. González-Ortega, F. J. Díaz-Pernas, M. Antón-Rodríguez,
M. Martínez-Zarzuela, J. F. Díez-Higuera, D. Boto-Giralda
Department of Signal Theory, Communications and Telematics Engineering, Telecommunications Engineering School, University of Valladolid
Valladolid, Spain
E-mails: {davgon, pacper, mirant, marmar, josdie, danbot}@tel.uva.es

***Abstract*—In this paper, a vision-based system to detect the eyelid closure for driver alertness monitoring is presented. Similarity measures with three eye templates (open, nearly close, and close) were calculated from many different features, such as 1-D and 2-D histograms and horizontal and vertical projections, of a big set of rectangular eyes images. Two classifiers, Multi-Layer Perceptron and Support Vector Machine, were intensively studied to select the best with the sequential forward feature selection. The system is based on the selected Multi-Layer Perceptron classifier, which is used to measure PERCLOS (percentage of time eyelids are close). The monitoring system is implemented with a consumer-grade computer and a webcam with passive illumination, runs at 55 fps, and achieved an overall accuracy of 95.75% with videos with different users, environments and illumination. The system can be used to monitor driver alertness robustly in real time.**

Keywords—eyelid closure detection, multi-layer perceptron, support vector machine, sequential forward selection, driver alertness monitoring.

I. INTRODUCTION

Statistics show that between 10% and 20% of all the traffic accidents in Europe are due to drivers with a reduced vigilance level caused by fatigue [1]. These figures show the importance that driver alertness monitoring applications can have to decrease the number of traffic accidents. Fatigue measurement is a difficult problem as there are few direct measures and most of them are measures of the outcomes of the fatigue rather than of fatigue itself. An important physiological measure that has been studied to detect fatigue is eye motion. Several eye motions were used to measure fatigue such as blink rate, blink duration, long closure rate, blink amplitude, saccade rate and peak saccade velocity. PERCLOS measure is the percentage of eyelid closure over the time and reflects slow eyelid closures rather than blinks [2]. A PERCLOS drowsiness metric was established as the proportion of time in a minute that the eyes are at least 80 percent closed [2] and it is a reliable and widely used driver performance and bio-behavioral measure [3].

Many vision-based methods have been proposed for eye tracking and eyelid closure detection. They can be based on active infrared or passive illumination. With infrared illumination, pupils can be detected by a simple thresholding of the difference between the dark and the bright pupil images [4] and factors such as the brightness, size of the pupils and the external illumination can influence performance. Approaches using standard cameras and passive illumination in cluttered scenes have also been presented. Orazio et al. [5] proposed a neural classifier to recognize the eyes in the image selecting two candidates regions that might contain the eyes by using iris geometrical information and symmetry. Smith et al. [6] presented a system to detect eye blinking and eye closure based on color statistics. Królak et al. [7] proposes a system which uses two active contours, one from each eye, for eye blink detection from previous eye tracking.

In this paper, we present a robust real-time system to detect the eye state of the driver with a consumer-grade computer, an inexpensive Universal Serial Bus camera, and passive illumination. While previous approaches have proposed many different methods and features to fulfill eye and eyelid closure detection, our contribution focuses on the calculation of multiple features of different nature in the eye region to distinguish three eye states, open, nearly close and close, in the eye regions. The use of three eye states is necessary to calculate PERCLOS as in [2] accurately. We accomplished the discrimination among these states of a non-rigid object with high variability such as the eye with a multiple-feature scheme. Another contribution of the work is the performance comparison of two classifiers, the Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM), with features selected with Sequential Forward Selection (SFS), in the eye state detection task.

The rest of the paper is organized as follows. Section II presents our approach to detect the eye state. Experimental results are described in Section III. Finally, Section IV draws the conclusions about the proposed driver alertness monitoring system.

II. OUR APPROACH

We aimed to monitor the driver alertness with the PERCLOS measure. PERCLOS is meaningful in highway driving when the driver face is mainly in frontal position with respect to a camera placed on the car dashboard and there is low head motion. For the face detection, we use the

978-0-7695-3872-3/09 $26.00 © 2009 IEEE
DOI 10.1109/ISDA.2009.226

frontal face detector based on the AdaBoost algorithm and with Haar-like features available in the framework of the Intel Open Source Computer Vision Library (OpenCV) [8]. This detector has a great performance with frontal and near frontal faces.

Rectangular pair of eyes region location is determined using face and head anthropometry measures from the output given by the face detection. As the face detection is mainly based on the eyes, they can be accurately located. Assuming that eyelid closure is simultaneous in both eyes, the eye state has to be determined in each frame of a video sequence to compute the PERCLOS measure. Therefore, we had to discriminate among open and close eye state and also the nearly close state, which is characteristic of a high level of drowsiness of the driver. In this state, the proportion of visible iris is below 20% of its total height. Due to the large variability of the eye appearance and dynamics, our approach was not to extract an individual feature or a small set of features as individual features are influenced by the fact that a feature with a very wide class of invariance looses the power to discriminate among other differences. On the contrary, we have extracted many features of different nature to achieve robustness to illumination, presence or absence of structural components such as glasses, and facial expression.

The extracted features were: 1-D grayscale histograms, 2-D color histograms, horizontal and vertical projections, and the entire rectangular eye images. Regarding grayscale images, an eye can be characterized by the intensity of two regions, one corresponds to the iris and the other to the sclera so 1-D grayscale histograms are extracted. 2-D histograms of the chromaticity components (HS) of the HSV color space were extracted to discriminate between an open and a close eye with some illumination invariance as skin colors have a certain invariance regarding chromaticity components [9]. Horizontal and vertical projection functions of a grayscale image I(x,y) in intervals $[y_1,y_2]$ and $[x_1,x_2]$ are expressed in (1) and (2), respectively. Fig. 1 shows the normalized vertical projections of open, nearly close, and close eye images in the range [0,255]. Zhou et al. [10] proposed the horizontal and vertical generalized projection functions as expressed in (3) and (4), where $IPF_h^{'}$ and VPF_h depend on IPF_h , $IPF_v^{'}$ and VPF_v depend on IPF_v and $0 \le \alpha \le 1$ is used to control the relative contribution of $IPF^{'}$ and VPF . We calculated IPF_h , IPF_v and GPF_h and GPF_v of the rectangular eye regions with α values of 0, 0.4, 0.6 and 1.

$$IPF_h\left(y\right)=\sum_{x=x_1}^{x_2} I(x,y) \tag{1}$$

$$IPF_v\left(x\right)=\sum_{y=y1}^{y2} I(x,y) \tag{2}$$

$$GPF_h\left(y\right)=\left(1-\alpha\right)\cdot IPF_h^{'}\left(y\right)+\alpha\cdot VPF_h\left(y\right) \tag{3}$$

$$GPF_v\left(x\right)=\left(1-\alpha\right)\cdot IPF_v^{'}\left(x\right)+\alpha\cdot VPF_v\left(x\right) \tag{4}$$

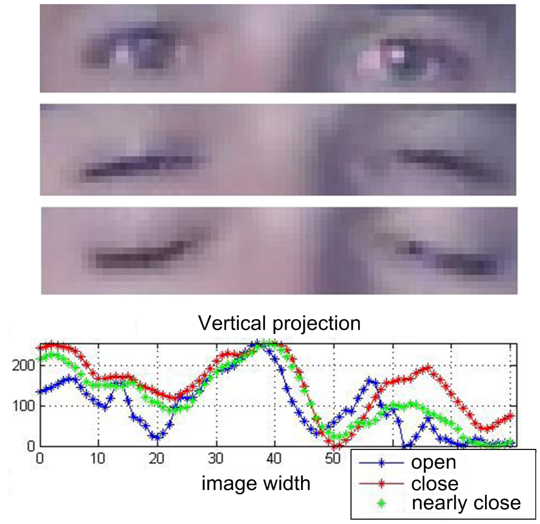


Figure 1. Vertical projections of open, nearly close, and close eye images.

The entire image is also used to measure the similarity with the templates.

Once we have the information of an eye image captured in a feature set, there are two possibilities from endowing them with meaning: make a unilateral interpretation from the feature set or compare the feature set with some elements on the basis of a similarity function. As proposed in [11], the complexity of the eye detection problem makes it necessary the use of a similarity function to achieve the desired accuracy in analyzing eyelid closure with non-strictly frontal face, in motion, and with illumination changes. The elements to compare the feature set of an eye image are the templates of open eye, nearly close eye and close eye.

A. *Similarity functions*

With the information obtained from the eye regions, a big number of similarity measures between the eye regions and the three templates were extracted. Our approach is to compute a set of different similarity measures so that the best measures be selected later. First, we obtained four histogram-based similarity measures: correlation, χ^2 , intersection and Bhattacharyya distance. For two histograms $Q=\{q_i\}$ and $D=\{d_i\}$, each one containing n bins, the similarity measures are defined as follows.

The correlation distance is defined as

$$d_C(Q,D)=\frac{\sum_{i=1}^{n} q_i^{'}\cdot d_i^{'}}{\sqrt{\sum_{i=1}^{n} q_i^{'2}\cdot d_i^{'2}}} \tag{5}$$

$h_i^{'}$ for a histogram $H=\{h_i\}$ with n bins is expressed as

$$h_i' = h_i - \frac{\sum_{k=1}^{n} h_k}{n} \tag{6}$$

The χ^2 distance is a statistical index showing how likely is for one distribution to get drawn from the population represented by the other and is defined as

$$d_{\chi^2}(Q,D) = \sum_{i=1}^{n} \frac{(q_i - d_i)^2}{q_i + d_i} \tag{7}$$

The histogram intersection is defined as

$$d_{HI}(Q,D) = \sum_{i=1}^{n} min(q_i, d_i) \tag{8}$$

A modified version of the Bhattacharyya distance was used, defined as

$$d_B(Q,D) = \sqrt{1 - \sum_{i=1}^{n} \frac{\sqrt{q_i \cdot d_i}}{\sqrt{\sum_{i=1}^{n} q_i \cdot \sum_{i=1}^{n} d_i}}} \tag{9}$$

The following four template matching methods to measure the similarity are applied to the entire eye regions and the projection functions of the regions: correlation matching, correlation coefficient matching and the normalized versions of the two matching methods. Unlike the histogram-based measures, these methods compute the spatial distribution of pixels. The correlation matching method multiplicatively matches the template T against the image I to obtain the matching result as expressed in (10). The correlation coefficient matching method R_{ccoeff} matches a template relative to its mean against the image relative to its mean. The normalized version of the correlation and the correlation coefficient matching methods are obtained dividing them by the same normalization coefficient. They are useful as they can help reduce the effects of lighting differences between the template and the image. Fig. 2 shows the correlation (R_{cor}) and the correlation coefficient matching (R_{ccoeff}) similarity measures between the vertical projections of 150 open eye video frames and the three templates. Only the correlation matching method has clear discriminative values depending on the template (open, nearly close and close) for almost all the frames in Fig. 2.

Each similarity measure between an eye region and a template represents an eye region feature.

$$R_{cor}(x,y) = \sum_{x',y'} [T(x',y') \cdot I(x+x', y+y')]^2 \tag{10}$$

B. Feature Selection and Classifiers

After computing a big number of features, dimensionality reduction is necessary. Many features are not suitable to classify eye images because noisy and redundant inputs can have a bad influence on the classification performance [12]. Dimensionality reduction approaches can be classified in feature extraction and feature selection algorithms. Unlike feature extraction, feature selection allows to make inferences on how input variables affect the model results. We have adopted the suboptimal SFS method, which selects the best single feature and then adds one feature at a time which in combination with the previously selected features maximizes the criterion function. SFS is computationally efficient [13].

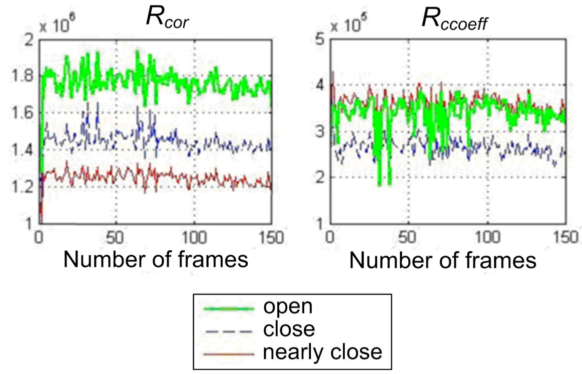


Figure 2. R_{cor} and R_{ccoeff} similarity measures between the vertical projections of 150 open eye images and the three templates.

We used and compared two classifiers widely used in data classification and pattern recognition: MLP and SVM. MLP is a type of Artificial Neural Network (ANN). ANNs, which are inspired by biological neural networks, are composed of neural-like units connected together through input and output paths which have adjustable weights [14]. The MLP is an ANN which has been very successful in a variety of applications, producing results that are at least competitive and often exceed other existing approaches. SVM is an algorithm based on the margin-maximization principle which has become increasingly popular for a wide range of machine learning tasks, such as face recognition and text categorization, and has good generalization ability.

III. Experimental Results

The experimental results include several phases: the calculation of features of a big number of eye images, the performance comparison of MLP and SVM classifiers with the SFS, and finally the testing of the eye state monitoring system.

A. MLP and SVM comparative results

We collected 792 rectangular eye images with different people and illumination conditions in the three states: open, nearly close and close. Fig. 3 shows examples of images in the three states. 1-D and 2-D histograms, horizontal and vertical projection functions (IPF_h and IPF_v), and horizontal and vertical generalized projection functions (GPF_h and GPF_v with α values of 0, 0.4, 0.6 and 1) of these eye images were calculated. Then, 180 similarity measures between the image histograms, image projection functions, and the entire images and those from the templates of open, nearly close and close eye regions were computed. Different templates were selected for each person presented in the 792 eye images. The 180 similarity measures are divided in: 24 similarity measures (#1-24) between the 1-D histograms of the grayscale eye images and templates, 24 similarity measures (#25-48) between the 2-D histograms of the HS components in the HSV color space of the eye images and templates, 120 similarity measures (#49-168) between the projection functions of the grayscale eye images and templates, and 12 similarity measures (#169-180) from the matching of the eye images and the templates. The 180 similarity measures are the features of each rectangular eye image.

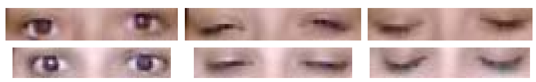


Figure 3. Eyes images in the three states: open, nearly close and close.

With the 180 similarity measures of the 792 rectangular eye images and SFS, an MLP was trained and validated using the leave-one-out method, which consists in averaging the result of doing as many training stages as images, leaving out only one image from the training set which is used to cross-validate. Backpropagation training algorithm and one hidden layer were used. In the training stage, the nearly close and close eyes are assigned the same output (close eye output) according to the PERCLOS measure. Fig. 4(a) shows the performance of the MLP as a function of the number of selected features. Three performance measures appear in Fig. 4(a): overall accuracy, specificity, and sensitivity, which characterize a binary classification. True positives are the nearly close or close eyes correctly classified by the network and false positives are the open eyes wrongly classified by the network. The maximum overall accuracy is achieved with 17 features. These 17 features have numbers from the most to the least discriminative: #15, 11, 30, 40, 44, 17, 21, 101, 121, 177, 4, 125, 123, 157, 126, 165, and 127. We used 501 additional images, different from the previous 792, to select the optimal number of features using the Receiver Operating Characteristics (ROC) curves. Fig. 5 shows two ROC curves and their Area Under the ROC Curve (AUC), which is the common method to compare classifiers [15]. The biggest value of AUC (0.9389) occurs with 15 features. The most sensitive parameters that influenced the MLP performance were the number of neurons in the hidden layer and the number of training iterations. Fig. 4(b) shows the MLP performance as a function of the number of neurons in the hidden layer. With 15 features selected with the SFS and five iterations, the best performance was 95.91%, achieved with five neurons in the hidden layer.

Then, SVM was trained and validated with the 180-feature set of the first 792 rectangular eye images using the leave-one-out method and SFS. Fig. 6(a) shows the performance of SVM as a function of the number of selected features. Unlike the 15 features selected with the MLP, only 6 features, with numbers #16, 12, 32, 28, 27, and 4, reached an SVM overall accuracy not significantly improved with more features. The SVM classifier requires the choice of kernel, the kernel-associated parameter (γ) and the ν parameter. In the same way as with MLP, 501 eye images, different from the previous 792, were used to select the SVM parameters. We chose the RBF as the kernel function. A set of classifiers were created with a range of values of γ, ν and the number of training iterations. Fig. 6(b) shows the SVM performance as a function of the γ parameter. Optimal value of γ was 56. Optimal value of ν was 0.1. The number of training iterations is important to avoid overfitting and 32 was the optimal value. Finally the SVM classifier achieved an overall accuracy of 94.28% with leave-one-out cross validation, which is outperformed by the 95.91% overall accuracy of the MLP classifier with 15 features.

The MLP classifier with 15 features was selected for the eye state monitoring system.

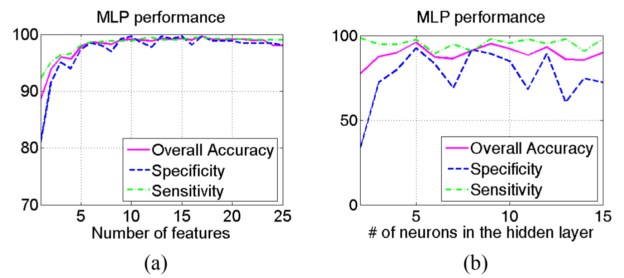


Figure 4. MLP performance for the eye state detection as a function of: (a) the number of features selected with SFS; (b) the number of neurons in the hidden layer.

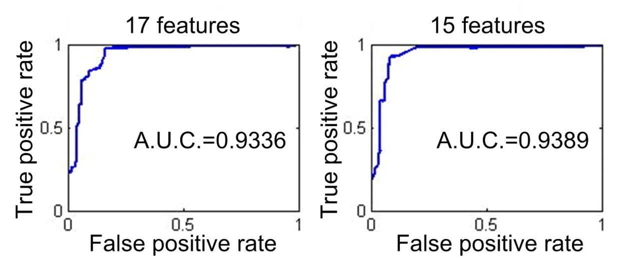


Figure 5. ROC curves of the MLP classifier for eye state detection using 17 and 15 features.

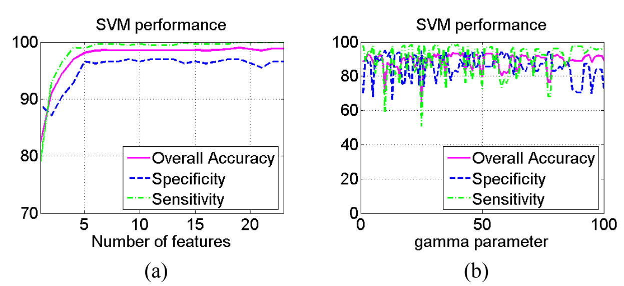


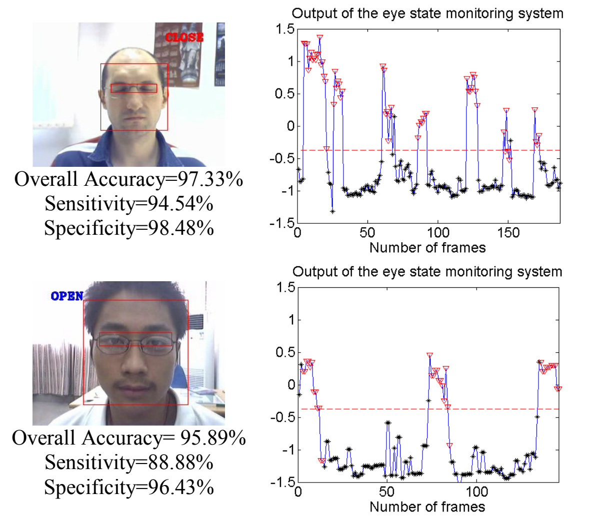
Figure 6. SVM performance for the eye state detection as a function of: (a) the number of features selected with the SFS; (b) gamma (γ) parameter.

B. Eye state monitoring system

We tested the system with eight videos, four from our private database taken in the laboratory and inside a car while a user is driving in real conditions and four from the publicly available ZJU Eyeblink database [16]. There are not public databases with videos recorded inside a car in real conditions to our knowledge. The eight videos have an image resolution of 320×240 pixels, with a face in frontal or near frontal position and a wide range of conditions: translation and scale changes, some people wear glasses and different illumination conditions. Fig. 7 shows the output of the eye state monitoring system for two videos, one from our private database and the other from the ZJU Eyeblink database. The dashed lines indicate the threshold value of the MLP classifier, which is -0.375. If the output is bigger than the threshold, the eyes are classified as close, otherwise they are classified as open. In Fig. 7, the manually ground truth eye state is indicated with a triangle, if the eyes are closed or nearly closed, or with an asterisk, if the eyes are open, in each frame of the video sequences. Overall accuracy, sensitivity and specificity are also indicated in Fig. 7. Errors mainly occur in the transitions between the close or the nearly close and the open eye state due to the inherent difficulty in classifying an eye when it is not completely open. This type of errors will not be significant for the driver alertness monitoring with the PERCLOS measure, unlike the situation of nearly close eyes being wrongly detected as open eyes in many consecutive frames. The average overall accuracy, sensitivity and specificity are 95.75%, 93.38%, 96.50%, respectively, for all the eight videos.

Processing times were taken with an Intel Core 2 Duo processor at 3 GHz and 4 GB RAM, with the software implemented in Visual C++. Each frame takes 18 ms on average, out of which 14 ms are taken for the face detection and the remaining time is mainly devoted to the eye region feature calculation and the classification. AdaBoost-based face detection is applied to image subwindows varying in position and size from a minimum size. We adapted this minimum size depending on the face size in the previous frame, with width and height 2/3 the face width and height in the previous frame. These processing times give rise to a frame rate of 55 frames per second (fps) so that real-time driver vigilance monitoring is achieved. Frame rate was limited by the 30 fps given by the two cameras used in the experiments: Logitech Quick Cam Zoom and Philips SPC315NC.

Fig. 8 shows frames of two video sequences, one from the Eyeblink database and one in real driving conditions. Face detection and eye region location are drawn in the frames together with the output of our eye state monitoring system. Our system monitors the driver alertness through the PERCLOS measure. The maximum value of the PERCLOS measure allowed for secure driving can be adapted to each driver and the level of alertness required for driving (high, medium, low). As soon as the threshold value is overcome, alarm state will be triggered. The driver will have to react and return to a secure driving state to abandon the alarm state.



(a) (b)

Figure 7. Output of the eye state classifier for two video sequences: (a) a frame from the video sequences and system performance figures; (b) output of the eye state monitoring system.

Figure 8. Output of the driver alertness monitoring system with two video sequences.

IV. CONCLUSIONS

In this paper, a robust real-time computer vision-based system to detect the eyelid closure for driver alertness monitoring is presented. Eye state detection is based on an MLP classifier using as pattern features different similarity measures between the eye region and one of the three templates: open, nearly close and close eyes. The calculation of the PERCLOS drowsiness measure makes it necessary the use of the three templates. 15 features were selected with the SFS and the AUC of the ROC curves: one is image-based, eight are histogram-based, and six are projection-based similarity measures. MLP outperforms SVM in our classification problem. It shows the MLP competitiveness in a wide range of applications. The selection of features of different nature stressed the complexity of characterizing such a deformable and highly variable object as the eye.

Our driver alertness monitoring system was tested with 8 videos. Average overall accuracy was 95.75% with no significant decrease in accuracy with oriental people of the videos from the ZJU Eyeblink database and the videos recorded in real driving conditions. This states the robustness of the system to changes of people, illumination, and environment. The system can be used to compute PERCLOS measure together with the consecutive time of non-face and close eye state detection for monitoring driver vigilance. It takes 18 ms per frame with 320×240 pixels. Our system runs in a standard computer and camera, which makes it easy the implementation of the vision-based security system inside a car.

ACKNOWLEDGMENT

This work was partially supported by the Spanish Ministry of Science and Innovation under project TIN2007-67236 and by the Education Department of Junta de Castilla y León (Spain) under project VA045A07.

REFERENCES

[1] Awake Consortium (IST 2000-28062), "AWAKE - System for effective assessment of driver vigilance and warning according to traffic risk estimation", 2001–2004. http://www.awake-eu.org.

[2] D.F. Dinges and R. Grace, "PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance", U.S. Department of Transportation, Federal Highway Administration, Office of Motor Carriers, Publication number FHWA-MCRT-98-006, 1998.

[3] Q. Wang, J. Yang, M. Ren, and Y. Zheng, "Driver fatigue detection: a survey", Proc of the World Congress on Intelligent Control and Automation, pp. 8587-8591, 2006.

[4] Q. Ji, Z. Zhu, and P. Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue", IEEE Trans. Vehicular Technol., vol. 53, no. 4, pp. 1052-1068, 2004.

[5] T. D'Orazio, M. Leo, C. Guaragnella, and A. Distante, "A visual approach for driver inattention detection", Pattern Recognit., vol. 40, no. 8, pp. 2341-2355, 2007.

[6] P. Smith, M. Shah, and N. da Vitoria Lobo, "Determining driver visual attention with one camera", IEEE Trans. Intell. Transp. Syst., vol. 4, no. 4, pp. 205-218, 2003.

[7] A. Królak and P. Strumillo, "Vision-based eye blink monitoring system for human-computer interfacing", Proc. of the Conf. on Human Systems Interactions, pp. 994-998, 2008.

[8] Open Computer Vision Library. http://sourceforge.net/projects/opencvlibrary, 2009.

[9] L. Sigal, S. Sclaroff, and V. Athitsos, "Skin color-based video segmentation under time-varying illumination", IEEE Trans. Pattern Analysis Machine Intell., vol. 26, no. 7, 2004.

[10] Z.-H. Zhou and X. Geng, "Projection functions for eye detection", Pattern Recognit., vol. 37, no. 5, pp. 1049-1056, 2004.

[11] G.R. Bradski, "Real time face and object tracking as a component of a perceptual user interface", Proc. of the IEEE Workshop on Appl. of Computer Vision, pp. 214-219, 1998.

[12] P.A. Devijver and J. Kittler, "Pattern Recognition: A Statistical Approach", Prentice Hall, 1982.

[13] A.K. Jain, R.P.W. Duin, and J. Mao, "Statistical pattern recognition: a review", IEEE Trans. Pattern Analysis Machine Intell., vol. 22, no. 1, pp. 4-37, 2000.

[14] S. Haykin, "Neural networks and learning machines", Prentice Hall, 3rd ed., 2008.

[15] T. Fawcett, "An introduction to ROC analysis", Pattern Recognit. Letters, vol. 27, no. 8, pp. 861-874, 2006.

[16] G. Pan, L. Sun, Z. Wu, S. Lao. "Eyeblink-based anti-spoofing in face recognition from a generic webcamera". Proc. of the IEEE Int'l Conf. on Computer Vision, pp. 1-8, 2007.