Visual Monitoring of Driver Inattention

Luis M. Bergasa, Jesús Nuevo, Miguel A. Sotelo, Rafael Barea, and Elena Lopez

Department of Electronics, University of Alcala, CAMPUS. 28805 Alcalá de Henares (Madrid), Spain,
bergasa@depeca.uah.es, jnuevo@depeca.uah.es, sotelo@depeca.uah.es, barea@depeca.uah.es,
elena@depeca.uah.es

1 Introduction

The increasing number of traffic accidents due to driver inattention has become a serious problem for society. Every year, about 45,000 people die and 1.5 million people are injured in traffic accidents in Europe. These figures imply that one person out of every 200 European citizens is injured in a traffic accident every year and that around one out 80 European citizens dies 40 years short of the life expectancy. It is known that the great majority of road accidents (about 90–95%) are caused by human error. More recent data has identified inattention (including distraction and falling asleep at the wheel) as the primary cause of accidents, accounting for at least 25% of the crashes [15]. Road safety is thus a major European health problem. In the "White Paper on European Transport Policy for 2010," the European Commission declares the ambitious objective of reducing by 50% the number of fatal accidents on European roads by 2010 (European Commission, 2001).

According to the U.S. National Highway Traffic Safety Administration (NHTSA), falling asleep while driving is responsible for at least 100,000 automobile crashes annually. An annual average of roughly 70,000 nonfatal injuries and 1,550 fatalities results from these crashes [32, 33]. These figures only cover crashes happening between midnight and 6 a.m., involving a single vehicle and a sober driver traveling alone, including the car departing from the roadway without any attempt to avoid the crash. These figures underestimate the true level of the involvement of drowsiness because they do not include crashes at daytime hours involving multiple vehicles, alcohol, passengers or evasive maneuvers. These statistics do not deal with crashes caused by driver distraction either, which is believed to be a larger problem. Between 13 and 50% of crashes are attributed to distraction, resulting in as many as 5,000 fatalities per year. Increasing use of in-vehicle information systems (IVISs) such as cell phones, GPS navigation systems, satellite radios and DVDs has exacerbated the problem by introducing additional sources of distraction. That is, the more IVISs the more sources of distraction from the most basic task at hand, i.e., driving the vehicle. Enabling drivers to benefit from IVISs without diminishing safety is an important challenge.

This chapter presents an original system for monitoring driver inattention and alerting the driver when he is not paying adequate attention to the road in order to prevent accidents. According to [40] the driver inattention status can be divided into two main categories: distraction detection and identifying sleepiness. Likewise, distraction can be divided in two main types: visual and cognitive. Visual distraction is straightforward, occurring when drivers look away from the roadway (e.g., to adjust a radio). Cognitive distraction occurs when drivers think about something not directly related to the current vehicle control task (e.g., conversing on a hands-free cell phone or route planning). Cognitive distraction impairs the ability of drivers to detect targets across the entire visual scene and causes gaze to be concentrated in the center of the driving scene. This work is focused in the sleepiness category. However, sleepiness and cognitive distraction partially overlap since the context awareness of the driver is related to both, which represent mental occurrences in humans [26].

The rest of the chapter is structured as follows. In Sect. 2 we present a review of the main previous work in this direction. Section 3 describes the general system architecture, explaining its main parts. Experimental

L.M. Bergasa et al.: *Visual Monitoring of Driver Inattention*, Studies in Computational Intelligence (SCI) **132**, 19–37 (2008)
www.springerlink.com

results are shown in Sect. 4. Section 5 discusses weaknesses and improvements of the system. Finally, in Sect. 6 we present the conclusions and future work.

2 Previous Work

In the last few years many researchers have been working on systems for driver inattention detection using different techniques. The most accurate techniques are based on physiological measures like brain waves, heart rate, pulse rate, respiration, etc. These techniques are intrusive, since they need to attach some electrodes on the drivers, causing annoyance to them. Daimler-Chrysler has developed a driver alertness system, Distronic [12], which evaluates the EEG (electroencephalographic) patterns of the driver under stress. In advanced safety vehicle (ASV) project by Toyota (see in [24]), the driver must wear a wristband in order to measure his heart rate. Others techniques monitor eyes and gaze movements using a helmet or special contact lens [3]. These techniques, though less intrusive, are still not acceptable in practice.

A driver's state of vigilance can also be characterized by indirect vehicle behaviors like lateral position, steering wheel movements, and time to line crossing. Although these techniques are not intrusive they are subject to several limitations such as vehicle type, driver experience, geometric characteristics, state of the road, etc. On the other hand, these procedures require a considerable amount of time to analyze user behaviors and thereby they do not work with the so-called micro-sleeps: when a drowsy driver falls asleep for some seconds on a very straight road section without changing the lateral position of the vehicle [21]. To this end we can find different experimental prototypes, but at this moment none of them has been commercialized. Toyota uses steering wheel sensors (steering wheel variability) and pulse sensor to record the heart rate, as mentioned above [24]. Mitsubishi has reported the use of steering wheel sensors and measures of vehicle behavior (such as lateral position of the car) to detect driver drowsiness in their advanced safety vehicle system [24]. Daimler Chrysler has developed a system based on vehicle speed, steering angle and vehicle position relative to road delimitation (recorded by a camera) to detect if the vehicle is about to leave the road [11]. Volvo Cars recently announced its Driver Alert Control system [39], that will be available on its high-end models from 2008. This system uses a camera, a number of sensors and a central unit to monitor the movements of the car within the road lane and to assess whether the driver is drowsy.

People in fatigue show some visual behaviors easily observable from changes in their facial features like eyes, head, and face. Typical visual characteristics observable from the images of a person with reduced alertness level include longer blink duration, slow eyelid movement, smaller degree of eye opening (or even closed), frequent nodding, yawning, gaze (narrowness in the line of sight), sluggish facial expression, and drooping posture. Computer vision can be a natural and non-intrusive technique for detecting visual characteristics that typically characterize a driver's vigilance from the images taken by a camera placed in front of the user. Many researches have been reported in the literature on developing image-based driver alertness using computer vision techniques. Some of them are primarily focused on head and eye tracking techniques using two cameras [27, 38]. In [34] a system called FaceLAB developed by the company Seeing Machines is presented. The 3D pose of the head and the eye-gaze direction are calculated accurately. FaceLAB also monitors the eyelids, to determine eye opening and blink rates. With this information the system estimates the driver's fatigue level. According to FaceLab information, the system operates day and night but at night the performance of the system decreases. All systems explained above rely on manual initialization of feature points. The systems appear to be robust but the manual initialization is a limitation, although it simplifies the whole problem of tracking and pose estimation.

There are other proposals that use only a camera. In [6] we can find a 2D pupil monocular tracking system based on the differences in color and reflectivity between the pupil and iris. The system monitors driving vigilance by studying the eyelid movement. Another successful system of head/eye monitoring and tracking for drowsiness detection using one camera, which is based on color predicates, is presented in [37]. This system is based on passive vision techniques and its functioning can be problematical in poor or very bright lighting conditions. Moreover, it does not work at night, when the monitoring is more important.

In order to work at nights some researches use active illumination based on infrared LED. In [36] a system using 3D vision techniques to estimate and track the 3D line of sight of a person using multiple cameras

is proposed. The method relies on a simplified eye model, and it uses the Purkinje images of an infrared light source to determine eye location. With this information, the gaze direction is estimated. Nothing about monitoring driver vigilance is presented. In [23] a system with active infrared LED illumination and a camera is implemented. Because of the LED illumination, the method can easily find the eyes and based on them, the system locates the rest of the facial features. They propose to analytically estimate the local gaze direction based on pupil location. They calculate eyelid movement and face orientation to estimate driver fatigue. Almost all the active systems reported in the literature have been tested in simulated environments but not in real moving vehicles. A moving vehicle presents new challenges like variable lighting, changing background and vibrations that must be taken into account in real systems. In [19] an industrial prototype called *Copilot* is presented. This system uses infrared LED illumination to find the eyes and it has been tested with truck's drivers in real environments. It uses a simple subtraction process to find the eyes and it only calculates a validated parameter called PERCLOS (percent eye closure), in order to measure driver's drowsiness. This system currently works under low light conditions. Recently, Seeing Machines has presented a commercial system called driver state sensor (DSS) [35] for driver fatigue detection in transportation operation. The system utilizes a camera and LEDs for night working. It calculates PERCLOS and obtains a 93% correlation with drowsiness.

Systems relying on a single visual cue may encounter difficulties when the required visual features cannot be acquired accurately or reliably, as happens in real conditions. Furthermore, a single visual cue may not always be indicative of the overall mental condition [23]. The use of multiple visual cues reduces the uncertainty and the ambiguity present in the information from a single source. The most recent research in this direction use this hypothesis. The European project AWAKE (2001–2004) [22] proposes a multi-sensor system adapted to the driver, the vehicle, and the environment in an integrated way. This system merges, via an artificial intelligent algorithm, data from on-board driver monitoring sensors (such as an eyelid camera and a steering grip sensor) as well as driver behavior data (i.e., from lane tracking sensor, gas/brake and steering wheel positioning). The system must be personalized for each driver during a learning phase. Another European project, SENSATION (2004–2007) [16] is been currently founded to continue research of the AWAKE project in order to obtain a commercial system. The European project AIDE – Adaptive Integrated Driver–Vehicle InterfacE (2004–2008) [15] works in this direction as well.

This chapter describes a real-time prototype system based on computer vision for monitoring driver vigilance using active infrared illumination and a single camera placed on the car dashboard. We have employed this technique because our goal is to monitor a driver in real conditions (vehicle moving) and in a very robust and accurate way mainly at nights (when the probability to crash due to drowsiness is the highest). The proposed system does not need manual initialization and monitors several visual behaviors that typically characterize a person's level of alertness while driving. In a different fashion than other previous works, we have fused different visual cues from one camera using a fuzzy classifier instead of different cues from different sensors. We have analyzed different visual behaviors that characterize a drowsy driver and we have studied the best fusion for optimal detection. Moreover, we have tested our system during several hours in a car moving on a motorway and with different users. The basics of this work were presented by the authors in [5].

3 System Architecture

The general architecture of our system is shown in Fig. 1. It consists of four major modules: (1) Image acquisition, (2) Pupil detection and tracking, (3) Visual behaviors and (4) Driver monitoring. Image acquisition is based on a low-cost CCD micro-camera sensitive to near-IR. The pupil detection and tracking stage is responsible for segmentation and image processing. Pupil detection is simplified by the "bright pupil" effect, similar to the red eye effect in photography. Then, we use two Kalman filters in order to track the pupils robustly in real-time. In the visual behaviors stage we calculate some parameters from the images in order to detect some visual behaviors easily observable in people in fatigue: slow eyelid movement, smaller degree of eye opening, frequent nodding, blink frequency, and face pose. Finally, in the driver monitoring stage we

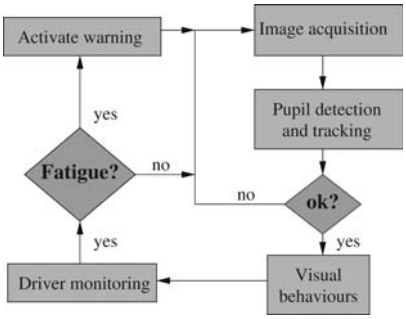


Fig. 1. General architecture

(a) Image obtained with inner IR ring

(b) Image obtained with outer IR ring

(c) Difference image

Fig. 2. Captured images and their subtraction

fuse all the individual parameters obtained in the previous stage using a fuzzy system, yielding the driver inattentiveness level. An alarm is activated if this level is over a certain threshold.

3.1 Image Acquisition System

The purpose of this stage is to acquire the video images of the driver's face. In this application the acquired images should be relatively invariant to light conditions and should facilitate the eye detection and tracking (good performance is necessary). The use of near-IR illuminator to brighten the driver's face serves these goals [25]. First, it minimizes the impact of changes in the ambient light. Second, the near-IR illumination is not detected by the driver, and then, this does not suppose an interference with the user's driving. Third, it produces the bright pupil effect, which constitutes the foundation of our detection and tracking system. A bright pupil is obtained if the eyes are illuminated with an IR illuminator beaming light along the camera optical axis. At the IR wavelength, the retina reflects almost all the IR light received along the path back to the camera, and a bright pupil effect will be produced in the image. If illuminated off the camera optical axis, the pupils appear dark since the reflected light of the retina will not enter the camera lens. An example of the bright/dark pupil effect can be seen in Fig. 2. This pupil effect is clear with and without glasses, with contact lenses and it even works to some extent with sunglasses.

Figure 3 shows the image acquisition system configuration. It is composed by a miniature CCD camera sensitive to near-IR and located on the dashboard of the vehicle. This camera focuses on the driver's head for detecting the multiple visual behaviors. The IR illuminator is composed by two sets of IR LEDs distributed symmetrically along two concentric and circular rings. An embedded PC with a low cost frame-grabber is used for video signal acquisition and signal processing. Image acquiring from the camera and LED excitation

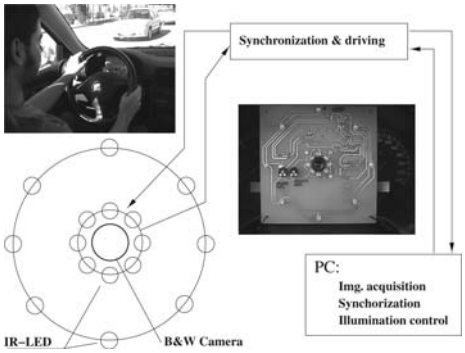


Fig. 3. Block diagram of the prototype

(a) Out-of-the-road lights effect (b) Vehicle lights effect (c) Sunlight effect (d) Sunlight effect with filter

Fig. 4. Effects of external lights in the acquisition system

is synchronized. The LED rings illuminate the driver's face alternatively, one for each image, providing different lighting conditions for almost the same scene.

Ring sizes has been empirically calculated in order to obtain a dark pupil image if the outer ring is turned on and a bright pupil image if the inner ring is turned on. LEDs in the inner ring are as close as possible to the camera, in order to maximize the "bright pupil" effect. The value of the outer ring radius is a compromise between the resulting illumination, that improves as it is increased, and the available space in the car's dashboard. The symmetric position of the LEDs in the rings, around the camera optical axis, cancels shadows generated by LEDs. The inner ring configuration obtains the bright pupil effect because the center of the ring coincides with the camera optical axis, working as if there were an only LED located on the optical axis of the lens. The outer ring provides ambient illumination that is used for contrast enhancing. In spite of those LEDs producing the dark pupil effect, a glint can be observed on each pupil.

The explained acquisition system works very well under controlled light conditions, but real scenarios present new challenges that must be taken into account. Lighting conditions were one of the most important problems to solve in real tests. As our system is based on the reflection of the light emitted by the IR LEDs, external light sources are the main source of noise. Three main sources can be considered, as are depicted in Fig. 4: artificial light from elements just outside the road (such as light bulbs), vehicle lights, and sun light. The effect of lights from elements outside the road mainly appears in the lower part of the image (Fig. 4a) because they are situated above the height of the car and the beam enters the car with a considerable angle. Then, this noise can be easily filtered. On the other hand, when driving on a double direction road, vehicle lights directly illuminate the driver, increasing the pixels level quickly and causing the pupil effect to disappear (Fig. 4b). Once the car has passed, the light level reduces very fast. Only after a few frames, the automatic gain controller (AGC) integrated in the camera compensates the changes, so very light and dark images are obtained, affecting the performance of the inner illumination system.

Regarding the sun light, it only affects at day time but its effect changes as function of the weather (sunny, cloudy, rainy, etc.) and the time of the day. With the exception of the sunset, dawn and cloudy days, sun light hides the inner infrared illumination and then the pupil effect disappears (Fig. 4c). For minimizing interference from light sources beyond the IR light emitted by our LEDs, a narrow band-pass filter, centered at the LED wavelength, has been attached between the CCD and the lens. This filter solved the problem of artificial lights and vehicle light almost completely, but it adds a new drawback for it reduces the intensity of the image, and then the noise is considerably amplified by the AGC. The filter does not eliminate the sun light interference, except for cases when the light intensity is very low. This is caused by the fact that the power emitted by the sun in the band of the filter is able to hide the inner illumination. An image of this case, taken by the sunset, is depicted in Fig. 4d. A possible solution for this problem could be the integration of IR filters in the car glasses.

3.2 Pupil Detection and Tracking

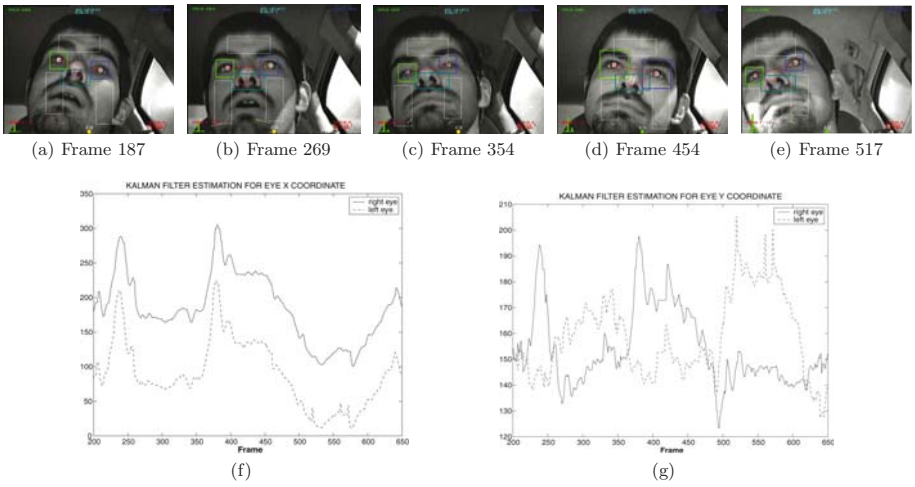
This stage starts with pupil detection. As mentioned above, each pair of images contains an image with bright pupil and another one with a dark pupil. The first image is then digitally subtracted from the second to produce the difference image. In this image, pupils appear as the brightest parts in the image as can be seen in Fig. 2. This method minimizes the ambient light influence by subtracting it in the generation of the difference image. This procedure yields high contrast images where the pupils are easily found. It can be observed that the glint produced by the outer ring of LEDs usually falls close to the pupil, with the same grey level as the bright pupil. The shape of the pupil blob in the difference image is not a perfect ellipse because the glint cuts the blob, affecting the modeling of the pupil blobs and, consequently, the calculation depending on it, as will be explained later. This is the reason why the system only uses subtracted images during initialization, and when light conditions are poor (this initialization time varies depending on the driver and light conditions, but it was below 5 s for all test). In other cases, only the image obtained with the inner ring is processed, increasing accuracy and reducing computation time.

Pupils are detected on the resulting image, by searching the entire image to locate two bright blobs that satisfy certain constraints. The image is binarized, using an adaptive threshold, for detecting the brighter blobs in the image.

A standard 8-connected components analysis [18] is then applied to the binarized difference image to identify binary blobs that satisfy certain size and shape constraints. The blobs that are out of some size constraints are removed, and for the others an ellipse model is fit to each one. Depending on their size, intensity, position and distance, best candidates are selected, and all the possible pairs between them are evaluated. The pair with the highest qualification is chosen as the detected pupils, and its centroids are returned as the pupil positions.

One of the main characteristics of this stage is that it is applicable to any user without any supervised initialization. Nevertheless, the reflection of the IR in the pupils under the same conditions varies from one driver to another. Even for the same driver, the intensity depends on the gaze point, head position and the opening of the eye. Apart from those factors, lighting conditions change with time, which modifies the intensity of the pupils. On the other hand, the size of the pupils also depends on the user, and the distance to the camera. To deal with those differences in order to be generic, our system uses an adaptive threshold in the binarization stage. The parameters of the detected pupils are used to update the statistics that set thresholds and margins in the detection process. Those statistics include size, grey level, position and apparent distance and angle between pupils, calculated over a time window of 2 s. The thresholds also get their values modified if the pupils are not found, widening the margins to make more candidates available to the system.

Another question related to illumination that is not usually addressed in the literature is the sensitivity of the eye to IR emission. As the exposure time to the IR source increases, its power has to be reduced in order to avoid damaging the internal tissues of the eye. This imposes a limit on the emission of the near-IR LEDs. To calculate the power of our system, we have followed the recommendations of [1], based on IEC 825-1 and CENELEC 60825-1 infrared norms. With these limitations, no negative effects have been reported in the drivers that collaborated in the tests.



(a) Frame 187 (b) Frame 269 (c) Frame 354 (d) Frame 454 (e) Frame 517

(f) (g)

Fig. 5. Tracking results for a sequence

To continuously monitor the driver it is important to track his pupils from frame to frame after locating the eyes in the initial frames. This can be done efficiently by using two Kalman filters, one for each pupil, in order to predict pupil positions in the image. We have used a pupil tracker based on [23] but we have tested it with images obtained from a car moving on a motorway. Kalman filters presented in [23] works reasonably well under frontal face orientation with open eyes. However, it will fail if the pupils are not bright due to oblique face orientations, eye closures, or external illumination interferences. Kalman filter also fails when a sudden head movement occurs because the assumption of smooth head motion has not been fulfilled. To overcome this limitation we propose a modification consisting on an adaptive search window, which size is determined automatically, based on pupil position, pupil velocity, and location error. This way, if Kalman filtering tracking fails in a frame, the search window progressively increases its size. With this modification, the robustness of the eye tracker is significantly improved, for the eyes can be successfully found under eye closure or oblique face orientation.

The state vector of the filter is represented as $\mathbf{x}_t = (\mathbf{c}_t, \mathbf{r}_t, \mathbf{u}_t, \mathbf{v}_t)$, where $(\mathbf{c}_t, \mathbf{r}_t)$ indicates the pupil pixel position (its centroid) and $(\mathbf{u}_t, \mathbf{v}_t)$ is its velocity at time t in $\mathbf{c}$ and $\mathbf{r}$ directions, respectively. Figure 5 shows an example of the pupil tracker working in a test sequence. Rectangles on the images indicate the search window of the filter, while crosses indicate the locations of the detected pupils. Figure 5f, g draws the estimation of the pupil positions for the sequence under test. The tracker is found to be rather robust for different users without glasses, lighting conditions, face orientations and distances between the camera and the driver. It automatically finds and tracks the pupils even with closed eyes and partially occluded eyes, and can recover from tracking-failures. The system runs at 25 frames per second.

Performance of the tracker gets worse when users wear eyeglasses because different bright blobs appear in the image due to IR reflections in the glasses, as can be seen in Fig. 6. Although the degree of reflection on the glasses depends on its material and the relative position between the user's head and the illuminator, in the real tests carried out, the reflection of the inner ring of LEDs appears as a filled circle on the glasses, of the same size and intensity as the pupil. The reflection of the outer ring appears as a circumference with bright points around it and with similar intensity to the pupil. Some ideas for improving the tracking with glasses are presented in Sect. 5. The system was also tested with people wearing contact lenses. In this case no differences in the tracking were obtained compared to the drivers not wearing them.

Fig. 6. System working with user wearing glasses

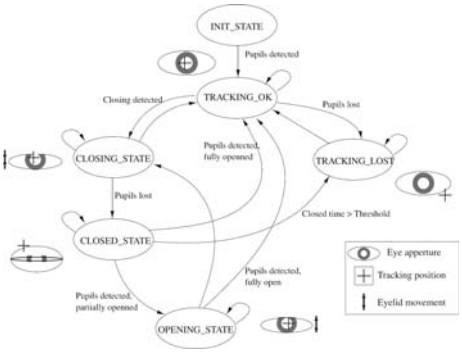


Fig. 7. Finite state machine for ocular measures

3.3 Visual Behaviors

Eyelid movements and face pose are some of the visual behaviors that reflect a person's level of inattention. There are several ocular measures to characterize sleepiness such as eye closure duration, blink frequency, fixed gaze, eye closure/opening speed, and the recently developed parameter PERCLOS [14, 41]. This last measure indicates the accumulative eye closure duration over time excluding the time spent on normal eye blinks. It has been found to be the most valid ocular parameter for characterizing driver fatigue [24]. Face pose determination is related to computation of face orientation and position, and detection of head movements. Frequent head tilts indicate the onset of fatigue. Moreover, the nominal face orientation while driving is frontal. If the driver faces in other directions for an extended period of time, it is due to visual distraction. Gaze fixations occur when driver's eyes are nearly stationary. Their fixation position and duration may relate to attention orientation and the amount of information perceived from the fixated location, respectively. This is a characteristic of some fatigue and cognitive distraction behaviors and it can be measured by estimating the fixed gaze. In this work, we have measured all the explained parameters in order to evaluate its performance for the prediction of the driver inattention state, focusing on the fatigue category.

To obtain the ocular measures we continuously track the subject's pupils and fit two ellipses, to each of them, using a modification of the LIN algorithm [17], as implemented in the OpenCV library [7]. The degree of eye opening is characterized by the pupil shape. As eyes close, the pupils start getting occluded by the eyelids and their shapes get more elliptical. So, we can use the ratio of pupil ellipse axes to characterize the degree of eye opening. To obtain a more robust estimation of the ocular measures and, for example, to distinguish between a blink and an error in the tracking of the pupils, we use a Finite State Machine (FSM) as we depict in Fig. 7. Apart from the *init_state*, five states have been defined: *tracking_ok*, *closing*, *closed*, *opening* and *tracking_lost*. Transitions between states are achieved from frame to frame as a function of the width-height ratio of the pupils.

The system starts at the *init_state*. When the pupils are detected, the FSM passes to the *tracking_ok* state indicating that the pupil's tracking is working correctly. Being in this state, if the pupils are not detected in a frame, a transition to the *tracking_lost* state is produced. The FSM stays in this state until the pupils are correctly detected again. In this moment, the FSM passes to the *tracking_ok* state. If the width-height ratio of the pupil increases above a threshold (20% of the nominal ratio), a closing eye action is detected and the FSM changes to the *closing_state*. Because the width-height ratio may increase due to other reasons, such as segmentation noise, it is possible to return to the *tracking_ok* state if the ratio does not constantly increase.

When the pupil ratio is above the 80% of its nominal size or the pupils are lost, being in *closing_state*, a transition of the FSM to *closed_state* is provoked, which means that the eyes are closed. A new detection of the pupils from the *closed_state* produces a change to *opening_state* or *tracking_ok* state, depending on the degree of opening of the eyelid. If the pupil ratio is between the 20 and the 80% a transition to the *opening_state* is produced, if it is below the 20% the system pass to the *tracking_ok* state. Being in *closed_state*, a transition to the *tracking_lost* state is produced if the closed time goes over a threshold. A transition from opening to closing is possible if the width-height ratio increases again. Being in *opening_state*, if the pupil ratio is below the 20% of the nominal ratio a transition to *tracking_ok* state is produced.

Ocular parameters that characterize eyelid movements have been calculated as a function of the FSM. PERCLOS is calculated from all the states, except from the *tracking_lost* state, analyzing the pupil width-height ratio. We consider that an eye closure occurs when the pupil ratio is above the 80% of its nominal size. Then, the eye closure duration measure is calculated as the time that the system is in the *closed_state*. To obtain a more robust measurement of the PERCLOS, we compute this running average. We compute this parameter by measuring the percentage of eye closure in a 30-s window. Then, PERCLOS measure represents the time percentage that the system is at the *closed_state* evaluated in 30 s and excluding the time spent in normal eye blinks. Eye closure/opening speed measures represent the amount of time needed to fully close the eyes or to fully open the eyes. Then, eye closure/opening speed is calculated as the time during which pupil ratio passes from 20 to 80% or from 80 to 20% of the nominal ratio, respectively. In other words, the time that the system is in the *closing_state* or *opening_state*, respectively. Blink frequency measure indicates the number of blinks detected in 30 s. A blink action will be detected as a consecutive transition among the following states: closing, closed, and opening, given that this action was carried out in less than a predefined time. Many physiology studies have been carried out on the blinking duration. We have used the recommendation value derived in [31] but this could be easily modified to conform to other recommended value. Respecting the eye nominal size used for the ocular parameters calculation, it varies depending on the driver. To calculate its correct value a histogram of the eyes opening degree for the last 2,000 frames not exhibiting drowsiness is obtained. The most frequent value on the histogram is considered to be the nominal size. PERCLOS is computed separately in both eyes and the final value is obtained as the mean of both.

Besides, face pose can be used for detecting fatigue or visual distraction behaviors among the categories defined for inattentive states. The nominal face orientation while driving is frontal. If the driver's face orientation is in other directions for an extended period of time it is due to visual distractions, and if it occurs frequently (in the case of various head tilts), it is a clear symptom of fatigue. In our application, the precise degree of face orientation for detecting this behaviors is not necessary because face poses in both cases are very different from the frontal one. What we are interested in is to detect whether the driver's head deviates too much from its nominal position and orientation for an extended period of time or too frequently (nodding detection).

This work provides a novel solution to the coarse 3D face pose estimation using a single un-calibrated camera, based on the method proposed in [37]. We use a model-based approach for recovering the face pose by establishing the relationship between 3D face model and its two-dimensional (2D) projections. A weak perspective projection is assumed so that face can be approximated as a planar object with facial features, such as eyes, nose and mouth, located symmetrically on the plane. We have performed a robust 2D face tracking based on the pupils and the nostrils detections on the images. Nostrils detection has been carried out in a way similar to that used for the pupils' detection. From these positions the 3D face pose is estimated, and as a function of it, face direction is classified in nine areas, from *upper left* to *lower right*.

This simple technique works fairly well for all the faces we tested, with left and right rotations specifically. A more detailed explanation about our method was presented by the authors in [5]. As the goal is to detect whether the face pose of the driver is not frontal for an extended period of time, this has been computed using only a parameter that gives the percentage of time that the driver has been looking at the front, over a 30-s temporal window.

Nodding is used to quantitatively characterize one's level of fatigue. Several systems have been reported in the literature to calculate this parameter from a precise estimation of the driver's gaze [23, 25]. However, these systems have been tested in laboratories but not in real moving vehicles. The noise introduced in real environments makes these systems, based on exhaustive gaze calculation, work improperly. In this work, a new technique based on position and speed data from the Kalman filters used to track the pupils and the FSM is proposed. This parameter measures the number of head tilts detected in the last 2 min. We have experimentally observed that when a nodding is taking place, the driver closes his or her eyes and the head goes down to touch the chest or the shoulders. If the driver wakes up in that moment, raising his head, the values of the vertical speed of the Kalman filters will change their sign, as the head rises. If the FSM is in *closed_state* or in *tracking_lost* and the pupils are detected again, the system saves the speeds of the pupils trackers for ten frames. After that, the data is analyzed to find if it conforms to that of a nodding. If so, the first stored value is saved and used as an indicator of the "magnitude" of the nodding.

Finally, one of the remarkable behaviors that appear in drowsy drivers or cognitively distracted drivers is fixed gaze. A fatigued driver looses the focus of the gaze, not paying attention to any of the elements of the traffic. This loss of concentration is usually correlated with other sleepy behaviors such as a higher blink frequency, a smaller degree of eye opening and nodding. In the case of cognitive distraction, however, fixed gaze is decoupled from other clues. As for the parameters explained above, the existing systems calculate this parameter from a precise estimation of the driver's gaze and, consequently, experience the same problems. In order to develop a method to measure this behavior in a simple and robust way, we present a new technique based on the data from the Kalman filters used to track the pupils.

An attentive driver moves his eyes frequently, focusing to the changing traffic conditions, particularly if the road is busy. This has a clear reflection on the difference between the estimated position from the Kalman filters and the measured ones.

Besides, the movements of the pupils for an inattentive driver present different characteristics. Our system monitors the position on the x coordinate. Coordinate y is not used, as the difference between drowsy and awake driver is not so clear. The fixed gaze parameter is computed locally over a long period of time, allowing for freedom of movement of the pupil over time. We refer here to [5] for further details of the computation of this parameter.

This fixed gaze parameter may suffer from the influence of vehicle vibrations or bumpy roads. Modern cars have reduced vibrations to a point that the effect is legible on the measure. The influence of bumpy roads depends on their particular characteristics. If bumps are occasional, it will only affect few values, making little difference in terms of the overall measure. On the other hand, if bumps are frequent and their magnitude is high enough, the system will probably fail to detect this behavior. Fortunately, the probability for a driver to get distracted or fall asleep is significantly lower in very bumpy roads. The results obtained for all the test sequences with this parameter are encouraging. In spite of using the same a priori threshold for different drivers and situations, the detection was always correct. Even more remarkable was the absence of false positives.

3.4 Driver Monitoring

This section describes the method to determine the driver's visual inattention level from the parameters obtained in the previous section. This process is complicated because several uncertainties may be present. First, fatigue and cognitive distractions are not observable and they can only be inferred from the available information. In fact, this behavior can be regarded as the result of many contextual variables such as environment, health, and sleep history. To effectively monitor it, a system that integrates evidences from multiple sensors is needed. In the present work, several fatigue visual behaviors are subsequently combined to form an inattentiveness parameter that can robustly and accurately characterize one's vigilance level.

The fusion of the parameters has been obtained using a fuzzy system. We have chosen this technique for its well known linguistic concept modeling ability. Fuzzy rule expressions are close to expert natural language. Then, a fuzzy system manages uncertain knowledge and infers high level behaviors from the observed data. As an universal approximator, fuzzy inference system can be used for knowledge induction processes. The objective of our fuzzy system is to provide a driver's inattentiveness level (DIL) from the fusion of several ocular and face pose measures, along with the use of expert and induced knowledge. This knowledge has been extracted from the visual observation and the data analysis of the parameters in some simulated fatigue behavior carried out in real conditions (driving a car) with different users. The simulated behaviors have been done according to the physiology study of the US Department of Transportation, presented in [24]. We do not delve into the psychology of driver visual attention, rather we merely demonstrate that with the proposed system, it is possible to collect driver information data and infer whether the driver is attentive or not.

The first step in the expert knowledge extraction process is to define the number and nature of the variables involved in the diagnosis process according to the domain expert experience. The following variables are proposed after appropriate study of our system: PERCLOS, eye closure duration, blink frequency, nodding frequency, fixed gaze and frontal face pose. Eye closing and opening variables are not being used in our input fuzzy set because they mainly depend on factors such as segmentation and correct detection of the eyes, and they take place in the length of time comparable to that of the image acquisition. As a consequence, they are very noisy variables. As our system is adaptive to the user, the ranges of the selected fuzzy inputs are approximately the same for all users. The fuzzy inputs are normalized, and different linguistic terms and its corresponding fuzzy sets are distributed in each of them using induced knowledge based on the hierarchical fuzzy partitioning (HFP) method [20]. Its originality lies in not yielding a single partition, but a hierarchy including partitions with various resolution levels based on automatic clustering data. Analyzing the fuzzy partitions obtained by HFP, we determined that the best suited fuzzy sets and the corresponding linguistic terms for each input variable are those shown in Table 1. For the output variable (DIL), the fuzzy set and the linguistic terms were manually chosen. The inattentiveness level range is between 0 and 1, with a normal value up to 0.5. When its value is between 0.5 and 0.75, driver's fatigue is medium, but if the DIL is over 0.75 the driver is considered to be fatigued, and an alarm is activated. Fuzzy sets of triangular shape were chosen, except at the domain edges, where they were semi-trapezoidal.

Based on the above selected variables, experts state different pieces of knowledge (rules) to describe certain situations connecting some symptoms with a certain diagnosis. These rules are of the form *"If condition, Then conclusion"*, where both premise and conclusion use the linguistic terms previously defined, as in the following example:

- **IF** *PERCLOS* is *large* **AND** *Eye Closure Duration* is *large*, **THEN** *DIL* is *large*

In order to improve accuracy and system design, automatic rule generation and its integration in the expert knowledge base were considered. The fuzzy system implementation used the licence-free tool Knowledge Base Configuration Tool (KBCT) [2] developed by the Intelligent Systems Group of the Polytechnics University of Madrid (UPM). A more detailed explanation of this fuzzy system can be found in [5].

Table 1. Fuzzy variables

Variable	Type	Range	Labels	Linguistic terms
PERCLOS	In	[0.0, 1.0]	5	Small, medium small, medium, medium large, large
Eye closure duration	In	[1.0–30.0]	3	Small, medium, large
Blink freq.	In	[1.0–30.0]	3	Small, medium, large
Nodding freq.	In	[0.0–8.0]	3	Small, medium, large
Face position	In	[0.0–1.0]	5	Small, medium small, medium, medium large, large
Fixed gaze	In	[0.0–0.5]	5	Small, medium small, medium, medium large, large
DIL	Out	[0.0–1.0]	5	Small, medium small, medium, medium large, large

4 Experimental Results

The goal of this section is to experimentally demonstrate the validity of our system in order to detect fatigue behaviors in drivers. Firstly, we show some details about the recorded video sequences used for testing, then, we analyze the parameters measured for one of the sequences. Finally, we present the performance of the detection of each one of the parameters, and the overall performance of the system.

4.1 Test Sequences

Ten sequences were recorded in real driving situations over a highway and a two-direction road. Each sequence was obtained for a different user. The images were obtained using the system explained in Sect. 3.1. The drivers simulated some drowsy behaviors according to the physiology study of the US Department of Transportation presented in [24]. Each user drove normally except in one or two intervals where the driver simulated fatigue. Simulating fatigue allows for the system to be tested in a real motorway, with all the sources of noise a deployed system would face. The downside is that there may be differences between an actual drowsy driver and a driver mimicking the standard drowsy behavior, as defined in [24]. We are currently working on testing the system in a truck simulator.

The length of the sequences and the fatigue simulation intervals are shown in Table 2. All the sequences were recorded at night except for sequence number 7 that was recorded at day, and sequence number 5 that was recorded at sunset. Sequences were obtained with different drivers not wearing glasses, with the exception of sequence 6, that was recorded for testing the influence of the glasses in real driving conditions.

4.2 Parameter Measurement for One of the Test Sequences

The system is currently running on a PC Pentium4 (1.8 Ghz) with Linux kernel 2.6.18 in real time (25 pairs of frames/s) with a resolution of 640×480 pixels. Average processing time per pair of frames is 11.43 ms. Figure 8 depicts the parameters measured for sequence number 9. This is a representative test example with a duration of 465 s where the user simulates two fatigue behaviors separated by an alertness period. As can be seen, until second 90, and between the seconds 195 and 360, the DIL is below 0.5 indicating an alertness state. In these intervals the PERCLOS is low (below 0.15), eye closure duration is low (below the 200 ms), blink frequency is low (below two blinks per 30-s window) and nodding frequency is zero. These ocular parameters indicate a clear alert behavior. The frontal face position parameter is not 1.0, indicating that the predominant position of the head is frontal, but that there are some deviations near the frontal position, typical of a driver with a high vigilance level. The fixed gaze parameter is low because the eyes of the driver are moving caused by a good alert condition. DIL increases over the alert threshold during two intervals (from 90 to 190 and from 360 to 565 s) indicating two fatigue behaviors. In both intervals the PERCLOS increases from 0.15 to 0.4, the eye closure duration goes up to 1,000 ms, and the blink frequency parameter

Table 2. Length of simulated drowsiness sequences

Seq. Num.	Drowsiness behavior time (s)	Alertness behavior time (s)	Total time (s)
1	394 (two intervals: 180 + 214)	516	910
2	90 (one interval)	210	300
3	0	240	240
4	155 (one interval)	175	330
5	160 (one interval)	393	553
6	180 (one interval)	370	550
7	310 (two intervals: 150 + 160)	631	941
8	842 (two intervals: 390 + 452)	765	1,607
9	210 (two intervals: 75 + 135)	255	465
10	673 (two intervals: 310 + 363)	612	1,285

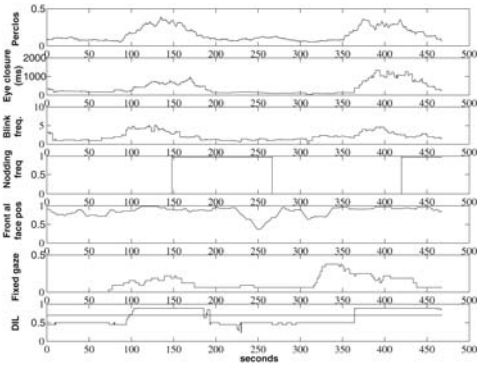


Fig. 8. Parameters measured for the test sequence number 9

Table 3. Parameter measurement performance

Parameters	Total % correct
PERCLOS	93.1
Eye closure duration	84.4
Blink freq.	79.8
Nodding freq.	72.5
Face pose	87.5
Fixed gaze	95.6

increases from 2 to 5 blinks. The frontal face position is very close to 1.0 because the head position is fixed and frontal. The fixed gaze parameter increases its value up to 0.4 due to the narrow gaze in the line of sight of the driver. This last variation indicates a typical loss of concentration, and it takes place before other sleepy parameters could indicate increased sleepiness, as can be observed. The nodding is the last fatigue effect to appear. In the two fatigue intervals a nodding occurs after the increase of the other parameters, indicating a low vigilance level. This last parameter is calculated over a temporal window of 2 min, so its value remains stable most of the time.

This section described an example of parameter evolution for two simulated fatigue behaviors of one driver. Then, we analyzed the behaviors of other drivers in different circumstances, according to the video tests explained above. The results obtained are similar to those shown for sequence number 9. Overall results of the system are explained in what follows.

4.3 Parameter Performance

The general performance of the measured parameters for a variety of environments with different drivers, according to the test sequences, is presented in Table 3. Performance was measured by comparing the algorithm results to results obtained by manually analyzing the recorded sequences on a frame-by-frame basis. Each frame was individually marked with the visual behaviors the driver exhibited, if any. Inaccuracies of this evaluation can be considered negligible for all parameters. Eye closure duration is not easy to evaluate accurately, as the duration of some quick blinks is around 5–6 frames at the rate of 25 frames per second (fps), and the starting of the blink can fall between two frames. However, the number of quick blinks is not big enough to make further statistical analysis necessary.

For each parameter the total correct percentage for all sequences excluding sequence number 6 (driver wearing glasses) and sequence number 7 (recorded during the day) is depicted. Then, this column shows the parameter detection performance of the system for optimal situations (driver without glasses driving at

night). The performance gets considerably worse by day and it dramatically decreases when drivers wear glasses.

PERCLOS results are quite good, obtaining a total correct percentage of 93.1%. It has been found to be a robust ocular parameter for characterizing driver fatigue. However, it may fail sometimes, for example, when a driver falls asleep without closing her eyes. Eye closure duration performance (84.4%) is a little worse than that of the PERCLOS, because the correct estimation of the duration is more critical. The variation on the intensity when the eye is partially closed with regard to the intensity when it is open complicates the segmentation and detection. This causes the frame count for this parameter to be usually less than the real one. These frames are considered as closed time. Measured time is slightly over the real time, as a result of delayed detection. Performance of blink frequency parameter is about 80% because some quick blinks are not detected at 25 fps. Then, the three explained parameters are clearly correlated almost linearly, and PERCLOS is the most robust and accurate one.

Nodding frequency results are the worst (72.5%), as the system is not sensible to noddings in which the driver rises her head and then opens her eyes. To reduce false positives, the magnitude of the nodding (i.e., the absolute value of the Kalman filter speed), must be over a threshold. In most of the non-detected noddings, the mentioned situation took place, while the magnitude threshold did not have any influence on any of them. The ground truth for this parameter was obtained manually by localizing the noddings on the recorded video sequences. It is not correlated with the three previous parameters, and it is not robust enough for fatigue detection. Consequently, it can be used as a complementary parameter to confirm the diagnosis established based on other more robust methods.

The evaluation of the face direction provides a measure of alertness related to drowsiness and visual distractions. This parameter is useful for both detecting the pose of the head not facing the front direction and the duration of the displacement. The results can be considered fairly good (87.5%) for a simple model that requires very little computation and no manual initialization. The ground truth in this case was obtained by manually looking for periods in which the driver is not clearly looking in front in the video sequences, and comparing their length to that of the periods detected by the system. There is no a clear correlation between this parameter and the ocular ones for fatigue detection. This would be the most important cue in case of visual distraction detection.

Performance of the fixed gaze monitoring is the best of the measured parameters (95.6%). The maximum values reached by this parameter depend on users' movements and gestures while driving, but a level above 0.05 is always considered to be an indicator of drowsiness. Values greater than 0.15 represent high inattentiveness probability. These values were determined experimentally. This parameter did not have false positives and is largely correlated with the frontal face direction parameter. On the contrary, it is not clearly correlated with the rest of the ocular measurements. For cognitive distraction analysis, this parameter would be the most important cue, as this type of distraction does not normally involve head or eye movements. The ground truth for this parameter was manually obtained by analyzing eye movements frame by frame for the intervals where a fixed gaze behavior was being simulated. We can conclude from these data that fixed gaze and PERCLOS are the most reliable parameters for characterizing driver fatigue, at least for our simulated fatigue study.

All parameters presented in Table 3 are fused in the fuzzy system to obtain the DIL for final evaluation of sleepiness. We compared the performance of the system using only the PERCLOS parameter and the DIL(using all of the parameters), in order to test the improvements of our proposal with respect to the most widely used parameter for characterizing driver drowsiness. The system performance was evaluated by comparing the intervals where the PERCLOS/DIL was above a certain threshold to the intervals, manually analyzed over the video sequences, in which the driver simulates fatigue behaviors. This analysis consisted of a subjective estimation of drowsiness by human observers, based on the Wierwille test [41].

As can be seen in Table 4, correct detection percentage for DIL is very high (97%). It is higher than the obtained using only PERCLOS, for which the correct detection percentage is about the 90% for our tests. This is due to the fact that fatigue behaviors are not the same for all drivers. Further, parameter evolution and absolute values from the visual cues differ from user to user. Another important fact is the delay between the moment when the driver starts his fatigue behavior simulation and when the fuzzy system detects it. This is a consequence of the window spans used in parameter evaluation. Each parameter responds to a

Table 4. Sleepiness detection performance

Parameter	Total % correct
PERCLOS	90
DIL	97

different stage in the fatigue behavior. For example, fixed gaze behavior appears before PERCLOS starts to increase, thus rising the DIL to a value where a noticeable increment of PERCLOS would rise an alarm in few seconds. This is extensible to the other parameters. Using only the PERCLOS would require much more time to activate an alarm (tens of seconds), especially if the PERCLOS increases more slowly for some drivers. Our system provides an accurate characterization of a driver's level of fatigue, using multiple visual parameters to resolve the ambiguity present in the information from a single parameter. Additionally, the system performance is very high in spite of the partial errors associated to each input parameter. This was achieved using redundant information.

5 Discussion

It has been shown that the system's weaknesses can be almost completely attributed to the pupil detection strategy, because it is the most sensitive to external interference. As it has been mentioned above, there are a series of situations where the pupils are not detected and tracked robustly enough. Pupil tracking is based on the "bright pupil" effect, and when this effect does not appear clearly enough on the images, the system can not track the eyes. Sunlight intensity occludes the near-IR reflected from the driver's eyes. Fast changes in illumination that the Automatic Gain Control in the camera can not follow produce a similar result. In both cases the "bright pupil" effect is not noticeable in the images, and the eyes can not be located. Pupils are also occluded when the driver's eyes are closed. It is then not possible to track the eyes if the head moves during a blink, and there is an uncertainty of whether the eyes may still be closed or they may have opened and appeared in a position on the image far away from where they were a few frames before. In this situation, the system would progressively extend the search windows and finally locate the pupils, but in this case the measured duration of the blink would not be correct. Drivers wearing glasses pose a different problem. "Bright pupil" effect appears on the images, but so do the reflections of the LEDs from the glasses. These reflections are very similar to the pupil's, making detection of the correct one very difficult.

We are exploring alternative approaches to the problem of pupil detection and tracking, using methods that are able to work 24/7 and in real time, and that yield accurate enough results to be used in other modules of the system. A possible solution is to use an eye or face tracker that does not rely on the "bright pupil" effect. Also, tracking the whole face, or a few parts of it, would make it possible to follow its position when eyes are closed, or occluded.

Face and eye location is an extensive field in computer vision, and multiple techniques have been developed. In recent years, probably the most successful have been texture-based methods and machine learning. A recent survey that compares some of these methods for eye localization can be found in [8]. We have explored the feasibility of using appearance (texture)-based methods, such as Active Appearance Models (AAM) [9]. AAM are generative models, that try to parameterize the contents of an image by generating a synthetic image as close as possible to the given one. The synthetic image is obtained from a model consisting of both appearance and shape. These appearance and shape are learned in a training process, and thus can only represent a constrained range of possible appearances and deformations. They are represented by a series of orthogonal vectors, usually obtained using Principal Component Analysis (PCA), that form a base in the appearance and deformation spaces.

AAMs are linear in both shape and appearance, but are nonlinear in terms of pixel intensities. The shape of the AAM is defined as the coordinates of the v vertices of the shape

$$\mathbf{s} = (x_1, y_1, x_2, y_2, \cdots, x_v, y_v)^t \tag{1}$$

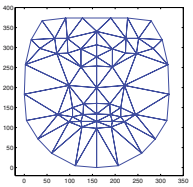


Fig. 9. A triangulated shape

and can be instantiated from the vector base simply as:

$$\mathbf{s} = \mathbf{s_0} + \sum_{i=1}^{n} p_i \cdot \mathbf{s_i} \tag{2}$$

where $\mathbf{s_0}$ is the *base shape* and $\mathbf{s_i}$ are the *shape vectors*. Appearance is instantiated in the same way

$$A(\mathbf{x}) = A_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \cdot A_i(\mathbf{x}) \tag{3}$$

where $A_0(\mathbf{x})$ is the *base appearance*, $A_i(\mathbf{x})$ are the *appearance vectors* and λ_i are the weights of these vectors.

The final model instantiation is obtained by warping the appearance $A(\mathbf{x})$, whose shape is $\mathbf{s_0}$, so it conforms to the shape $\mathbf{s}$. This is usually done by triangulating the vertices of the shape, using Delaunay [13] or another triangulation algorithm, as shown in Fig. 9. The appearance that falls in each triangle is affine warped independently, accordingly to the position of the vertices of the triangle in $\mathbf{s_0}$ and $\mathbf{s}$.

The purpose of fitting the model to a given image is to obtain the parameters that minimize the error between the image I and the model instance:

$$\sum_{\mathbf{x} \in \mathbf{s}_0} \left[A_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i A_i(\mathbf{x}) - I(\mathbf{W}(\mathbf{x}; \mathbf{p})) \right]^2 \tag{4}$$

where $\mathbf{W}(\mathbf{x}; \mathbf{p})$ is a warp defined over the pixel positions $\mathbf{x}$ by the shape parameters $\mathbf{p}$.

These parameters can be then analyzed to gather interesting data, in our case, the position of the eyes and head pose. Minimization is done using the Gauss–Newton method, or some efficient variations, such as the *inverse compositional algorithm* [4, 28].

We tested the performance and robustness of the Active Appearance Models on the same in-car sequences described above. AAMs perform well in sequences where the IR-based system did not, such as sequence 6, where the driver is wearing glasses (Figs. 10a, b), and is able to work with sunlight (10c), and track the face under fast illumination changes (10d–f). Also, as the model covers most of the face, the difference between a blink and a tracking loss is clearer, as the model can be fitted when eyes are either open or closed.

On our tests, however, AAM was only fitted correctly when the percentage of occlusion (or self-occlusion, due to head turns) of the face was below 35% of the face. It was also able to fit with low error although the position of the eyes was not determined with the required precision (i.e., the triangles corresponding to the pupil were positioned closer to the corner of the eye than to the pupil). The IR-based system could locate and track an eye when the other eye was occluded, which the AAM-based system is not able to do. More detailed results can be found on [30].

Overall results of face tracking and eye localization with AAM are encouraging, but the mentioned shortcomings indicate that improved robustness is necessary. Constrained Local Models (CLM) are models closely related to AAM, that have shown improved robustness and accuracy [10]. Instead of covering the whole face, CLM only use small rectangular patches placed in specific points that are interesting for its characteristic appearance or high contrast. Constrained Local Models are trained in the same way as AAMs, and both a shape and appearance vector bases are obtained.

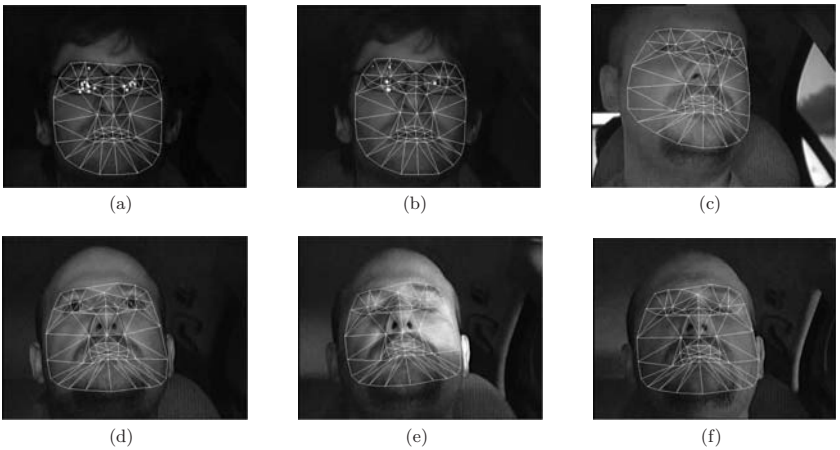


Fig. 10. Fitting results with glasses and sunlight

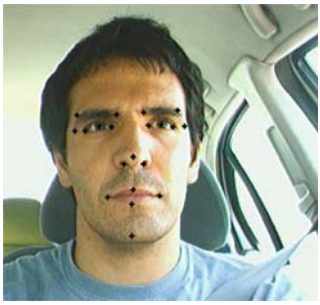


Fig. 11. A constrained local model fitted over a face

Fitting the CLM to an image is done in two steps. First, the same minimization that was used for AAMs is performed, with the difference that now no warping is applied over the rectangles. Those are only displaced over the image. In the second step, the correlation between the patches and the image is maximized, with an iterative algorithm, typically the Nelder–Mead simplex algorithm [29].

The use of small patches and the two-step fitting algorithm make CLM more robust and efficient than AAM. See Fig. 11 for an example. The CLM is a novel technique that performs well in controlled environments, but that has to be thoroughly tested in challenging operation scenarios.

6 Conclusions and Future Work

We have developed a non-intrusive prototype computer vision system for real-time monitoring of driver's fatigue. It is based on a hardware system for real time acquisition of driver's images using an active IR illuminator and the implementation of software algorithms for real-time monitoring of the six parameters that

better characterize the fatigue level of a driver. These visual parameters are PERCLOS, eye closure duration, blink frequency, nodding frequency, face pose and fixed gaze. In an attempt to effectively monitor fatigue, a fuzzy classifier was implemented to merge all these parameters into a single Driver Inattentiveness Level. Monitoring distractions (both visual and cognitive) would be possible using this system. The system development has been discussed. The system is fully autonomous, with automatic (re)initializations if required. It was tested with different sequences recorded in real driving condition with different users during several hours. In each of them, several fatigue behaviors were simulated during the test. The system works robustly at night for users not wearing glasses, yielding accuracy of 97%. Performance of the system decreases during the daytime, especially in bright days, and at the moment it does not work with drivers wearing glasses. A discussion about improvements of the system in order to overcome these weaknesses has been included.

The results and conclusions obtained support our approach to the drowsiness detection problem. In the future the results will be completed with actual drowsiness data. We have the intention of testing the system with more users for long periods of time, to obtain real fatigue behaviors. With this information we will generalize our fuzzy knowledge base. Then, we would like to improve our vision system with some of the techniques mentioned in the previous section, in order to solve the problems of daytime operation and to improve the solution for drivers wearing glasses. We also plan to add two new sensors (a steering wheel sensor and a lane tracking sensor) for fusion with the visual information to achieve correct detection, especially at daytime.

Acknowledgements

This work has been supported by grants TRA2005-08529-C02-01 (MOVICON Project) and PSE-370100-2007-2 (CABINTEC Project) from the Spanish Ministry of Education and Science (MEC). J. Nuevo is also working under a researcher training grant from the Education Department of the Comunidad de Madrid and the European Social Fund.

References

1. Inc. Agilent Technologies. *Application Note 1118: Compliance of Infrared Communication Products to IEC 825-1 and CENELEC EN 60825-1*, 1999.
2. J.M. Alonso, S. Guillaume, and L. Magdalena. KBCT, knowledge base control tool, 2003. URL http://www.mat.upm.es/projects/advocate/en/index.htm
3. Anon. Perclos and eyetracking: Challenge and opportunity. Technical report, Applied Science Laboratories, Bedford, MA, 1999. URL http://www.a-s-l.com
4. S. Baker and I. Matthews. Lucas-Kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(3):221–255, March 2004.
5. L.M. Bergasa, J. Nuevo, M.A. Sotelo, R. Barea, and M.E. Lopez. Real-time system for monitoring driver vigilance. *Intelligent Transportation Systems, IEEE Transactions on Intelligent Transportation Systems*, 7(1):63–77, 2006.
6. S. Boverie, J.M. Leqellec, and A. Hirl. Intelligent systems for video monitoring of vehicle cockpit. In *International Congress and Exposition ITS. Advanced Controls and Vehicle Navigation Systems*, pp. 1–5, 1998.
7. G. Bradski, A. Kaehler, and V. Pisarevsky. Learning-based computer vision with intel's open source computer vision library. *Intel Technology Journal*, 09(02), May 2005.
8. P. Campadelli, R. Lanzarotti, and G. Lipori. Eye localization: a survey. In *NATO Science Series*, 2006.
9. T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. *IEEE Transaction on Pattern Analysis an Machine Intelligence*, 23:681–685, 2001.
10. D. Cristinacce and T. Cootes. Feature Detection and Tracking with Constrained Local Models. *Proceedings of the British Machine Vision Conf*, 2006.
11. DaimerChryslerAG. The electronic drawbar, June 2001. URL http://www.daimlerchrysler.com
12. DaimlerChrysler. Driver assistant with an eye for the essentials. URL http://www.daimlerchrysler.com/dccom
13. B. Delaunay. Sur la sphere vide. *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk*, 7: 793–800, 1934.
14. D. Dinges and F. Perclos: A valid psychophysiological measure of alertness as assesed by psychomotor vigilance. Technical Report MCRT-98-006, Federal Highway Administration. Office of motor carriers, 1998.

15. European Project FP6 (IST-1-507674-IP). AIDE – Adaptive Integrated Driver-Vehicle Interface, 2004–2008. URL http://www.aide-eu.org/index.html
16. European Project FP6 (IST-2002-2.3.1.2). Advanced sensor development for attention, stress, vigilance and sleep/wakefulness monitoring (SENSATION), 2004–2007. URL http://www.sensation-eu.org
17. A.W. Fitzgibbon and R.B. Fisher. A buyer's guide to conic fitting. In *Proceedings of the 6th British Conference on Machine Vision*, volume 2, pp. 513–522, Birmingham, United Kingdom, 1995.
18. D.A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, 2003.
19. R. Grace. Drowsy driver monitor and warning system. In *International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Aug 2001.
20. S. Guillaume and B. Charnomordic. A new method for inducing a set of interpretable fuzzy partitions and fuzzy inference systems from data. *Studies in Fuzziness and Soft Computing*, 128:148–175, 2003.
21. H. Ueno, M. Kaneda, and M. Tsukino. Development of drowsiness detection system. In *Proceedings of Vehicle Navigation and Information Systems Conference*, pp. 15–20, 1994.
22. AWAKE Consortium (IST 2000-28062). *System for Effective Assessment of Driver Vigilance and Warning According to Traffic Risk Estimation – AWAKE*, Sep 2001–2004. URL http://www.awake-eu.org
23. Q. Ji and X. Yang. Real-time eye, gaze and face pose tracking for monitoring driver vigilance. *Real-Time Imaging*, 8:357–377, Oct 2002.
24. A. Kircher, M. Uddman, and J. Sandin. Vehicle control and drowsiness. Technical Report VTI-922A, Swedish National Road and Transport Research Institute, 2002.
25. D. Koons and M. Flicker. IBM Blue Eyes project, 2003. URL http://almaden.ibm.com/cs/blueeyes
26. M. Kutila. *Methods for Machine Vision Based Driver Monitoring Applications*. Ph.D. thesis, VTT Technical Research Centre of Finland, 2006.
27. Y. Matsumoto and A. Zelinsky. An algorithm for real-time stereo vision implementation of head pose and gaze direction measurements. In *Proceedings of IEEE 4th International Conference Face and Gesture Recognition*, pp. 499–505, Mar 2000.
28. I. Matthews and S. Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, November 2004.
29. J.A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7(4):308–313, 1965.
30. J. Nuevo, L.M. Bergasa, M.A. Sotelo, and M. Ocana. Real-time robust face tracking for driver monitoring. *Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE*, pp. 1346–1351, 2006.
31. L. Nunes and M.A. Recarte. *Cognitive demands of hands-free phone conversation while driving*, Chap. F5, pp. 133–144. Pergamon, Oxford, 2002.
32. P. Rau. Drowsy driver detection and warning system for commercial vehicle drivers: Field operational test design, analysis and progress, NHTSA, 2005.
33. D. Royal. Volume I – Findings; National Survey on Distracted and Driving Attitudes and Behaviours, 2002. Technical Report DOT HS 809 566, The Gallup Organization, March 2003.
34. Seeing Machines. Facelab transport, August 2006. URL http://www.seeingmachines.com/transport.html
35. Seeing Machines. Driver state sensor, August 2007. URL http://www.seeingmachines.com/DSS.html
36. W. Shih and Liu. A calibration-free gaze tracking technique. In *Proceedings of 15th Conference Patterns Recognition*, volume 4, pp. 201–204, Barcelona, Spain, 2000.
37. P. Smith, M. Shah, and N.Da.V. Lobo. Determining driver visual attention with one camera. *IEEE Transaction on Intelligent Transportation Systems*, 4(4):205–218, 2003.
38. T. Victor, O. Blomberg, and A. Zelinsky. Automating the measurement of driver visual behaviours using passive stereo vision. In *Proceedings of Intelligent Conference Series Vision in Vehicles VIV9*, Brisbane, Australia, Aug 2001.
39. Volvo Car Corporation. Driver alert control. URL http://www.volvocars.com
40. W. Wierwille, L. Tijerina, S. Kiger, T. Rockwell, E. Lauber, and A. Bittne. Final report supplement – task 4: Review of workload and related research. Technical Report DOT HS 808 467(4), USDOT, Oct 1996.
41. W. Wierwille, Wreggit, Kirn, Ellsworth, and Fairbanks. Research on vehicle-based driver status/performance monitoring; development, validation, and refinement of algorithms for detection of driver drowsiness, final report; technical reports & papers. Technical Report DOT HS 808 247, USDOT, Dec 1994. URL www.its.dot.gov