Abstract—We present a method to monitor driver distraction based on a stereo camera to estimate the face pose and gaze of a driver in real-time. A coarse eye direction is composed with the face pose estimation to obtain the gaze and driver’s fixation area in the scene, a parameter which gives much information about the distraction pattern of the driver. The system does not require any subject-specific calibration, it is robust to fast and wide head rotations and works in low lighting conditions.

The system provides some consistent statistics which help psychologists to assess the driver distraction patterns under influence of different In-Vehicle Information Systems (IVIS). These statistics are objective, as the drivers are not required to report their own distraction states. The proposed gaze fixation system has been tested on a set of challenging driving experiments directed by a team of psychologists in a naturalistic driving simulator. This simulator mimics conditions present in real driving, including weather changes, manoeuvring and distractions due to IVIS. Professional drivers participated in the tests.

Index Terms—Distraction monitoring, driver, inattention, gaze fixation, PRC, IVIS, naturalistic simulator

I. INTRODUCTION

Driving inattention is a major factor in traffic crashes. In the EU-27, 38,900 people died in 2008 in traffic accidents and 34,500 people lost their lives in 2009 [1]. That year, over 1.25 million accidents took place and more than 1.5 million people were injured [2]. On the other hand, the National Highway Traffic Safety Administration (NHTSA) estimates that approximately 25% of police-reported crashes involve some form of driving inattention, including fatigue and distraction [3].

Driving distraction is more diverse and implies a more risky factor that fatigue and it is present in over half of inattention involved crashes, resulting in as many as 5000 fatalities and $40 billion in damages each year [4].

Driving distraction is defined by the American Automobile Association Foundation for Traffic Safety (AAA FTS) as occurring “when a driver is delayed in the recognition of information needed to safely accomplish the driving task because some event, activity, object or person within or outside the vehicle compelled or tended to induce the driver’s shifting attention away from the driving task” [5]. Thirteen types of potentially distracting activities are listed [4]. Since the distracting activities take many forms, NHTSA classifies distraction into 4 categories from the view of the driver’s functionality [3]: visual distraction, cognitive distraction, auditory distraction (e.g., responding to a ringing cell phone), and biomechanical distraction (e.g., manually adjusting the radio volume). Many distracting activities can involve more than one of these components (e.g., talking to a phone while driving creates a biomechanical, auditory and cognitive distraction).

Increasing use of in-vehicle information systems (IVIS) such as cell phones, GPS, DVD players and other on-board devices has exacerbated the problem by introducing additional sources of distraction [6]. Enabling drivers to benefit from IVIS without diminishing safety is an important challenge.

One promising strategy to mitigate the effects of distraction involves monitoring and classifying the driver state and then using this classification to adapt the IVIS. Driver inattention monitoring has been an active research field for decades, mainly focusing in fatigue, and various methods have been proposed. Some auto companies have already installed some fatigue monitoring systems in their high-end vehicles. Yet, there is still a great need to develop a more reliable and fully functional system using cost efficient methods for a real driving context.

To date, realistic studies that provide information on the impact of distracting activities have been developed as small scale studies. An effort is needed to study distraction under naturalistic situations. Simulation is an optimal method of experimentation to acquire knowledge of driver’s behaviour, close to a real scenario but without the safety risks of having inattentive drivers in an open road. The simulation methodologies applied in Europe to the road transport sector research are demonstrating their profitability and efficiency [7]. The main objective through the simulation is to immerse the driver in his normal work environment. In order to do this, a fully equipped cockpit is required to perform the driving task. Previous work scoping the prediction of drivers behaviour mostly rely on lane position and vehicle sensors. Also, a video signal of the driver is frequently used but manual annotation or very simple head position estimation is used [8].

In this paper, an automatic gaze fixation system for the evaluation of IVIS-induced distraction is presented. Driver’s gaze fixation is estimated using a non-intrusive vision-based approach. The system has been tested with professional drivers in a naturalistic simulator running tests directed by a team of psychologists.

We show results of the performance of our system and consistent statistics in order to infer distraction behaviour of the drivers during these exercises.

The remaining of the paper is organised as follows. In section II, a review of the main approaches of the state of the art is presented. Sections III and IV present our 3D face pose estimation and eye direction estimation proposals. Evaluation of our gaze fixation system performance and an analysis of the statistics proposed to study distractions are addressed in Section V. This paper closes with conclusions and future work, presented in Section VI.
II. DRIVER DISTRACTION MONITORING APPROACHES

In the literature there are three main groups of works, according to the measurements they used to detect distractions: biological signals, driving signals and driver images.

Biological signals include electroencephalogram (EEG), electrocardiogram (ECG), etc. These signals are collected through electrodes in contact with the skin of the human body and consequently they are intrusive systems [9]. Only few works, focusing in cognitive distractions, have been reported in the literature using this approach [10]. Most of them have only been tested in operational environments.

Vehicle signals reflect driver’s action, so driver’s state can be characterised in an indirect way. Force on pedals, vehicle velocity changes, steering wheel motion, lateral position or lane changes are normally used in this category. The advantage of these approaches is that signal acquisition is easier. This is the reason why the few commercial systems existing nowadays use this technique [11], [12]. However, they are subject to several limitations such as vehicle type, driver experience or road geometric characteristics. Moreover, some results showed that the accuracy using this approach varied between individuals [13].

Approaches based on image processing are effective because the occurrence of distraction is reflected on the driver’s face appearance and head and eyes activity. Different kinds of cameras and analysis algorithms have been employed in this approach. We group them according to the cameras they adopted, including visible spectrum monochrome cameras, IR cameras or stereo cameras.

A. Methods based on visible spectrum camera

The simplest and most affordable hardware setup is a visible spectrum image acquisition system, at the cost of requiring more complicated algorithms to compensate for the lack of data when compared with infrared or stereo systems. Rongben et al. [14] used skin colour to segment the face. This method needs initialisation and is not robust to lighting conditions or user race. Sun et al. [15] detected the face using adaptive boosting, locating the eyes using a template matching method, and estimated gaze combining Hough transform and gradient direction. Eye activities do not only contain fatigue information but also distraction information.

A commercial eyetracker is also employed in some research works. Blaschke et al. [16] use an off-the-shelf eye-tracker to get head pose and eye gaze signal. They model the visual distraction level as a time dependency of the visual focus, with the assumption that visual distraction increases with time as the driver looks away from the road, but decreases nearly instantaneously when the driver refocuses on the road. Based on the pose and eye signals they propose a two-stage detection method: first, the instantaneous distraction level is estimated, and then a classifier determines if the current level corresponds to a distracted driver.

B. Methods based on IR camera

Many researchers have adopted image acquisition systems based on infrared illumination (IR). The use of IR serves three purposes: it minimises the impact of different ambient lighting conditions, it produces the bright pupil effect and it increases illumination without disturbing the driver. Because of bright pupil effect the eye can be detected more easily, eliminating the face segmentation and reducing computation times.

Cudalbu et al. [17] use a headband with IR reflective markers to estimate the head pose, from which they get a 6 DOF head pose with the average error of 0.2 degrees. Together with this headband, they use a simplified 3D eyeball model to estimate the gaze orientation with an accuracy between 1 and 3 degrees. Jiao and He [18] propose a Round Template Two Values Matching algorithm to locate the bright pupil, which obtains an accuracy of 96.4% but takes 1.01 seconds per frame on a PIII 800MHz computer.

Some commercial products measuring driver’s states are already available on the market, such as SmartEye AntiSleep [9] and SeeingMachines DSS [19], [10]. Both of them use two IR illuminators to enhance their robustness to lighting conditions, and employ only one camera to give 3D information. However, they are focusing on detecting fatigue, not distraction, and they are still limited to some well controlled environments.

C. Methods Based on Stereo Camera

Stereo cameras are also employed to estimate driver state. In [20] two standard web-cameras are employed to make a 3D image acquisition system. They extract face from the disparity map on the assumption that the driver face has smaller depth than background. After face region is extracted they perform embedded HMM to recognise the forehead, eyes, nose, mouth and chin, from which the driving fatigue level can be estimated. The commercial products based on 3D camera technology such as Smart Eye Pro [9] and Seeing Machines faceLAB [21], can provide measurements of head pose, eyebrow, eye, nose, and mouth.

Commercial products are still limited to some well controlled environments, and more importantly no technical information about their algorithms and effectiveness has been published.

D. Contributions

Most of the works in the state of the art were designed for visual distractions, but only a few for cognitive, auditory or biomechanical distraction. In this paper, we introduce a novel gaze fixation system which could potentially be used to extrapolate distractions from any kind, provided all the information from the driving exercises and simulator. Our approach comprises a face pose estimation system and a gaze classification system. Both are non-intrusive and designed to work in very low illumination conditions. The face pose estimation works in the full yaw rotation range. In contrast with existing commercial solutions, it does not require any subject-specific calibration, and could then be used in commercial vehicles. With only two cameras, the system is robust to fast and wide head rotations. The gaze is classified in several areas in the cabin and the road, which may draw the attention of the driver when a distraction is taking place. We use this system in a naturalistic simulator to study changes in driver behaviour.
due to distractions caused by IVIS. Our system is able to generate objective distraction statistics automatically. It does not require user self-reports, which are subjective, or additional input from experts. The experimental setup was designed by a team of psychologists, and it is described in detail in this paper.

III. 3D FACE POSE ESTIMATION

In this section we describe the main characteristics of our face pose estimation method. A detailed description of the face pose estimation can be found in [22]. A diagram of the implemented face pose estimation system is shown on Figure 1.

A requirement is that the pose estimation must be user independent. The proposed face pose estimation system is based on tracking a set of features which are automatically detected on the subject’s face with a calibrated stereo rig.

The features are selected from high-contrast regions of the face using a Harris interest point detector [23]. Taking both camera views, the features are arranged in the form of a sparse 3D face model. This user-specific 3D model provides a prior to estimate the eye parameters.

The appearance of the features in the face changes when the head rotates, and our method proposes a feature template registering using a novel mixed-view technique: the samples of the features taken by both cameras are used to build a joint appearance model, so that views of the face from one camera can be used to anticipate what the other camera will see as the face rotates. This strategy was inspired by the work of Nuevo [24]. From the detected 2D position of the features, the pose is estimated using Levenberg-Marquardt [25].

A model extension process adds new features to the model when the face rotates exposing new areas to the camera. This model extension and the shared template registering allow for a face pose estimation over the full yaw rotation range, from -90° to +90°. The online model creation process is subject to errors. We correct the 3D feature coordinates of the model with a bundle adjustment optimisation [26], which is executed at certain key frames after initialisation and after the addition of new features.

IV. EYE DIRECTION ESTIMATION

To obtain the gaze fixation areas of the driver, we need to take into account both the face pose and the eye direction. Gaze estimation has been applied in HCI and in studies of cognition, which also have attracted interests from marketing research. A survey of the techniques can be found in Hansen and Ji [27]. A common step of all these techniques is the need of person-specific calibration, and recent developments focus on reducing the length of this step. Guerstein and Eizenman [28] presented a system based on calibrated light sources that achieved accuracies of 1°, requiring the subjects to look at a single point. Saliency maps of the images shown to the users have been used as a prior to estimate the eye parameters. Sugano et al. [29] reported errors of 6°, and more recently Chen and Qi [30] presented a probabilistic approach with errors below 2°.

The approach we take is substantially different from these methods. We aim to classify the gaze fixation point in a discrete number of areas of interest, and thus we are only interested in a rough estimation of the gaze. Compared to the works mentioned above, our setup is much harder as the head can move freely and quickly in a wide range and the illumination is low. Also, we can not do any offline user calibration and we can not precompute the saliency maps on the scene. Our only prior is the position of the different areas, which is much less informative than a saliency map because it is not based on human cognition.

We calculate the eye direction, $\vec{e}$, with respect to the face model coordinate system. Consequently, the gaze $G$ is computed as the composition of face pose and eye direction, as we can see in figure 2. It can be expressed as a vector $\vec{g}$ and an origin $T_g$ as

$$G = \{T_g, \vec{g}\}, \quad T_g = \vec{T} + R \cdot \vec{e}_{off}, \quad \vec{g} = R \cdot \vec{e}, \quad (1)$$

where $P = \{\vec{T}, R\}$ is the face pose (translation to the centre of the face and rotation), and $\vec{e}_{off}$ is the distance between the centre of the eyes and the centre of the face model, $\vec{T}$.

At the gaze estimation step, face pose is already known because it has been calculated during the previous stage, so $P$ is given by the face pose translation vector. If head rotation is wide enough, eyes might not be visible, so the gaze is calculated using only the known face pose.

A. Algorithm description

The steps of the coarse algorithm for gaze estimation are shown in figure 3. It consists of the following steps:

1) Automatic 3D face model creation: 3D face model is first created, as explained in section III.

Fig. 1: Architecture of the face pose estimation algorithm.
Fig. 2: Difference between gaze and face pose

1. Automatic 3D face model creation
2. Eye features detection
3. 3D face pose estimation
4. Eye tracking
5. Pupil centre localisation
6. Eye direction calculation
7. Gaze estimation
8. Fixation classification

Fig. 3: Main blocks of the gaze estimation algorithm.

2) Initial eye features detection: At the model creation stage, some characteristic features around eyes are detected within the face using the Stacked Timmed Active Shape Models (STASM) algorithm [31] and added to the face model. Figure 4 shows these features.

3) 3D face pose estimation: In a loop, face pose is estimated from frame to frame as described in Section III.

4) Eyes tracking: The eye features are tracked frame to frame using the same technique as for the rest of the model points. This gives the approximate position of the pupils.

5) Pupil centre localisation: Pupil position is located for each eye using the integral projections algorithm and a Gaussian approximation, obtaining an approximated localisation error of 20% of the pupil size. See figure 5.

6) Eye direction calculation: The eye direction is calculated from the relative displacement of the pupils centre with respect to its original 3D coordinates.

7) Gaze estimation: Gaze is computed rectifying the face pose estimation with the eye direction estimation, according to Equation (1).

8) Fixation classification: Fixation is calculated as the intersection point between the gaze vector and the simulator scene. The fixation point is classified among a set of interest areas where the subject can be looking at.

Figure 6 shows an example of the face pose and gaze estimation monitoring application. At the bottom of the image is shown the fixation point on the scene.

B. Gaze fixation and classification

The objective of this classification is to determine the fixation area of the driver, to know where of a set of key areas she/he can be looking at. The 11 different fixation areas defined for this project are shown on figure 7, and are:

1) Front: The road itself and traffic ahead. Victor and Joanne, basing on experiments on various simulators and on
V. DISTRACTION ANALYSIS USING GAZE ESTIMATION

This section presents the tests and results of our non-intrusive approach to monitoring driver’s distraction. Fixation in the scene is calculated in order to infer driver’s distraction state. Different distraction tasks or activities were inferred in a naturalistic simulator and a study of the incidence of these distracting tasks in the driver’s behaviour was carried out.

Professional drivers were invited to drive the truck through a few scenarios carefully designed by a team of psychologists from the Safety and Human Factors Investigation and Training Centre (ESM) [33], who later examined the generated data to extrapolate behaviour. The scenarios were designed and prepared to require a high level of attention from the driver, and some tasks were intentionally programmed during the driving activity to stress the driver in order to study his/her behaviour under such conditions. The simulator cabin was fully equipped with a variety of IVIS. Experiment layout, driver behaviour studies’ results and conclusions are presented.

A. Experimental environment

Different aspects must be considered in the experimental environment: the physical simulator layout, the camera vision system for gaze estimation, the experiments setup and the subjects.

1) Naturalistic driving simulator: The experiments were performed in the research facilities at CEIT [34], (San Sebastián, Spain), in a room with controlled light and sound environment. The given naturalistic simulator TUTOR [35], shown in figure 8a, consists of a real truck cabin, motorised to simulate movement and equipped with common IVIS. The cabin is assembled on a movement platform with 6 degrees of freedom on which drivers can feel the vehicle accelerating, braking, its centrifugal force, etc. The devices send information to the host, located at the Instructor Position (PI), where the psychologists can control the whole simulator, analyse all the data and reproduce stored simulations. Main computers are placed in the PI, located near the cabin. A dedicated computer processes face pose and gaze estimation using the algorithms presented in this paper, and sends this information to the PI, where the psychologists can access to the data.

The visualisation system is made of three back-projection panels with a total surface of 22 m². The fact that the screens have no marked separation plus the geometry of the image system makes for a flawless overall impression. Moreover, two computer screens are used as rear mirrors, attached to both sides of the cabin.

The cabin is fully equipped, and contains a GPS, a hands-free, the on-board computer and a tachograph. These are some of the key locations that the gaze estimator must differentiate.

2) Vision system: The stereo rig is located inside the cabin, over the dashboard, between the windscreen and the driving wheel, facing the driver at a distance between 70 and 100 cm. The baseline is 23 cm, and cameras are slightly turned inward, to better point to the driver’s face, closer to the rig. Figures 8 b) and c) show the camera layout inside the naturalistic simulator cabin. The cameras [36] are 1392 × 1040 pixels, and capture images at of 30 fps, with a Firewire 800b interface. The custom-built rig and IR illumination board are attached to the windscreen using suction pads, and rests on foam over the dashboard to reduce vibration. The hardware layout can vary without affecting system operation, as long as the stereo rig is calibrated and referenced to the world coordinates, fixation area locations are approximately known, and the cameras are
ahead of the driver, not necessarily in front. Our cameras mount a 9 mm lenses, and capture is synchronised with the IR pulsed illumination. The capture system is coded using C++ and GNU libraries and tools, running on a Core 2 Duo processor commanded by Ubuntu. This vision system is an evolution of ther monocular one developed by the group for drowsiness detection [37].

3) Experiments setup: To design the experimental protocol, the team of psychologists built on the following initial hypothesis: “The potential driver distraction due to IVIS is determined by the level of attentional demand required by them while driving, decreasing the effectiveness of the primary task: driving.”

By analysing the professional drivers behaviour, the basic and most representative features in the context of this activity are identified [38]. Some scenarios, types of vehicles, incidents, on-board systems utilisation and critical situations are selected to infer distraction in drivers. Thus, the professional drivers behaviour should be generically represented. Taking into consideration this basis, which involves observing and information recording during the driving activity, the next step was to define the simulation exercises.

Experiments were designed with the goal of refuting the initial hypothesis of the research regarding the potential distraction of four different on-board systems which are commonly used in professional driving. These devices were digital tachograph, GPS, hands-free and on-board computer. Under these conditions, four scenarios were created: mountain, inter-city, urban and long-distance. Different exercises setup were prepared for each scenario, each containing different tasks, events, weather conditions and IVIS requirements.

According to Victor et al. [32], three different tasks are of special importance to study distraction: visual tasks, auditory tasks, and cognitive tasks. During the experiments, visual tasks require to use the GPS. Auditory ones involve making a call to the hands-free telephone, and engaging in a trivial conversation. For the last one, a cognitive task is enforced in one of the exercises by making a phone call in which the driver is asked to describe the route from one point to another on a city he knew. During the exercises, events are inserted proximal to tasks, such as motor, tires or ABS breakdown, sudden braking of the precedent vehicle, broken down vehicles on the road, vehicles running a red light, etc. A summary of the different exercises setup is shown in table I.

These tests were implemented using 16 different exercises: five based on the inter-city scenario, four on the mountain scenario, three on the urban and the last four on the long-distance one. The defined procedure to evaluate these exercises consists on different drivers driving through different scenarios.

The first exercise of each scenario was the control exercise which corresponds to the exercise undertaken by each driver without external perturbations. This exercise provides a reference for the other exercises, in terms of driver gaze fixation patterns on the controls, to which the results from distractions can be compared.

4) Subjects: 12 professional drivers from different gender, age and experience were selected, having different participants for each test configuration to detect the user dependent behaviour variables. Each driver drove in two exercises on each scenario (the control exercise plus another), and duration varies from 20 to 45 minutes, depending on the scenario.

Previous studies with similar conditions used groups from 7 to 30 participants [39], [40]. All subjects were informed of the purpose of the experiment and the security procedures in the simulator facilities.

B. Ground-truth generation

To assess the performance of our system we obtained the ground-truth (GT) for the face pose and gaze. We obtained the pose ground-truth with a calibration chessboard, attached to a helmet the subjects wore during the experiments. The pose of the chessboard was obtained using camera calibration techniques, with an average error below 1°. The ground-truth data was obtained for 6 users on sequences over 10 minutes long.

With respect to gaze, we used a video database of 15 videos more than 10 minutes long each. We have obtained the ground-truth for the gaze fixation classification by manually labelling the driver’s fixation area on each moment.

C. Face pose experimental results

Firstly we analyse the general performance of our face pose estimation system. Mean pose estimation error in the three
angle rotations are shown in Table II. Mean error is computed by rotation ranges for each direction. Pitch and roll rotations present a smaller output range than for yaw since there are no wider head rotations for these directions on the driving exercises. Because the registering technique can not be applied to pitch variations, the system error is higher in this direction, and it can be observed how it increases for angles $\alpha_{pitch} > 30^\circ$. Still, the bundle adjustment (BA) slightly improves the results. Evaluation of the error in a wider pitch and roll range is not possible because significant rotations are not natural while driving.

<table>
<thead>
<tr>
<th>Rotation</th>
<th>BA?</th>
<th>$\alpha &lt; 15^\circ$</th>
<th>$\alpha &lt; 30^\circ$</th>
<th>$\alpha &lt; 45^\circ$</th>
<th>$\alpha &lt; 45^\circ$</th>
</tr>
</thead>
<tbody>
<tr>
<td>yaw</td>
<td>no</td>
<td>1.92</td>
<td>2.44</td>
<td>6.72</td>
<td>12.83</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>0.98</td>
<td>1.54</td>
<td>3.04</td>
<td>8.54</td>
</tr>
<tr>
<td>pitch</td>
<td>no</td>
<td>3.82</td>
<td>7.86</td>
<td>8.59</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>1.81</td>
<td>4.70</td>
<td>6.34</td>
<td>-</td>
</tr>
<tr>
<td>roll</td>
<td>no</td>
<td>1.27</td>
<td>2.06</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>yes</td>
<td>1.16</td>
<td>1.75</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

TABLE II: Mean face pose estimation error. The error is divided into $yaw$, $pitch$ and $roll$, and evaluated in different ranges of the absolute rotation angle in the ground truth, $\alpha$.

As we can see, our proposal has a very low error thanks to the BA corrections. The error remains low for the full range $\pm 90^\circ$ of yaw rotations. These results show equal or lower errors than other important works in the literature [41], [42], [43], despite we tested or system with more challenging scenarios. The low-lighting conditions, along with the fast head movements make the face appear very blurry when the head is moving side to side. Any other tracking system would lose track of the head under these conditions. On the other hand, we can keep tracking the face thanks to the extended face model and the mixed-view technique.

The tests used in the different works presented in Table II differ, but a qualitative comparison among them is possible taking into account the performance results published for each work.

D. Gaze fixation experimental results

Table III presents the gaze fixation classification error. The error has been calculated as the relation between the incorrect classifications, and the total number of frames (from the GT) that the driver has gaze fixation to an specific area. A False negative is when we do not detect that the driver is looking at an specific fixation area. A False Positive happens when the classifier says that the driver is looking at fixation area, but actually is looking at another area. For reference, we divide the false positives into the different areas where the driver was really looking at. As it can be observed in the table, the classification errors for the Front fixation area, rear mirrors and on-board computer are below 5%. This are the areas which give more information on the distraction state of the driver. The left and right signals are the most difficult to detect, since they are very proximal to the front fixation area.

E. Statistics results to infer distractions

In this section we provide some consistent statistics, based on gaze fixations, useful to assess driver distraction. The
driver’s behaviour interpretation must be carried out by psychologists and is out of the scope of this paper.

As explained in Section V-A the drivers were asked to perform demanding tasks while driving such as answering the phone and following indications, receive corrections of previous indications to get to a place or to operate the on-board computer or the tachograph. In Table IV the recorded incidents while using the different IVIS are shown.

<table>
<thead>
<tr>
<th>Fixation area</th>
<th>False Negative</th>
<th>False Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(not detected)</td>
<td>(detected while looking to:)</td>
</tr>
<tr>
<td>Front</td>
<td>4.04%</td>
<td>0.82%</td>
</tr>
<tr>
<td>Left mirror</td>
<td>2.79%</td>
<td>-</td>
</tr>
<tr>
<td>Right mirror</td>
<td>3.22%</td>
<td>0.21%</td>
</tr>
<tr>
<td>Computer</td>
<td>4.96%</td>
<td>0.34%</td>
</tr>
<tr>
<td>GPS</td>
<td>11.16%</td>
<td>1.20%</td>
</tr>
<tr>
<td>Hands-free</td>
<td>10.78%</td>
<td>1.70%</td>
</tr>
<tr>
<td>Tachograph</td>
<td>5.45%</td>
<td>0.15%</td>
</tr>
<tr>
<td>Left and Right signals</td>
<td>13.95%</td>
<td>1.30%</td>
</tr>
</tbody>
</table>

TABLE III: Gaze classification error percentage

Most of the accidents occurred while receiving or executing the indications from the GPS or the hands-free device. As expected, the more cognitive demanding a task is the higher the risk of being involved in a distraction related incident. Results on table IV also show that using the tachograph and on-board computer is not very cognitive demanding since it only requires quick looks, similar to a mirror checking. The challenge now is to study driver reactions to different IVIS and find out the optimal IVIS location and the optimal way to deliver indications and warnings to them. Hereafter, we analyse two concrete cases for the most critical IVIS. Using our system, statistics are automatically generated while many other works in the state of the art are based on user self-report, which is subjective.

1) **Hands-free device**: Most incidents were registered while the driver was using the hands-free device. This, together with the fact that 85% of people admitted using a cell phone while driving at least occasionally makes phone use one of the most dangerous activities while driving. The recorded consequences of using the hands-free device are: an increase in the corrections of the in-lane position, short braking, non constant speed and fixed gaze.

Figure 9a shows gaze focalization of the driver for a piece of the exercise D2 starting at the moment when a phone call is received. Continuous looks to the hands-free device (H) and no checking to the vehicle speed (C) or rear-view mirrors (M) are detected while answering the phone call. On the other hand, on the control exercise (D1) for the same piece, depicted in Figure 9b, the driver remains attentive to the road (F) checking the vehicle speed and the rear-view mirrors from time to time.

2) **GPS**: The second higher cause of incidents was the use of the GPS device. In our experiment, incidents were caused while setting up the GPS device and while following the indications. When setting up the GPS the loss of attention to the road was the main cause of incidents. During these losses of attention the gaze was fixed on the GPS device for long periods of time. However, the indicators received from the GPS also caused an increment in the incident rate, which is mainly due to a decrease in the attention paid to the road information (traffic signs and overhead panels) and a loss of attention to the road with frequent looks to the GPS device.

Figure 10 depicts gaze focalization of the driver for a piece of the exercise B2 versus the control exercise for the same piece. In this case the driver follows guidance GPS instructions. The lack of attention paid to the road information while using the GPS, makes the driver to start an overtaking manoeuvre while approximating the exit he has been asked to take. The driver is overtaking a very slow moving vehicle when GPS informs him of taking the next exit. To accomplish this instruction, a two-lane changing is required in less than 4 seconds. Instead, when driving without external perturbations, the driver checks the traffic signs to reach his target. The driver is aware that he has to take the next exit in advance and decides to wait behind the slow moving vehicle.

As figure 10a depicts, reiterative checks of the rear-view mirrors (M) and the vehicle speed (C) are performed in the
To help psychologists understand the distraction pattern of a driver, two more measurements can be taken, apart from reaction time. Traditionally, a glance based measure has been used. This technique measures duration of individual fixations on different zones, frequency, number of glances or total task duration. However, these measurements heavily depend on the task, the driver experience, and other factors. In [32] authors found that the Percent Road Centre (PRC) measurement is more stable across users and different experiments. PRC measures how much time is spent looking at the road centre area while performing a task. This zone includes the road, signalling and visual elements proximal to the road. We have analysed this parameter in our experiments.

Figure 11 shows the fixation time percents in the different zones of the scene during the execution of a task. This figure was generated after the analysis of the results obtained with the gaze estimation algorithm and tasks schedule during the exercises. The first column shows the average percents for the control experiments (exercises A1, B1, C1 and D1). The second column depicts distractions inferred using the GPS, obtained from exercises A2, A3, B2, B4 and C2. The third column depicts distractions inferred by tasks requiring talking by the hands-free phone, on exercises A4, A5, B3, D2 and D4. The last one represents a cognitive task on exercise D3, which induced distraction with a phone call to explain a route.

We have found PRC to be a good parameter to assess driver distractions. In contrast with other parameters used in the state of the art works, this is automatically calculated. Moreover, we also found that the distraction pattern inferred for the different IVIS is different. Unlike most of the works in the state of the art, we analyse visual, cognitive and auditory distractions. While the GPS shows an important reduction of PRC, phone calls do not reduce PRC, rather instead, it slightly increases. On cognitive tasks, we could not infer any important variation of this parameter. However, the time used looking at the signalling and as well as the road proximities is reduced for all tasks. This behaviour is clearly observable on figure 11, and it is in line with the conclusions presented by [32].

It can be observed how the time that the driver spends looking at the mirrors, signals and on-board computer is drastically reduced for any of the task, in comparison with the control exercises. During phone calls, driver increases the time in which he is looking at the front, but reduces the fixations on the mirrors and signals. This could mean that the driver is not actually paying attention to the road, although more work would be needed to extract a precise conclusion.

VI. CONCLUSIONS AND FUTURE WORKS

We have applied a gaze fixation technique, based on face pose and gaze estimation algorithms, to monitor the distraction state of a driver in a naturalistic driving simulator. Our face pose estimation approach is based on a sparse 3D face model obtained from face features. A re-registering algorithm, online model extension and bundle adjustment process for error corrections allow a face tracking for the full yaw rotation range, $\pm 90^{\circ}$. The system has been tested under challenging conditions, such as low illumination, variety of drivers, heavy
TABLE V: Driver behaviour and reaction time statistics

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Exercise</th>
<th>Overpass speed limit [mph]</th>
<th>Reaction time to an event [seconds]</th>
<th>Obstacle‡</th>
<th>Mechanical fault‡</th>
<th>Answer a call‡</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>max min</td>
<td>min max</td>
<td>min max</td>
<td>min max</td>
</tr>
<tr>
<td>Inter-city</td>
<td>A1</td>
<td>0</td>
<td>25 5</td>
<td>(1) 5</td>
<td>2</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>A[2-5]</td>
<td>2</td>
<td>15 3</td>
<td>32 81</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Mountain</td>
<td>B1</td>
<td>2</td>
<td>19 5</td>
<td>(2) 0.6</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>B[2-4]</td>
<td>6</td>
<td>13 2</td>
<td>24 11s</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>Urban</td>
<td>C1</td>
<td>0</td>
<td>16 3</td>
<td>(3) 4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>C[2-3]</td>
<td>0</td>
<td>4 0</td>
<td>4 43</td>
<td>4</td>
<td>miss</td>
</tr>
<tr>
<td>Long-Distance</td>
<td>D1</td>
<td>1</td>
<td>34 12</td>
<td>(4) 4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>D[2-4]</td>
<td>4</td>
<td>20 0</td>
<td></td>
<td>-</td>
<td>3</td>
</tr>
</tbody>
</table>

* Reaction time after the event. ‡ Reaction time before the event. Moment at which the subject is aware of the obstacle and takes an action before colliding. (1) Mechanical fault in the air filter. Warning marked on the on-board computer. (2) Tyre blowout. Marked through audible sound and truck dynamic model changes. (3) ABS fault. Warning marked on the on-board computer. (4) No mechanical faults scheduled.

IVIS utilisation and fast and wide head movements. Under this conditions the system has proved good results. It shows a mean error below 1° for rotations in the ±15° range, and 1.54° in the ±30° range, improving the results of other works in the literature.

An eye direction estimation method has been added to the face pose to generate the gaze estimation. The eye direction is based on the pupils displacement with respect to their original positions. Gaze fixation information can be used to help psychologists to assess the distraction state of the driver. Performance of our gaze fixation system has been analysed showing good figures for this application.

Different driving exercises directed by a team of psychologists were recorded in a naturalistic driving simulator, through 12 professional drivers, generating more than 15 hours of driving. Four scenes were created, and distractions induced by different IVIS were inferred and compared to the control exercises without external perturbations. This way, it was possible to compare the driver behaviour under non-distracted driving conditions versus distracting ones. This has been a challenge dataset to evaluate our system.

Reaction times and gaze focalisation behaviour patterns have been measured to draw conclusions about the capacity of our monitoring system to help psychologists to study driver’s reactions under different situations while using IVIS. At the time of writing, psychologists are finishing the interpretation of data provided by our system in order to assess driver distractions over the explained tests.

Further studies about the optimal location of the different IVIS and the way the information is delivered are being designed by psychologists to reduce the distraction. New tests with these modified IVIS will be performed on the simulator to evaluate the improvements of the new designs.

VII. ACKNOWLEDGMENTS

This work has been financed with funds from the Ministerio de Ciencia e Innovación through the projects: DRIVER-ALERT (TRA2008 - 03600), ADD-Gaze (TRA2011-29001-
CABINTEC (PSE-370000-2009-12), and with funds from the Comunidad de Madrid under project Robocity2030 (S2009/DPI-1559). Authors want to thank the drivers that participated in the experiments as well as psychologists form ESM and staff from CEIT for their contributions to this paper.

REFERENCES


Pedro Jiménez received the MS degree in Telecommunications Engineering from the University of Alcalá, Alcalá de Henares (Madrid) in 2005, and he received the PhD degree in March 2011 from the University of Alcalá. He started as a researcher at the University of Alcalá in 2004, and since then has tackled problems in Electronics, FPGAs and computer vision. In the last field, he has worked in automotive and face-tracking applications. Nowadays he is working in Spain for Gamesa in the field of Renewable energies.

Luis M. Bergasa received the M.S. degree in Electrical Engineering in 1995 from the Technical University of Madrid and the PhD degree in Electrical Engineering in 1999 from the University of Alcalá, Spain. He is currently a Full Professor at the Department of Electronics of the University of Alcalá. He has been the Head of the Department of Electronics in the period 2004-2010. He is the Coordinator of the RobeSafe Research Group from 2010. His research interests include real-time computer vision and its applications, particularly in the field of the assistant robotics and the intelligent transportation systems. He is the author of more than 150 refereed papers in journals and international conferences, and corresponding author of 6 national patents and 1 PCT patent. He is been recipient of 8 Spanish prizes related to Robotics and Automotive fields from 2004-2011. He is associate editor of the Physical Agents Journal and habitual reviewer in 12 journals included in the JCR index. He has served on Program/Organizing Committees in more than 15 conferences. He is IEEE member of the Robotics and Automation Society Technical Committee on Autonomous Ground Vehicles and Intelligent Transportation Systems. Since 2009, He is Research Director of Vision Safety Technologies Ltd, a spin-off company established to commercialize computer vision systems for road infrastructure inspection.

Jesus Nuevo obtained his PhD from the University of Alcalá (Madrid, Spain) in 2009. He has worked on driver monitoring systems using computer vision since 2004. He is now a OCE Post-Doctoral fellow at CSIRO ICT Centre Computer Vision group, in Brisbane, Australia. His interest are in theoretical and applied Computer Vision, specially in face tracking and Non-Rigid Structure from Motion.

Iván G. Daza received the M.S. degree in telecommunications engineering from the University of Alcalá (UAH), Madrid, Spain, in 2004. He received the PhD in November 2011, from University of Alcalá. He worked in drowsy drivers detection system. His research interests include computer vision, pattern recognition, machine learning, stochastic processes optimisation and control theory.

Noelia Hernández was born in Valladolid, Spain in 1984. She received the Bachelor of Computer Science in 2007 and the Masters degree in Advanced Electronic Systems (Intelligent Systems) in 2009, both from the University of Alcalá. She is currently working towards her PhD at the Electronics Department of the University of Alcalá. Her research interests include fuzzy control, WiFi localization systems, robotics, image processing and intelligent vehicles.