Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 40 (2007) 2341–2355
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

A visual approach for driver inattention detection

T. D'Orazio[a,*], M. Leo[a], C. Guaragnella[b], A. Distante[a]

[a]*Institute of Intelligent Systems for Automation, C.N.R. Via Amendola 122/D-I, 70126 Bari, Italy*
[b]*DEE-Electrics and Electronics Department, Politecnico di Bari, via E. Orabona 4, Bari 70125, Italy*

Received 25 July 2006; received in revised form 6 December 2006; accepted 16 January 2007

Abstract

Monitoring driver fatigue, inattention, and lack of sleep is very important in preventing motor vehicles accidents. A visual system for automatic driver vigilance has to address two fundamental problems. First of all, it has to analyze the sequence of images and detect if the driver has his eyes open or closed, and then it has to evaluate the temporal occurrence of eyes open to estimate the driver's visual attention level. In this paper we propose a visual approach that solves both problems. A neural classifier is applied to recognize the eyes in the image, selecting two candidate regions that might contain the eyes by using iris geometrical information and symmetry. The novelty of this work is that the algorithm works on complex images without constraints on the background, skin color segmentation and so on. Several experiments were carried out on images of subjects with different eye colors, some of them wearing glasses, in different light conditions. Tests show robustness with respect to situations such as eyes partially occluded, head rotation and so on. In particular, when applied to images where people have eyes closed the proposed algorithm correctly reveals the absence of eyes. Next, the analysis of the eye occurrence in image sequences is carried out with a probabilistic model to recognize anomalous behaviors such as driver inattention or sleepiness. Image sequences acquired in the laboratory and while people were driving a car were used to test the driver behavior analysis and demonstrate the effectiveness of the whole approach.

Keywords: Iris detection; Behavior analysis; Neural network; Mixture Gaussian model

1. Introduction

The detection of driver visual attention is very important for developing automatic systems that monitor driver inattention, driver fatigue, and sleepiness. A great number of fatalities occurring in motor vehicles could be avoided if these behaviors were detected and alarm signals were provided to the driver. The literature reports many attempts to develop safety systems for reducing the number of automobile accidents: these systems detect both the "driving" behavior by monitoring lane keeping, steering movements, acceleration, braking and gear changing [1], and also the "driver" behavior by such means as tracking the driver's head and the eye movements, monitoring the heart and breathing rates, the brain activity [2], and recognizing the torso and arm/leg motion [3]. Repeated experiments have shown that among all driver performance and bio-behavioral measures tested, the percentage of eyelid closure over time (Perclos) reliably predicts the most widely recognized psychophysiological index of loss of alertness. In this work we address the problem, crucial for automotive applications, of developing a robust eye recognition algorithm that can be used for detecting the Perclos measure but also for modelling different driver behaviors.

1.1. Related works

Many works presented in the literature propose real time methods for eye tracking based on active infrared (IR) illumination approaches. Some commercial products are coming to the market such as SmartEye and EyeAlert Fatigue Warning System [4,7] and also many Institutions are actively involved in automotive research projects. Different types of IR light sources have been devised to emit non-coherent energy synchronized with the camera frame rate, which generates bright and dark pupil images. Pupils can be detected by a simple thresholding

* Corresponding author. Tel.: + 39 080 5929442; fax: +39 080 5929460.
E-mail address: dorazio@ba.issia.cnr.it (T. D'Orazio).

0031-3203/$30.00
doi:10.1016/j.patcog.2007.01.018

of the difference between the dark and the bright pupil images [9–15]. The success of these active approaches depends on several factors: the brightness and size of the pupils, which are often functions of face orientation; external illumination interference; the distance of the subject from the camera; the need for stable lighting conditions (not strong sun light). In addition, glasses tend to disturb the IR light so much that the red eye effect may appear very weak. In recent years large improvements have been made in miniaturizing the cameras with the compact IR illuminator, in designing configurations that produce diffuse lighting, and in selecting allowable levels of IR irradiation. However, in many cases an initial calibration phase is still required during which the intensity of the active IR illuminators has to be tuned in order to operate in different natural light conditions, multiple reflections of glasses, and variable gaze directions.

Alternative approaches that use standard cameras with classical algorithms for eye detection in cluttered images have also been investigated [8]. SeeingMachine [5] proposes a couple of stereo cameras to determine the 3D position of matching features on the driver face. Starting from an initial calibration of the subject in a green room, the extracted 3D features are then used to capture the 3D pose of the person's face as well as the eye gaze direction, blink rates and eye closure [6]. Instead, in our work we investigate the subject of eye detection with monocular images within the visual spectrum and normal illumination. Generally the eye detection algorithms in cluttered images require two steps: locating face to extract eye regions and then eye detection from eye windows. The face detection problem has been addressed with different approaches: neural networks, principal components, independent components, skin color based methods, face models [16–18]. Each imposes some constraints: frontal view, expressionless images, limited variations of light conditions, hairstyle dependence, uniform background, and so on. An exhaustive review has been presented in Ref. [19].

Many works on eye or iris detections, assume either that eye windows have been extracted or rough face regions have already been located [20–22,24–27]. In Ref. [21], the eye detection method is performed within the possible eye region of the candidate face field. In this case it can be applied only after a face detection system has extracted a small number of candidates for eye regions. Left and right eye templates were used to detect eyes by a method that is unaffected by slight rotation, scaling, and translation (up to 10%). The algorithm proposed in Ref. [20] requires the detection of face regions in order to extract intensity valleys. Only at this point do the authors apply a template matching technique to extract iris candidates. The authors deal with the difficult problem of face region extraction both in intensity images and in color images.

Skin color models are strictly related to the considered images and cannot be so general as to be applied in every light condition and with different colors of skin. Region-growing methods or head contour methods on intensity images require strong constraints such as plain background. In Refs. [22,24] the first step is also a face detection algorithm, based on skin color segmentation of the input image with the constraints of there being only one face and a simple background; then the facial feature segmentation is based on gray value reliefs in Ref. [22] or on template matching of edge and color features in Ref. [24]. A probabilistic framework is used in Ref. [23] for locating precisely eyes in face areas extracted formerly by using a face detector. In Ref. [25] linear and non-linear filters have been used for eye detection: oriented Gabor wavelets form an approximation of the eye in gray level images; non-linear filters are applied to color images to determine the color distribution of the sclera region. In both cases a face detection step is applied, which assumes the face as the most prominent flash tone region in the image. The same algorithm has been used in Ref. [26] for tracking iris and eyelids in video.

In Ref. [27], lip and skin color predicates are used as a first step to segment lip region and skin regions in the image: the two holes above the lip region that satisfy some fixed size criteria are selected as the candidate for the eyes region. A hierarchical strategy is applied to track the eyes in a video sequence and evaluate the driver visual attention by finite state automata. In Ref. [28] the authors make use of multicues extracted from a gray level image to detect the eye windows within an a priori detected face region. The precise iris and eye corner locations are then detected by a variance projection function and an eye variance filter. In Ref. [29], face regions are also initially determined by using rules derived from quadratic polynomial models, then eye components are extracted after the segmentation of skin pixels and lips. In Ref. [30] the distribution of discriminant features that characterize eye patterns are learnt statistically and provided to probabilistic classifiers to separate eyes and non-eyes. Also in this case the eye localization method requires a face detection step based on hierarchical classifiers [31].

The use of eye detection algorithms in the visible spectrum for automotive applications is not straightforward for several reasons: the problem of face segmentation (distinguishing faces from a cluttered background) should not be avoided, as has been done in many papers by imaging faces against a uniform background; the common use of skin color information to segment the face region is based on computationally expensive initializations and is not so general as to be applicable in different light conditions and with different skin colors; finally, the more precise the location of the eye regions in an initial step, the more reliable the results of the subsequent eye detection algorithms.

The second problem that a visual system for automotive applications has to solve is to model the eye occurrence on image sequences to detect driver status. Substantial amount of data have been collected to study driver behavior from a suite of vehicle sensors and unobtrusively placed video cameras [32]. Human error is known to be a causal factor in many accidents. Inattention and fatigue play an important role in human errors since they affect cognitive aspects and impair perception and the ability to make a decision to react, and also degrade the actual performance of actions. Detecting the driver's state and the driver's fatigue level in particular is actually a difficult task. Research on this subject is still incomplete. At this time it is not possible to provide an exact quantitative assessment of fatigue.

There are also limits to the effectiveness of attention monitoring systems. Monitoring the gaze position or head orientation is one plausible method, but looking at something does not necessarily mean being aware of it [33]. These kinds of analysis require cognitive and physiology studies that are beyond the scope of this paper. However, recent study results on driver accidents demonstrate that in 93% of all rear-end crashes, the driver looked away from the road ahead at least once within 5 s of the crash. What this suggest is that many crashes occur because drivers do not anticipate the events or are unable to respond to these unanticipated events in a timely manner [34]. For this reason, redirecting the driver's eyes to the road ahead may be one way to decrease crash rates. In Ref. [27] a simple characterization of the driver behavior has been introduced by using three different finite state machines monitoring the eye closure rates and long-lasting head rotations to distinguish good visual attention, decreasing visual attention, and low visual attention. The finite state machines have to be initially defined and should be valid for any driver. The Perclos measure (the percentage of time a driver's eyes are closed) has been recognized as a valid measure of loss of alertness among drivers. Many commercial and experimental sensors [4,5,14] currently use the Perclos measure to evaluate driver fatigue and many research efforts are centered on the effects of warning systems on driver performance [35]. The Perclos measure is evaluated during a long observation period in order to assess if a critical level of fatigue has been reached: the higher the Perclos measure, the higher the driver fatigue level. The warning triggers are associated with Perclos calculated over at least 1 min.

1.2. Our contributions

The main objective of our work is to introduce a new visual approach to detect the eye occurrence in image sequences in the visible spectrum, and to learn the model of normal behavior of each driver during an initial training phase. We propose an eye detection algorithm, applicable in the real context of people driving a car, that does not assume predefined acquisition conditions on the background and skips the first segmentation step that extracts the face region, as is commonly described in the literature. The proposed approach works on the whole image, looking for regions that have the same geometrical configuration at the edges as those expected of the iris. Different iris radii are considered to take account of people having different eye dimensions; also, light variation is addressed to take account of the distance between the camera and the person. The search for similar regions is used to discard false positives that can occur in the image. Experimental results demonstrate that when the eyes are open they are correctly detected. However, when the eyes are closed or occluded, the algorithm provides false positives. To overcome this problem, we introduce a further step for the validation of the results: a neural classifier is trained to recognize the "eyes" or "not-eyes" classes using a large number of examples taken from images of different people. A large number of tests have been carried out on different people, with different eye colors and dimensions, some of them wearing glasses and with no constraints on hair style and background. Experiments have been carried out both in the laboratory with varying light conditions such as with natural or artificial illumination coming from only one side of the person (a half of the face lit by the sun), and in real situations of people driving cars. It should be noted that the system does not work with people wearing sun glasses (dark lenses) that occlude the eyes. But this is also the great limitation of any eye detection system (visual and IR based approaches). The only restriction we impose is that the distance between observed people and the camera does not greatly change, since our algorithm requires the knowledge of an approximate range of radius for the iris. In our opinion, this constraint is not at all restrictive to build a visual system for driver inattention detection: in this context the distance between a camera placed on the dashboard and the driver cannot greatly change. Besides, during the acquisition of video sequences, people are allowed to look around, move the head and assume any expression. If the eyes are visible in the image the algorithm is able to detect their presence independently of the gaze direction. Experimental tests demonstrate that, with a camera placed in the central part of the car dashboard, the eyes are clearly visible if the drivers are looking ahead or at the central rear view mirror, but if the drivers are looking in the side view mirror they are forced to rotate the head and the eyes are no longer visible. For this reason we suggest the use of two cameras placed on the dashboard as shown in Fig. 1. In this paper we report the results obtained on image sequences acquired with the central camera. But the integration with the results obtained with the second camera can be easily achieved.

Fig. 1. The two cameras placed on the car dashboard.

The proposed algorithm for eye detection can be used for monitoring the driver alertness by evaluating the Perclos measure during a long period of observation. In this work we propose a different adaptive approach that can be used both to detect driver inattention and to model the normal behavior of each driver and recognize anomalous situations. Two concurrent procedures evaluate the eye occurrences in different observation windows. The first one continuously evaluates the time the driver is looking away from the road. If the number of frames with not-eyes is greater than a fixed threshold an alarm signal is set off. The detection of more complex inattentive behaviors such as with people who shift their attention from the primary task of driving or people impaired by the increase of fatigue levels, requires longer observations of the eye occurrence and above all requires the analysis of behavioral parameters that can be different for different drivers. For

this reason we propose a concurrent procedure that examines the eye occurrence in an initial period of observation, during which a good attention level is assumed, learns some statistical parameters and builds a probabilistic model that characterizes the normal driver behavior. In our opinion driver behaviors can be modelled not only by considering the percentage of time a driver's eyes are closed but evaluating the eye closure duration together with the eye closure frequency. These two parameters contain more information than the percentage parameter: under the same percentage value different behaviors can be recognized with different combinations of eye closure durations and eye frequencies. A multivariate Gaussian mixture model was used in this work to characterize the two considered parameters. After the training phase, the same parameters are extracted from other observed sequences and the probability of having a normal behavior is evaluated: low values are indicative of anomalous situations and then an alarm signal can be set off.

The main contribution of this work is the introduction of a reliable eye detection approach that does not impose any constraints on the background and does not need any preprocessing to segment eye regions. It searches for the eye regions in the whole image, can work in sunlight and it performs well also in cases of people wearing glasses. In addition, the use of a learning phase to extract the parameters characterizing the normal behavior of each driver avoids the difficult problem of building a general model that has to be valid for any driver.

The rest of the paper is organized as follows: Section 2 describes the eye detection algorithm. The behavior analysis approach is detailed in Section 3. The results of different experiments for testing both the eye detection algorithm and the behavior analysis are reported in Section 4. Finally, in Section 5 conclusions are presented.

2. Eye detection

The eye detection algorithm consists of three steps. Initially, an iris detection algorithm, which uses a modified Hough transform, is applied to the whole image to detect the candidate region that might contain one eye. Then a symmetrical region is searched for in limited areas of the image. Finally an eye validation algorithm, based on a neural classifier is used to confirm the eye presence in the image. If the eyes are correctly found their position is used in the successive image to limit the search of the iris detection algorithm. In the following subsections we describe the details of each step.

2.1. Iris detection

The idea behind our work is quite simple: the eyes can be easily located in the image since the iris is always darker than the sclera no matter what color it is. In this way the edge of the iris is relatively easy to detect as a set of points that are placed on a circle. It is not possible to know the exact diameter of the iris since people have different iris dimensions and also the system has to manage variable distances between drivers and the camera. For this reason a range $[R_{min}, R_{max}]$ is set to tackle different iris radii. In this work we have used a circle detection operator based on the directional circle Hough transform that has been already used in a different application in Ref. [36]. In particular the algorithm is based on convolutions applied to the edge image. The operator that follows produces a maximal value when a circle is detected with a radius in the range $[R_{min}, R_{max}]$:

$$u(x, y) = \frac{\int\int_{D(x,y)} \vec{g}(\alpha, \beta) \cdot \vec{O}(\alpha - x, \beta - y)\, \mathrm{d}\alpha\, \mathrm{d}\beta}{2\pi(R_{max} - R_{min})}, \tag{1}$$

where the domain $D(x, y)$ is defined as

$$D(x, y) = \{(\alpha, \beta) \in \mathbb{R} \times \mathbb{R} \mid R_{\min}^2 \leqslant (\alpha - x)^2 + (\beta - y)^2 \leqslant R_{\max}^2\}. \tag{2}$$

$\vec{g}$ is the gradient vector

$$\vec{g}(x, y) = [\frac{\vartheta}{\vartheta x} I(x, y), \frac{\vartheta}{\vartheta y} I(x, y)]^{\mathrm{T}} \tag{3}$$

and $\vec{O}$ is the kernel vector

$$\vec{O}(x, y) = \left[\frac{\cos(\arctan(y/x))}{\sqrt{x^2 + y^2}}, \frac{\sin(\arctan(y/x))}{\sqrt{x^2 + y^2}}\right]^{\mathrm{T}}. \tag{4}$$

The kernel vector contains a normalization factor (the division by the distance of each point from the center of the kernel) that is fundamental in order to have the same values in the accumulation space when circles with different radii in the admissible range are found. Besides, the normalization ensures that the peak in the convolution result is obtained for the most complete circle and not for the greatest in the annulus.

The circle detection operator is applied to the whole image without any constraint on plain background or limitations on eye regions. The resulting maximum values represent the regions of the whole image that are the likeliest candidates to contain an eye. In Fig. 2 the resulting image of the convolution step together with the original image are shown. The peaks are obtained in the areas around the eyes but also in other zones of the face.

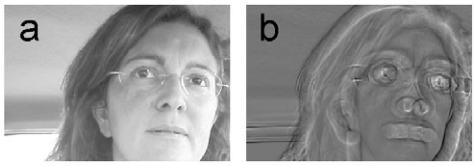

Fig. 2. The original gray level image (a) and the resulting convolution image (b).

2.2. Search for symmetric regions and correlation

Once the first maximum is obtained the candidate region that might contain one of the eyes is selected. Starting from this area the search for the second eye is applied only in the two opposite regions whose distances and orientations are compatible with the range of possible eye positions. Let (x_1, y_1) be the coordinate of the central point of the first maximum

Table 1
The procedure for searching the second eye

1. Search the maximum value M1 of the output convolution in the whole image
2. Search the second maximum M2 in the regions D that are candidate to contain the second eye
3. Compare the two region M1 and M2 using the measure MAE
3. IF $(MAE(M1, M2) < Threshold)$ THEN Output = M1,M2 ELSE M1 = 0 GO TO step 1

obtained. The second point (x_2, y_2), where the second eye could be found, belongs to the regions defined as follows:

$$D(x, y) = \{(x_2, y_2) \mid x_2 = x_1 \pm d, y_2 = \tan \alpha(x_2 - x_1) + y_1\}, \quad (5)$$

where $d \in [d_1, d_2]$ and $\alpha \in [-\pi/6, \pi/6]$.

In this way false positives that occur in the hair regions or on other face parts (such as nose, mouth, and so on) are easily discarded since their distances are not compatible with the distance between the eyes. Different similarity measures have been explored. The results obtained are quite similar. For this reason, since our main goal is to have frame rate performance, we decided to choose the one that required the least computational load: the mean absolute error (MAE). The MAE is evaluated on mirrored domains:

$$MAE = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}|a_{ij} - b_{i(M-j+1)}|}{N \cdot M}, \quad (6)$$

where a and b represent the points in the two regions whose dimensions are $N \times M$. The mirroring of the second region is necessary to evaluate the symmetry of the two candidate eyes. If this similarity measure is below a fixed threshold the two regions are considered the best match for eye candidates, otherwise a further search is activated. The whole procedure used for searching the eyes in the image is described in Table 1.

The results of the similarity measure are strictly dependent on the region dimensions: the smaller the regions, the higher the false positives found. Intuitively, two similar circular regions can be found in many areas of face images: among curly hair, at the sides of the mouth, on the forehead, and so on. However, if the two regions considered for the similarity measure are larger than the circular shape searched for in the first step, the probability of finding false matches decreases. The experiments show that the similarity constraint is very selective: a large number of wrong candidates are discarded since their similarities overcome the fixed threshold. The threshold has been selected experimentally by evaluating the similarity of a priori known eye regions. In addition, experimental results have demonstrated that when the light conditions are uniform, the similarity measure rarely fails. But one of the most common situations is that sunlight illuminates mainly one part of the face. In these cases the eyes have different appearance and the similarity measure will produce wrong answers. For this reason a normalization procedure is introduced on both the regions, subtracting the mean values from the pixel intensities.

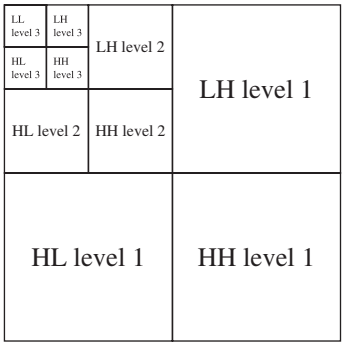

Fig. 3. The decomposition of the image with a 3-level wavelet transform.

After the normalization, the similarity measure can be applied since we are sure that the results are independent of the changing luminosity in the image.

2.3. Eyes validation by using a neural network

The two candidate regions that might contain the eyes have to be validated to detect if the eyes are closed or false detection has been found. The sub-images containing the result of the detection process are fed to a neural network trained to classify images as eye or not-eye. The main requirement of our application is the possibility to work in real time. This requirement can be satisfied if the algorithm of features extraction is able to store all the necessary information in a small set of coefficients and the algorithm for classification is skilful enough to take advantage of that in order to supply right answers in reduced times. In this work we have used a discrete wavelet transform that supplies a hierarchical representation of the image, implemented with the iterative application of two filters: a low-pass filter (approximation filter) and its complementary filter in frequency (detail filter). At each step the wavelet transform breaks the image into four sub-sampled or decimated images.

In Fig. 3 the result of a 3-level wavelet transform is shown. The capital letters in each sub-image represent the kind of filters that were applied on the image of the previous level: H stands for a high-pass filter, L stands for a low-pass filter. The first letter is the filter that is applied in the horizontal direction, while the second letter is the filter that is applied in the vertical direction. The band LL is a coarser approximation of the original image. The band LH and HL record the changes of the image along horizontal and vertical directions. The band HH shows the high-frequency components of the image. Decompositions can be iterated on the LL sub-bands. After applying a 3-level wavelet transform, an image is decomposed into sub-bands of different frequency components. Because of their simplicity, Haar filters were used for the implementation of the wavelet transform. The coefficients of the wavelet transform hold information about the texture and the shape of the object in the image. In this way it is possible to distinguish the eyes from other elements that could have one of the two aspects in common. The final result of a 3-level wavelet transform for the eye image is shown in Fig. 4. We have selected for the recognition process the coefficients of the third level of decomposition of the wavelet transform because

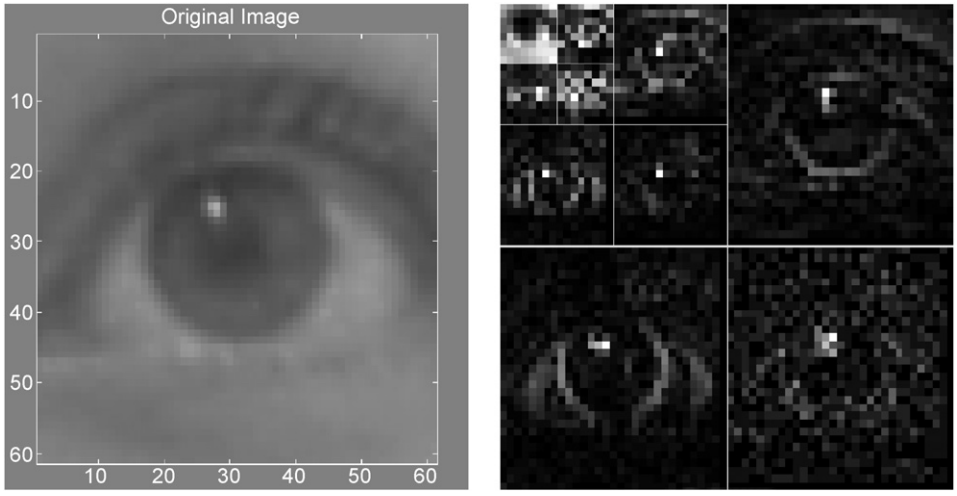

Fig. 4. The decomposition of the eye image. On the left the original image, on the right the result of the 3-level wavelet transform.

they mainly contain the information of shape and texture. The total number of features extracted is equal approximately to $\frac{1}{16}$ of the total number of pixels of the initial image.

The classifier consists of a neural network, which is simple to use and widely employed in pattern recognition problems, trained with a back propagation algorithm. The input layer has the number of nodes equal to the number of coefficients of the WT of the third level, the hidden layer has 80 nodes and the output layer has only one node. The neural network parameters were found experimentally, choosing the combination that produced the best results in terms of eye detection rate. In the learning phase the weight updating has been done using the gradient descent rule with a momentum term $\alpha = 0.1$ and learning rate $\eta = 0.5$. All the nodes of the network used the sigmoid activation function since it evaluates the output as a non-linear function of its inputs, it is differentiable, and it has been proven to be a good approximation of Boolean functions. A large number of examples of eyes and not-eyes were selected from images of different people and were fed to the neural network to train the classifier. The training phase of the neural network was performed just once, at the beginning. The same network was used in several experimental sessions to recognize the presence of the eyes with different people (not those used in the training phase), and light conditions. At this point if at least one of the two candidate regions is recognized as eye instance, the algorithm skips to the successive image of the sequence and uses the estimated eye positions to limit the search in the neighboring areas. Otherwise the search is applied to a larger window. The window dimension increases at each step by a constant quantity until the eyes are located or the window dimension equals the image dimension. As soon as the algorithm detects the presence of the eyes the search is restricted again to the initial small window.

3. Behavior analysis

Detecting the driver attention level with a visual system is a difficult task. An automatic system should consider many factors. The occurrence of the eyes in the image sequences needs to be evaluated to monitor the eye closure rates. Many studies into human sleep have been carried out to understand the physiological changes, such as accelerated respiration, decreased brain activity, eye movement, and muscle relaxation. The gaze direction needs to be evaluated in order to understand if people are looking straight ahead, into the rear-vision mirrors, or somewhere else. Also the head rotation needs to be considered to determine if the driver is paying adequate attention to the road. The aim of this paper is not to provide a model of the possible driver behaviors in order to recognize anomalous situations, but only to study if the simple analysis of the eye occurrence can be used to monitor driver inattention and send alarm signals when an abnormal behavior is detected. The problem of raising alarms in this application context is crucial: too frequent an alarm would be considered annoying. An automatic system used to highlight a lack of attention in the driver behavior has to be quick in the detection and robust, so that wrong estimates of alarm situations can be avoided as much as possible. However, the main point that must be considered is the unpredictability of the situation in which the driver is losing attention: there can be no examples of such a case to teach an automatic system in the training phase to recognize abnormal behavior.

What is available is the general statistics of the "normal driver behavior". The normal situation can be "learned" by an automatic system: using the specific driver peculiarities, its "normal case" parameters estimation can be carried out, and the statistics of parameters considered relevant to the driver attention assessment can be modelled. In this work the eye closure duration (referred afterward as ECD) and the frequency of eye closure (FEC) were selected as the most pertinent and measurable parameters. They represent a twofold refinement on the currently used Perclos measure as each is a factor in determining the Perclos. Our main objective was to show that a bespoke model of normal behavior for each person can be created with the two parameters, against which non-normal behavior can be identified using the same two parameters. As a secondary

objective we aimed to provide evidence that these two described parameters can work accurately identifying anomalous behaviors that were not detected by Perclos. Certainly, in real applications the proposed approach has to be integrated into more complex analysis systems, which consider measures coming from other sensors such as the steering wheel, the accelerator, the lane position, and so on [1].

In order to model the "normal behavior" statistics for each person we have used a mixture Gaussian model. Gaussian mixtures are a combination of a finite number of Gaussian distributions, largely used to model complex multidimensional distributions. Even if only two parameters make up our statistics, the forms of the underlying density functions are not known. Furthermore, high-dimensional densities cannot be simply represented as the product of 1D density functions. Mixture Gaussian models offer a potential tool to learn different complex distributions [37]: they can be used with arbitrary distributions and without the assumption that the forms of the underlying densities are known. A mixture of Gaussian can be written as a weighted sum of Gaussian densities. The 2D Gaussian probability density function (pdf) with mean vector μ and covariance Σ can be represented as follows:

$$g_{(\mu,\Sigma)}(x) = \frac{1}{\sqrt{2\pi}\cdot\sqrt{\det(\Sigma)}}\cdot \exp^{-1/2(x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)}. \quad (7)$$

A weighted mixture of K Gaussian can be written as

$$gm(x) = \sum_{k=1}^{K} w_k \cdot g_{(\mu_k,\Sigma_k)}(x), \quad (8)$$

where the weights are all positive and sum to one:

$$w_k \geqslant 0, \quad \sum_{k=1}^{K} w_k = 1 \ \text{ for } k \in \{1,\ldots,K\}. \quad (9)$$

By varying the number of Gaussian K, the weights w_k, and the parameters μ_k and Σ_k of each Gaussian density function, Gaussian mixtures can be used to describe any complex pdf. To find the parameters that optimally fit a certain pdf for a set of data, an iterative algorithm, the expectation maximization (EM), is used [38]. Starting with initial values of all parameters, they are re-estimated iteratively. In this work the algorithm of EM presented in Ref. [39] is used. It requires as input parameters the expected number K of Gaussian mixtures that should be used for the pdf and the data set that should be modelled by the pdf. After the construction of the normal behavior model it is possible to evaluate for different test sequences the probability of belonging to the estimated model: the lower the obtained values, the higher the probability of representing anomalous behaviors. In these cases an alarm signal is provided.

4. Experimental results

The experimental phase consists of two parts. First of all, the eye detection algorithm was tested, using both image sequences of different people taken in our laboratory in different light conditions and image sequences of people while driving a car. Then the algorithm for behavior analysis was tested. The eye closure parameters of two people were extracted from a long image sequence during which a normal behavior was assumed. Then the statistical model was generated and used to recognize anomalous behaviors differing from the learned model.

4.1. Eye detection results

In order to evaluate the eye detection algorithm performances, three sets of experiments were carried out. The first two were performed in our laboratory by using a Fire-wire camera that took 30 frame/s with a resolution of 720 × 575 pixels. In the first we considered diffuse lighting conditions, while in the second one only half the face was lit (as is often the case in real situations in sunlight). The third set of experiments was carried out on image sequences acquired in a car by using a Webcamera with a low resolution of 320 × 240 pixels and frame rate 15 frame/s. Two groups of people were used to extract images with different eye color, dimension and shape. We assumed that the distance between the camera and the person cannot change greatly (they remained seated in front of a camera), but we allowed for slight modification to the chair position as is normally the case. No constraints were imposed on the background.

Since the image resolution was different for the two experimental setups used, two different training phases of the neural network were necessary. In the images acquired in our laboratory the iris dimensions could vary within a range of [20,30] pixels. Then the dimension of each example image was fixed to 61×61 pixels, about twice the size of the iris dimension. In this case the training set for the neural classifier contains 452 negative examples (not-eye images) and 291 positive examples (eye images), obtained from images from the first three people. In the first row of Fig. 5 some images of eye examples are shown, while in the second row some not-eye examples extracted from other parts of the face and used in the training phase are shown. In the images acquired in the car the iris dimensions could vary within a range of [10,15] pixels (the dimension of each example image was fixed to 30 × 30 pixels). The training set for the neural classifier contains 30 negative examples and 85 positive examples obtained also in this case from images from the first three people.

The images of the people used in the three sets of experiments with the eye detection results superimposed on each frame are shown in Figs. 6–8. In Tables 2–4 the results of the eye detection algorithm are reported. The test images were divided into three sets: eyes completely open, eyes partially visible (in this set we have considered all the images in which people were looking off center, or the eyelids covered part of the iris), not-eyes (the eyes could be closed, or people had the head turned). The number of images used for each person is specified in the first column; in each row of the tables, the detection rates with the number of images for each set of eyes open, eyes partially open, and not-eyes are reported. In the columns the detections rates for the three sets of images are represented in terms of true positive (TP), i.e. the number of images in which the eyes are open

Fig. 5. In the first row some positive examples and in the second row some negative examples provided to the neural network in the training phase.

Fig. 6. Some results obtained on images taken in our laboratory in the first set of experiments with uniform lighting condition.

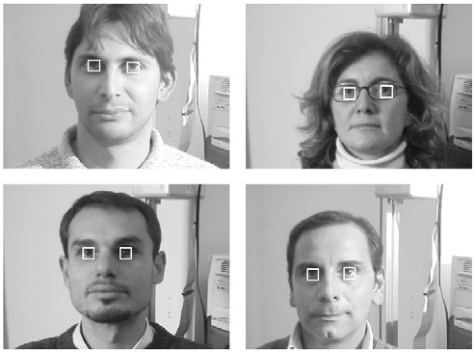

Fig. 7. Some results obtained on images taken in our laboratory in the second set of experiments with partially illuminated faces.

Fig. 8. Some results obtained on images acquired in the car by a low-resolution camera in the third set of experiments.

and the algorithm correctly recognizes the eyes, false positive (FP) the number of images in which the algorithm erroneously recognizes the eyes in wrong areas, and false negative (FN), i.e. the cases in which the algorithm fails to recognize the eyes. In the case of not-eye images the true negative (TN) detection rates that represents the percentage of images in which the algorithm correctly recognizes the absence of eyes are reported. In the last row of the tables the total detection rates are evaluated. The (*) notation indicates people included in the training of the neural network. It should be noted that in the second set of experiments in our laboratory with not-uniform lighting conditions the same neural classifier, trained with the images of the first experiment, was used; in particular some images of the person #3 with uniform lighting condition were included in the first training set.

On the set of images with open eyes the proposed algorithm reaches high detection rates, with a percentage of TP over 95% (first columns of Tables 2–4). Also the numbers of TNs are very high. It should be noted that performance is very good also when the algorithm is tested on different people from those included in the training set (see the last three rows of Tables 2–4 corresponding to the persons without the (*)). The third set of experiments demonstrates that even if the image quality and resolution is decreased, the proposed algorithm reaches high detection rates both in terms of percentage of TPs with open eyes, and in terms of percentage of TNs when the eyes are closed.

When the eyes are partially open the detection rates are about 70% (the TP of the set eyes partially open in the three tables). These situations happen when people are closing the eyes, or something is occluding the face. Lower values are expected as both the iris detection and the neural classifier can fail. The circle detection algorithm cannot identify the iris if their contours are not complete. The neural network, trained only with open eye examples, can fail to generalize when the images contain parts of the eyes. Some images of correct detection when the eyes are partially visible are shown in Fig. 9. It should be observed that when people have their heads turned several situations may occur. If one of the eyes is occluded (the head has a large rotation with respect to the camera) the similarity constraint will not allow the detection of the eyes. If some part of the second eye is visible and the gaze direction is toward the camera (people are still looking toward the camera even if the head is rotated), the iris contours are almost complete and

Table 2
Results obtained after the eye detection algorithm

	Open eyes			Eyes partially open			Not-eyes	
	TP	FP	FN	TP	FP	FN	TN	FN
Person 1* (%)	100	0	0	50	0	50	95	5
(95)	($\frac{73}{73}$)	($\frac{0}{73}$)	($\frac{0}{73}$)	($\frac{1}{2}$)	($\frac{0}{2}$)	($\frac{1}{2}$)	($\frac{19}{20}$)	($\frac{1}{20}$)
Person 2* (%)	100	0	0	75	0	25	100	0
(235)	($\frac{210}{210}$)	($\frac{0}{210}$)	($\frac{0}{210}$)	($\frac{3}{4}$)	($\frac{0}{4}$)	($\frac{1}{4}$)	($\frac{21}{21}$)	($\frac{0}{21}$)
Person 3* (%)	97.1	0.2	2.7	60	10	30	90	10
(454)	($\frac{402}{414}$)	($\frac{1}{414}$)	($\frac{11}{414}$)	($\frac{6}{10}$)	($\frac{1}{10}$)	($\frac{3}{10}$)	($\frac{27}{30}$)	($\frac{3}{30}$)
Person 4 (%)	97.7	0.6	1.7	62	5	33	96.3	3.7
(368)	($\frac{295}{302}$)	($\frac{2}{302}$)	($\frac{5}{302}$)	($\frac{24}{39}$)	($\frac{2}{39}$)	($\frac{13}{39}$)	($\frac{26}{27}$)	($\frac{1}{27}$)
Person 5 (%)	100	0	0	100	0	0	100	0
(266)	($\frac{235}{235}$)	($\frac{0}{235}$)	($\frac{0}{235}$)	($\frac{14}{14}$)	($\frac{0}{14}$)	($\frac{0}{14}$)	($\frac{17}{17}$)	($\frac{0}{17}$)
Person 6 (%)	98	1	1	91	0	9	89	11
(153)	($\frac{94}{96}$)	($\frac{1}{96}$)	($\frac{1}{96}$)	($\frac{20}{22}$)	($\frac{0}{22}$)	($\frac{2}{22}$)	($\frac{31}{35}$)	($\frac{4}{35}$)
Detection rate (%)	98.4	0.3	1.3	75	3	22	94.4	5.6

The (*) notation signs people included in the training set of the neural network.

Table 3
Results obtained after the eye detection algorithm on images with not-uniform light

	Open eyes			Eyes partially open			Not-eyes	
	TP	FP	FN	TP	FP	FN	TN	FN
Person 3* (%)	97.5	0	2.5	91	3.7	5.3	97	3
(465)	($\frac{341}{350}$)	($\frac{0}{350}$)	($\frac{9}{350}$)	($\frac{48}{53}$)	($\frac{2}{53}$)	($\frac{3}{53}$)	($\frac{69}{71}$)	($\frac{2}{71}$)
Person 4 (%)	98.3	0.1	1.6	76	0	24	97	3
(845)	($\frac{702}{715}$)	($\frac{1}{715}$)	($\frac{12}{715}$)	($\frac{75}{98}$)	($\frac{0}{98}$)	($\frac{23}{98}$)	($\frac{31}{32}$)	($\frac{1}{32}$)
Person6 (%)	81.5	9.5	9	31	3	66	89	11
(210)	($\frac{128}{157}$)	($\frac{15}{157}$)	($\frac{14}{157}$)	($\frac{11}{35}$)	($\frac{1}{35}$)	($\frac{23}{35}$)	($\frac{16}{18}$)	($\frac{2}{18}$)
Person 7 (%)	92	1	7	53	6	43	95	5
(135)	($\frac{90}{98}$)	($\frac{1}{98}$)	($\frac{7}{98}$)	($\frac{8}{16}$)	($\frac{1}{16}$)	($\frac{7}{16}$)	($\frac{20}{21}$)	($\frac{1}{21}$)
Detection rate (%)	95.5	1.3	3.2	70.3	2	27.7	95.7	4.3

The (*) notation indicates people included in the training set of the neural network.

Table 4
Results obtained after the eye detection algorithm on images acquired in the car

	Eyes open			Eyes partially open			Not-eyes	
	TP	FP	FN	TP	FP	FN	TN	FN
Person 1* (%)	97	0.5	2.5	83.6	3.9	12.5	94.3	5.7
(1024)	($\frac{785}{809}$)	($\frac{3}{809}$)	($\frac{21}{809}$)	($\frac{107}{128}$)	($\frac{5}{128}$)	($\frac{16}{128}$)	($\frac{82}{87}$)	($\frac{5}{87}$)
Person 2* (%)	97.8	0	2.2	62	0	38	100	0
(1082)	($\frac{794}{812}$)	($\frac{0}{812}$)	($\frac{18}{812}$)	($\frac{121}{195}$)	($\frac{0}{195}$)	($\frac{74}{195}$)	($\frac{75}{75}$)	($\frac{0}{75}$)
Person 3* (%)	98	0.5	1.5	80.8	2.8	16.4	96	4
(1276)	($\frac{848}{865}$)	($\frac{4}{865}$)	($\frac{13}{865}$)	($\frac{172}{213}$)	($\frac{6}{213}$)	($\frac{35}{213}$)	($\frac{190}{198}$)	($\frac{8}{198}$)
Person 4 (%)	85.5	0	14.5	53.7	1.0	45.3	98.5	1.5
(913)	($\frac{641}{750}$)	($\frac{1}{750}$)	($\frac{108}{750}$)	($\frac{51}{95}$)	($\frac{1}{95}$)	($\frac{43}{95}$)	($\frac{67}{68}$)	($\frac{1}{68}$)
Person 5 (%)	98.7	0	1.3	70.7	3.4	25.9	96.2	3.8
(937)	($\frac{597}{605}$)	($\frac{0}{605}$)	($\frac{8}{605}$)	($\frac{123}{174}$)	($\frac{6}{174}$)	($\frac{45}{174}$)	($\frac{152}{158}$)	($\frac{6}{158}$)
Person 6 (%)	92.3	0	7.7	70.1	7.7	13.2	95.8	4.2
(847)	($\frac{610}{661}$)	($\frac{0}{661}$)	($\frac{51}{661}$)	($\frac{72}{91}$)	($\frac{7}{91}$)	($\frac{12}{91}$)	($\frac{91}{95}$)	($\frac{4}{95}$)
Detection rate (%)	95.0	0.2	4.8	72.1	2.8	25.1	96.5	3.5

The (*) notation indicates people included in the training set of the neural network.

the system is able to detect at least one eye (the two similar circular regions are identified, and the neural network recognizes the most complete eye region). In the case of a different gaze direction (people are not looking toward the camera) and the irises are not clearly visible, then the system correctly detects not-eyes. In Fig. 10 the image of the these different cases are shown.

4.2. Behavior analysis results

The behavior analysis was carried out according to the scheme shown in Fig. 11. Two concurrent procedures were used to evaluate the eye occurrences in the sequences of eye detections (sequences of 0/1) both to detect simple inattention and to recognize different behaviors from the normal one learned for each driver. The inattention detection procedure required only short sequences of eye occurrence: if the number of frames with not-eyes is greater than a fixed threshold this means that the driver is not paying adequate attention to the road and then an alarm signal is set off. The threshold can be fixed according to the velocity of the car, since the effects of driver inattention can be more catastrophic when the speed increases.

The behavior analysis considered a larger sliding window of eye occurrence observations. The experiments were executed on real image sequences of two different drivers. For each of them, 20 min of observation were recorded: the first 10 min were used to learn the normal behaviors, while in the remaining 10 min the drivers alternated normal behavior with cell phone activities, passenger related tasks, and having a conversation. In Fig. 12(a) and (b) the Perclos measures corresponding to the normal behavior in training and the normal and anomalous test behaviors are plotted. In the graphs different line representations have been used to separate these situations.

Fig. 9. Some correct detection when the eyes are partially visible.

In Fig. 12(a) it is evident that anomalous behaviors cannot be detected by using the Perclos measure alone (except for the peaks over a threshold). In particular, note the similarity between the training phase and the passenger related task phase. Instead, the proposed observation of two different parameters EDC and FEC to learn the model of normal behavior of each driver gives the system greater power to separate normal from anomalous behaviors.

A sliding window containing 30 s of images was considered for extracting the statistical parameters of each driver. The observation window was shifted every 7.5 s, obtaining a total number of 77 windows in which the frequency and the duration parameters of the eye closures are evaluated. The plots of frequency versus duration for the two drivers are shown in Fig. 13. The mixture Gaussian models that best fit the data sets are learned by using the EM algorithm [39].

In Fig. 14 the mixture Gaussian models learned by using the EM algorithm are shown for the two people. In this experiment the number of Gaussians was fixed at three. It should be noted that, since the two driver had different normal behaviors, the two bidimensional models assume different shapes and also spread across different ranges. In order to assess

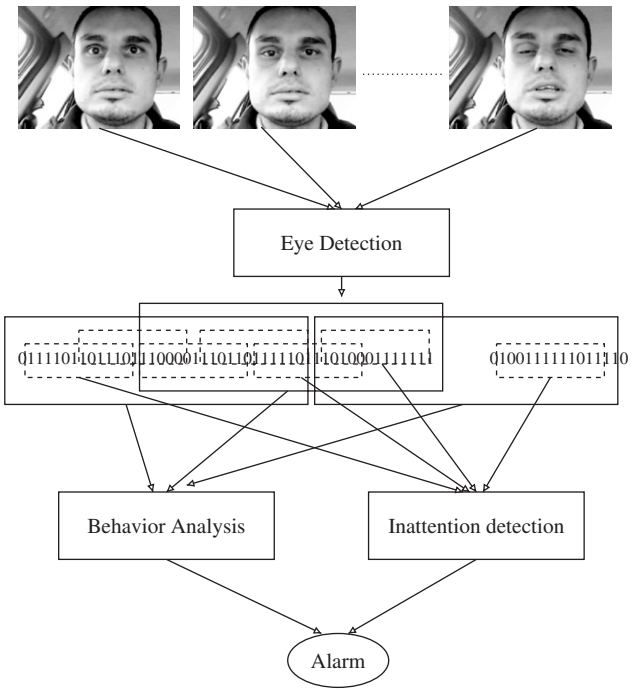

Fig. 11. The behavior analysis scheme.

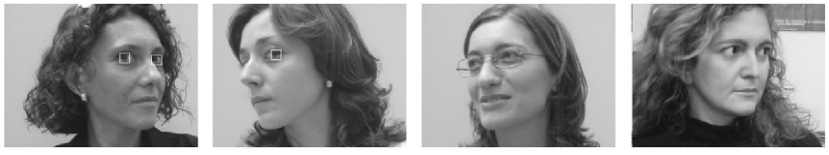

Fig. 10. Some results obtained with the head turned with different gaze directions. In the first two images the eyes have been detected. In the other two images the eyes have not been detected because people are not looking toward the camera.

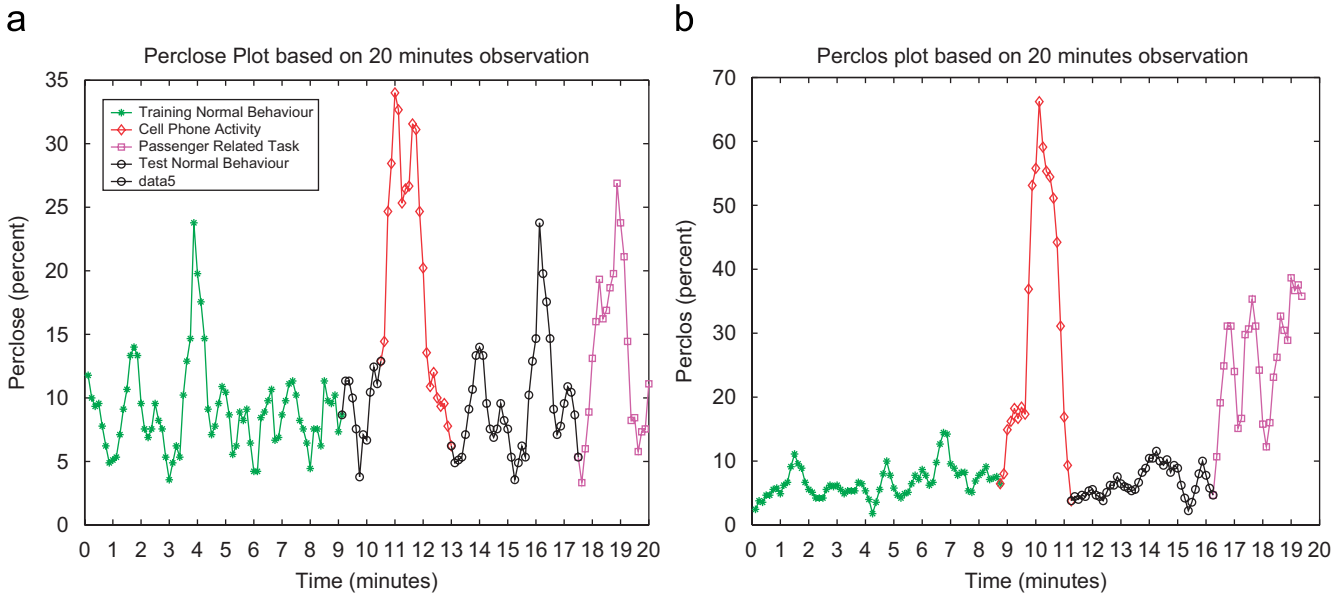

Fig. 12. The Perclos measure plotted over an observation period of 20 min. (a) First person. (b) Second person.

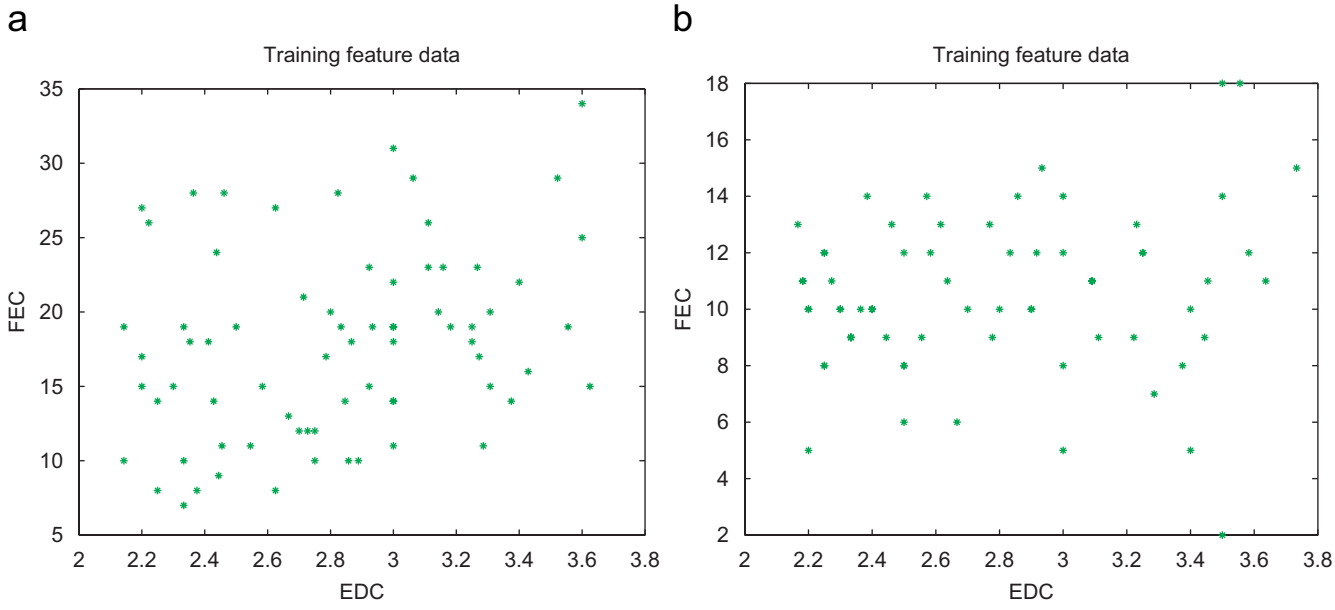

Fig. 13. The plot of the parameters EDC (x-axis) and FEC (y-axis) extracted during the learning phase. (a) First person. (b) Second person.

the effectiveness of the learned model the successive period of observation was processed, during which the two people had normal and anomalous behaviors. The duration and frequency parameters were extracted and their probability values were evaluated with respect to the mixture Gaussian model learned during the normal behavior. In Figs. 15(a) and 16(a) the test points corresponding to normal and anomalous behaviors are plotted with different shapes while in Figs. 15(b) and 16(b) the corresponding probabilities are shown. Also in this case the two people behave differently while they are involved in the same tasks. However, anomalous behaviors produced duration and frequency parameters which gave rise to low probability values when applied to the learned Gaussian models. The threshold that has to be used to cutoff the values corresponding to anomalous behaviors is obtained automatically by evaluating the 10% of the mean probability of the training values plotted in the learned model. In Fig. 17 we have plotted on the Perclos graph the points that are outside this threshold value for the first person. In particular note that the values corresponding to the cell phone activity (minutes 11–12) and the conversation with the passenger (minutes 18–19) are actually recognized as outliers with respect to the learned model, while the normal

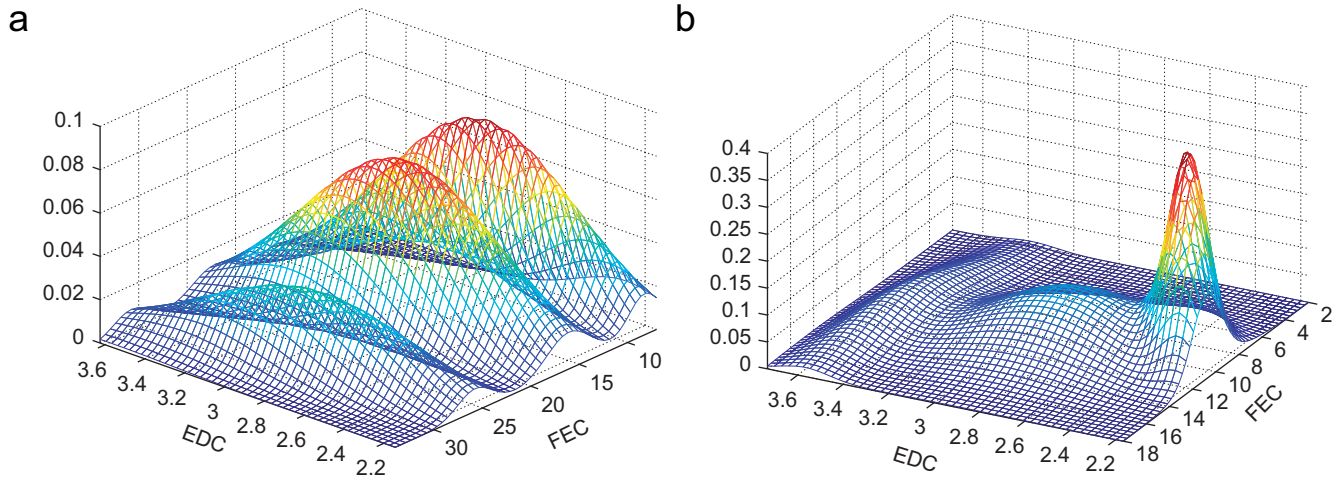

Fig. 14. The mixture Gaussian models learned with EM. (a) First person. (b) Second person.

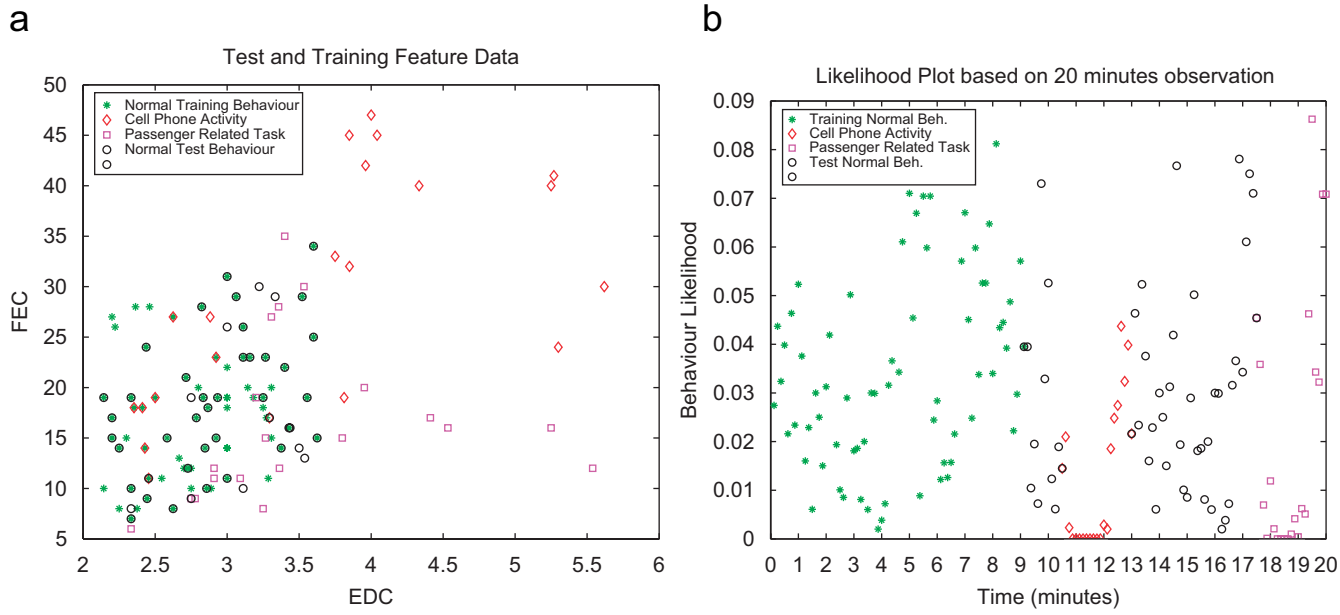

Fig. 15. First person. (a) The test data over-imposed on the training data. (b) The probability values of the test sequences on the learned mixture Gaussian model.

behavior, except for one point, produces probability values above the fixed threshold. The same results could not be obtained by using the Perclos values of Fig. 12(a).

The software was implemented by using Visual C++ on a Pentium IV 3.2 GHz without any processing specialized hardware. Although not all the code optimizations were completed the processing times are encouraging for a real time implementation. We evaluated the processing time when the search was applied to the whole image and when the search was applied to smaller windows during the tracking procedure. It should be noted that when the eyes are correctly found in the image, and the tracking procedure is active, the search is limited to small areas; as soon as the eyes are not found, the window dimension is gradually increased in the successive frames. For this reason the processing times quadratically increased with the number of frames in which the eyes were not found by the algorithm. In Fig. 18 the processing time for increasing window dimension is reported for the real experiments executed with a low-resolution camera in the car. The processing time for the search in the whole image is around 0.5 s.

In order to evaluate if the algorithm is able to process the image sequences acquired during 30 s of observation, we have plotted in Fig. 19 the curve delimiting the domain of the two parameters EDC and FEC that can be processed in real time (i.e. about 30 s). The combinations of (EDC, FEC) that lie under the curve can be processed in real time, while those above the line require more processing load. As an example, we considered the test sequences of Fig. 16(a): only a few test examples,

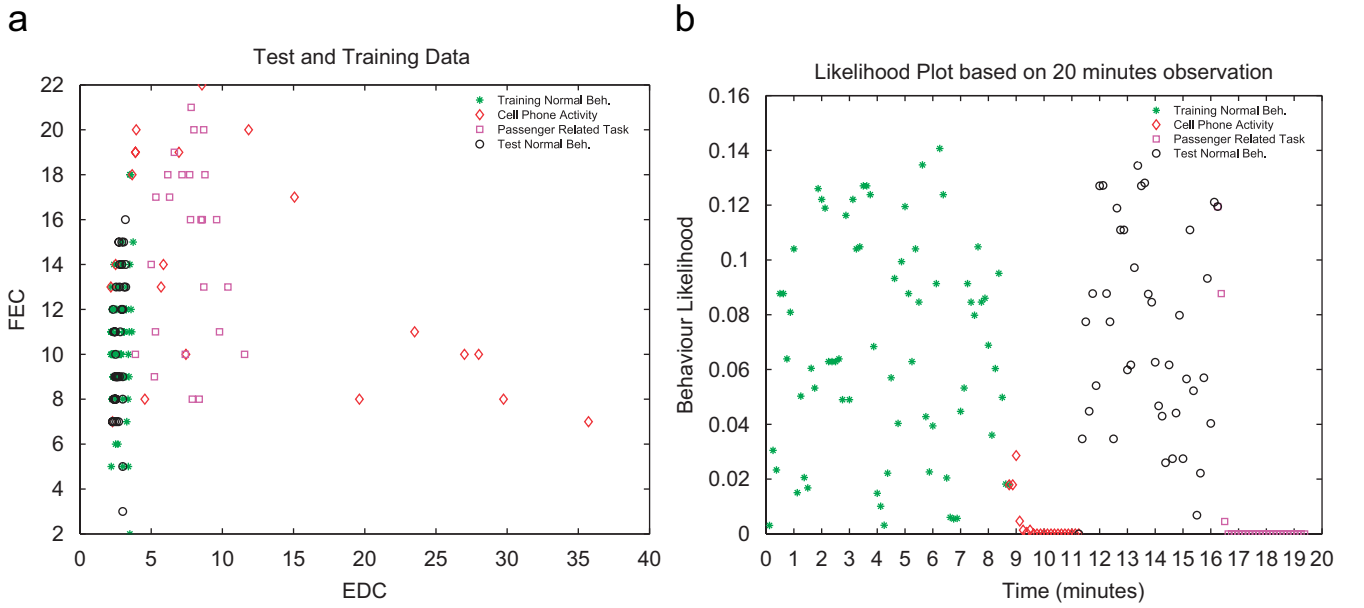

Fig. 16. Second person. (a) The test data over-imposed on the training data. (b) The probability values of the test sequences on the learned mixture Gaussian model.

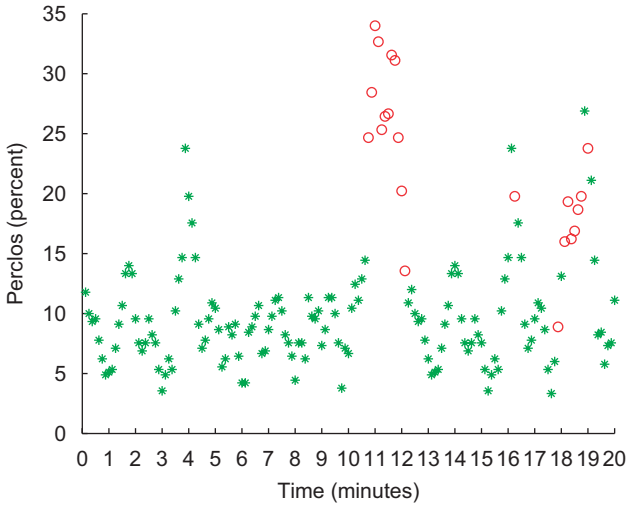

Fig. 17. The points of the Perclos graph that have been classified as anomalous by using a threshold on the learned bidimensional model of FEC and EDC: the green stars correspond to normal behaviors while the red circle correspond to anomalous behaviors.

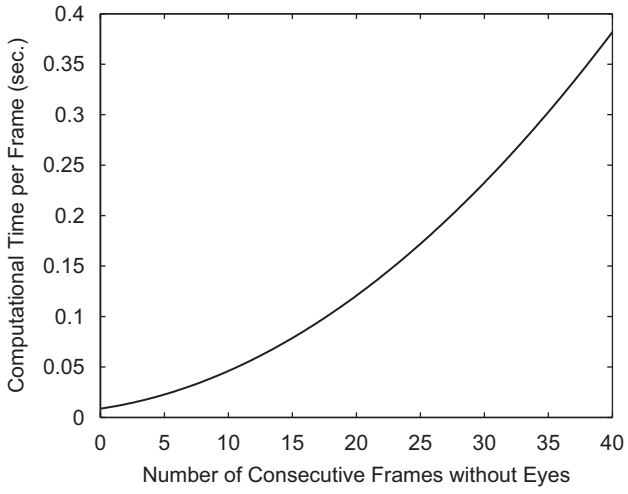

Fig. 18. The quadratic processing time for increasing window dimensions.

which represent highly anomalous situations, lie next to the curve of Fig. 19, all the other test sequences can be processed in less then 30 s. These evaluations were done also for the experiments executed in the laboratory with a high-resolution camera and comparable results were obtained without any code optimization. If some specialized hardware is introduced to speed up the code, the processing time for the whole image will be reduced and the curve of Fig. 19 will move toward the top right corner.

5. Conclusions

In this work we propose a visual framework that, together with other kinds of sensors, can be used for warning against driver inattention. Many research groups are working on the possibility of providing the car with sensors for monitoring the steering wheel, brakes, accelerator, lane keeping and so on. The prevailing method for detecting driver inattention involves the tracking of the driver's head and eyes. Active IR illumination approaches have been largely used for detecting eyes and are robust in controlled light conditions. However, they are inefficient in strong light conditions or when people wear glasses. Instead, in these conditions a visual camera, that observes the scene, has no problem in detecting the eyes. We propose a

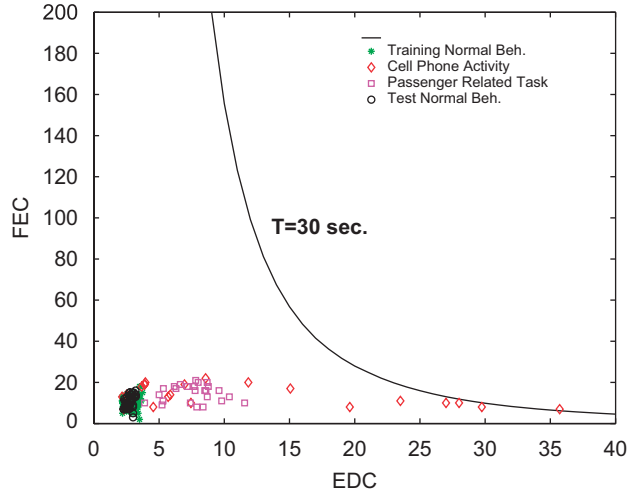

Fig. 19. The curve delimiting the real time domain. Dots represent the EDC and FEC parameters obtained for the test sequences. Most of the points falls under the curve for $T = 30$ s in the real time domain.

visual framework that consists of an eye detection step and a behavior analysis step. It searches for the eye occurrence in the image sequences and then it learns the model of normal behavior for each driver during an initial training phase. The main contribution of this work is in the innovative definition of the eye detection algorithm, applicable in real contexts of people driving a car, that does not assume predefined acquisition conditions on the background and skips the first segmentation step to extract the face region as commonly described in literature. It consists of three steps. First, the candidate region that might contain one eye is detected in the whole image matching the edge directions with an edge template of the iris. Then, the search for the second eye is applied in the two opposite regions whose distances and orientations are compatible with the range of possible eye positions. Finally, the two candidate regions are validated by using a neural classifier. Our system does not impose any constraint on the background and does not require any preprocessing step for face segmentation. High detection rates were obtained on tests carried out on six people in different light conditions.

The second step of the visual framework proposes a novel approach for behavior analysis. The idea of this approach is to let the system learn several models for different drivers. Indeed we assume that different people have different behavioral attitudes while driving which can be characterized by different head movements and eye blinking. Therefore, the system has to learn the concept of normality for each driver. The examination of the eye occurrence in an initial period of observation, during which a good attention level is assumed, is used to learn some statistical parameters and build a probabilistic model characterizing the normal driver behavior. The use of the two parameters FEC and EDC refines the system measures increasing performance over the commonly used Perclos measure.

In addition, the learning system could be left "always awake", to learn new behaviors, thus making the system adaptive. In the adaptive operation mode, the system can acquire information on line and save probabilistic borderline experiences as "new experiences". Such new data can be used to update the driver probabilistic model.

Of course in a real application the proposed approach has to be integrated into more complex analysis systems that consider measures coming from other sensors such as the steering wheel, the accelerator, the lane position, and so on [1].

Acknowledgment

The authors would like to thank Neil Owens for his helpful and critical suggestions during the preparation of this manuscript.

References

[1] K. Torkkola, N. Massey, C. Wood, Driver inattention detection through Intelligent Analysis of Readily Available Sensors, in: Proceedings of IEEE Conference on Intelligent Transportation Systems, Washington, DC, October 2004, pp. 326–331.
[2] G. Yang, Y. Lin, P. Bhattacharya, A driver fatigue recognition model using fusion of multiple features, in: IEEE International Conference on Systems, Man and Cybernetics, Hawai USA, vol. 2, October 2005, pp. 1777–1784.
[3] S. Park, M. Trivedi, Driver activity analysis for intelligent vehicles: issues and development framework, in: Proceedings of IEEE Intelligent Vehicles Symposium, Las Vegas, USA, June 2005, pp. 795–800.
[4] ⟨http://www.smarteye.se⟩.
[5] ⟨http://www.seeingmachine.com⟩.
[6] Y. Matsumoto, A. Zelinsky, An algorithm for real time stereo vision implementation of head pose and gaze direction measurement, in: Proceedings of the Fourth International Conference on Automatic Face and Gesture Recognition, Grenoble France, March 2000, pp. 499–504.
[7] ⟨http://www.eyealert.com⟩.
[8] E. Wahlstrom, O. Masoud, N. Papanikolopoulos, Monitoring Driver Activities ITS Institute, Research Reports No. CTS 04-05, University of Minnesota.
[9] C.H. Morimoto, D. Koons, A. Amir, M. Flickner, Pupil detection and tracking using multiple light sources, Image Vision Comput. 18 (4) (2000) 331–335.
[10] Z. Zhu, Q. Ji, Eye and gaze tracking for interactive graphic display, Mach. Vision Appl. 15 (3) (2004) 139–148.
[11] Z. Zhu, Q. Ji, P. Lan, Real time non intrusive monitoring and prediction of driver fatigue, IEEE Trans. Vehicular Technol. 53 (4) (2004) 1052–1068.
[12] A. Haro, M. Flickner, I. Essa, Detecting and tracking eyes by using their physiological properties, dynamics, and appearance, in: Proceedings of IEEE CVPR 2000, Hilton Head Island, South Carolina, June 2000, pp. 163–168.
[13] C. Cudalbu, B. Anastasiu, R. Radu, R. Cruceanu, E. Schmidt, E. Barth, Driver monitoring with a single high speed camera and IR illumination, in: International Symposium on Signals, Circuits and Systems, ISSCS 2005, Iasi Romania, vol. 1, July 2005, pp. 219–222.
[14] Q. Ji, X. Yang, Real time eye gaze and face pose tracking for monitoring driver vigilance, Real-Time Imaging 8 (2002) 357–377.
[15] E.M. Ayoob, R. Grace, A.M. Steinfeld, Driver-vehicle-interface (DVI) development of a drowsy driver detection and warning system for commercial vehicles, Technical Report CMU-RI-TR-05-46, Robotics Institute, Carnegie Mellon University, September 2005.
[16] H.A. Rowley, S. Baluja, T. Kanade, Neural network-based face detection, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1) (1998) 23–38.
[17] R. Hsu, M. Mottleb, A.K. Jain, Face detection in color images, IEEE Trans. Pattern Anal. Mach. Intell. 24 (5) (2002) 696–706.
[18] O. Jesorsky, K. Kirchberg, R. Frischholz, Robust face detection using the Hausdorff distance, in: Proceedings of the Third International Conference

on Audio and Video-based Biometric Person Authentication, Lecture Notes in Computer Science, vol. 2091, Halmstad, Sweden, Springer, Berlin, 6–8 June 2001, pp. 90–95.
[19] M.H. Yang, D. Kriegman, N. Ahuja, Detecting faces in images: a survey, IEEE Trans. Pattern Anal. Mach. Intell. 24 (1) (2002) 34–58.
[20] T. Kawaguchi, M. Rizon, Iris detection using intensity and edge information, Pattern Recognition 36 (2) (2003) 549–562.
[21] Y. Li, X.-l. Qi, Y.-j. Wang, Eye detection by using fuzzy template matching and feature-parameter based judgement, Pattern Recognition Lett. 22 (10) (2001) 1111–1124.
[22] S. Baskan, M. Bulut, V. Atalay, Projection based method for segmentation of human face and its evaluation, Pattern Recognition Lett. 23 (14) (2002) 1623–1629.
[23] Y. Ma, X. Ding, Z. Wang, N. Wang, Robust Precise eye location under probabilistic frame work, in: Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, Seoul Korea, May 2004, pp. 339–344.
[24] H. Kashima, H. Hongo, K. Kato, K. Yamamoto, A robust iris detection method of facial and eye movement, in: Proceedings of VI 2001 Vision Interface Annual Conference Ottawa, Canada, 7–9 June 2001, pp. 9–14.
[25] S. Sirohey, A. Rosenfiled, Eye detection in a face image using linear and nonlinear filters, Pattern Recognition 34 (7) (2001) 1367–1391.
[26] S. Sirohey, A. Rosenfiled, Z. Duric, A method of detection and tracking iris and eyelids in video, Pattern Recognition 35 (6) (2002) 1389–1401.
[27] P. Smith, M. Shah, N. da Vitoria Lobo, Determining driver visual attention with one camera, IEEE Trans. Intell. Transp. Syst. 4 (4) (2003) 205–218.
[28] G.C. Feng, P.C. Yuen, Multi cues eye detection on gray intensity image, Pattern Recognition 34 (5) (2001) 1033–1046.
[29] C.C. Chiang, W.K. Tai, M.T. Yang, Y.T. Huang, C. Jaung, A novel method for detecting lips, eyes and faces in real time, Real Time Imaging 9 (4) (2003) 277–287.
[30] P. Wang, M.B. Green, Q. Ji, J. Wayman, Automatic eye detection and its validation, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, USA, vol. 3, June 2005, pp. 164–164.
[31] P. Wang, Q. Ji, Learning Discriminant features for multiview face and eye detection automatic eye detection and its vadidation, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, USA, vol. 1, June 2005, pp. 373–379.
[32] V.L. Neale, T. Dingus, S. Klauer, J. Sudweeks, M. Goodman, An overview of the 100-car naturalistic study and findings, in: Proceedings of 19th International Conference on Enhanced Safety of Vehicles, Washington, DC, June 6–9, 2005.
[33] T. Von Jan,T. Karnahl, K. Seifert, J. Hilgenstock, R. Zobel, Don't sleep and drive-VW's fatigue detection technology, in: Proceedings of 19th International Conference on Enhanced Safety of Vehicles, Washington, DC, June 6–9, 2005.
[34] D.L. Smith, J. Chang, D. Cohen, J. Foley, R. Glasco, A simulation approach for evaluating the relative safety impact of driver distraction during secondary tasks, in: Proceedings of 12th World Congress on ITS, November 6–10, 2005, San Fancisco.
[35] P.S. Rau, Drowsy driver detection and warning system for commercial vehicle drivers: field operational test design, data analyses, and progress, in: Proceedings of 19th International Conference on Enhanced Safety of Vehicles, Washington, DC, June 6–9, 2005.
[36] T. D'Orazio, C. Guaragnella, M. Leo, A. Distante, A new algorithm for ball recognition using circle Hough transform and neural classifier, Pattern Recognition 37 (2004) 393–408.
[37] B.S. Everitt, D.J. Hand, Finite Mixture Distributions, Chapman & Hall, New York, 1981.
[38] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. Ser. B 39 (1) (1977) 1–38.
[39] J. Bilmes, A gentle tutorial on the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models, Technical Report ICSI-TR-97-021, International Computer Science Institute (ICSI), Berkeley, CA, 1997.

About the Author—TIZIANA D'ORAZIO recevied the degree in Computer Science from the Unviersity of Bari (Italy) in 1988. Since 1989 she has worked at the Institute of Signal and Image Processing of the Italian National Research Council (CNR). She is actually a researcher of the Institute of Intelligent Systems for Automation of the CNR. Her research interests include pattern recognition, artificial intelligence, image processing for robotic application and intelligent systems.

About the Author—MARCO LEO was born in Gallipoli, Lecce, Italy in 1974. He received the degree in Computer Science Engineering from the University of Lecce in 2001. Since then, he is a collaborator of research at the Italian National Research council (CNR), Institutue of Study of Intelligent Systems for Automation (ISSIA) in Bar, Italy. He is currently a researcher of the ISSIA. His research interests are in the area of image and signal processing, neural networks and pattern recognition.

About the Author—CATALDO GUARAGNELLA was born in Italy in 1964. He graduated in electronic engineering in 1990 at University of Bari, Italy, and received the Ph.D. degree in Telecommunications by the Politecnico di Bari in 1994. In 1996 he joined the Electrics and Electronics Department of Politecnico di Bari as an assistant professor in Telecommunications. His main research interests include signal and image and video processing/coding, motion estimation in video sequences and multidimensional signal processing.

About the Author—ARCANGELO DISTANTE received the degree in Computer Science from the University of Bari, Italy in 1976. He joined the National Nuclear Physics Institute until 1983 where he worked on various theoretical and computation aspects of 3D reconstruction and pattern recognition of nuclear events. Since 1984 he has been working with the Institute for Signal and Image Processing (IESI) of Italian National Research Council (CNR). Currently he is the coordinator of the Robot Vision Group and the Director of the Institute of Intelligent System for Automation (ISSIA-CNR). In 1996 he joined the University of Lecce where he is an associate professor in Theory and Practice of Image Processing at the Faculty of Engineering. His current research deals with computer vision, pattern recognition, machine learning, neural computation, robot navigation and architectures for computer vision. Dr. Arcangelo Distance is a member of IAPR and SPIE.