Towards low-cost systems for measuring visual cues of driver fatigue and inattention in automotive applications

Lars Bretzner and **Martin Krantz**
Smart Eye AB
Skräddaregatan 1
Göteborg, Sweden
bretzner@smarteye.se, mkz@smarteye.se

Abstract- Recent studies show that driver fatigue and inattention are major causes of fatal road accidents. A large number of these accidents could be avoided if the vehicles were equipped with a) sensors that reliably could monitor the attention and alertness cues of the drivers, and b) systems that, based on the sensor input, could warn the drivers in time. Today, camera-based systems can measure such attention cues, and are tested in automotive in-cabin installations with two or more cameras. Camera-based solutions are non-intrusive, they can measure head position and orientation as well as gaze direction and eyelid closure, and they can be used to identify the driver for other tasks. However until now, no camera-based system has met the requirements from the car manufacturers on compactness and cost. Here we present the principles behind Smart Eye AntiSleep, which is a compact one-camera system specially designed for automotive in-cabin applications. AntiSleep measures 3D head position, head orientation, gaze direction and eyelid closures. AntiSleep uses one camera and two IR-flashes. With its own illumination, AntiSleep is highly robust to all natural illumination conditions and can efficiently handle disturbing reflexes in eyeglasses. The system is fully automatic and requires no manual intervention during initialisation.

I. Introduction

TOGETHER with alcohol and speeding, driver fatigue and inattention are recognized to be the most causative factors of severe road accidents [1,2]. Large research and development efforts are spent on finding efficient ways to monitor driver alertness in road vehicles; trucks, buses and automobiles. There are many cues to driver alertness, ranging from heartbeat frequency to steering wheel motions and lane keeping, but far from all can be measured efficiently in a realistic automotive scenario. Three of the major cues are eyelid closure, gaze direction and head motion. These cues can be measured non-intrusively using recent video camera-based systems. Video camera-based systems offer a number of advantages in comparison to other techniques; in addition to measuring all the abovementioned cues, the systems can be used to identify the driver for different purposes, such as recalling personalized settings, preventing unauthorized drivers from starting the car etc. Current systems on the market use stereovision and require two or more cameras placed at a distance from each other. These multi-camera systems were originally mainly designed for laboratory settings, and they are less suited for large-scale in-cabin mounting with high costs of space and wiring. An automotive system as previously discussed must have small dimensions and be inexpensive in large volumes. This is why dual camera stereo systems are not likely to make it into serial production. It must also work on virtually all drivers even with glasses or sunglasses, without requiring any interaction from the driver. The accuracy of the gaze measurement should be such that at least four distinct gaze directions can be distinguished without prior calibration of the system (on road/main instrument/ left/right). Some algorithms for drowsiness detection use the speed of which the eye is opened or closed, and then a frame rate of 50 Hz or higher is needed. For such applications, the eyelid opening has to be measured with a resolution of about 1 mm. If the driver's face has been temporarily occluded, the re-finding time should be kept to a minimum once the face is visible to the camera.

While a mono-camera system would be preferable for cost reasons, it is a technical challenge to construct such a system that can deliver the driver alertness cues with sufficient accuracy and reliability. We will here present the principles behind the Smart Eye AntiSleep system, which we argue is major step towards a low-cost system for measuring visual cues of driver fatigue and inattention in automotive applications.

II. Technical Principles

A. Hardware: Camera and illumination

The AntiSleep system uses one standard video camera for image input. The functionality is based on computer vision analysis of the incoming video images. The current resolution of the b/w camera is VGA (640x480) and the frame rate is 60Hz. Details on the hardware set-up as well as performance figures can be found in [3]. The choice of hardware has been made as a step to meet the abovementioned constraints on cost

0-7803-9435-6/05/$20.00 ©2005 IEEE.

and size of each unit. The camera is chosen to have the same performance as very small, low-price cameras while it provides IEEE 1394 (FireWire) connection to any PC and laptop to facilitate in-cabin tests and close integration of fatigue/inattention analysis software. Figure 1 shows the camera unit, which can be mounted in e.g. the dashboard, the top of center stack or at the A-pillar.

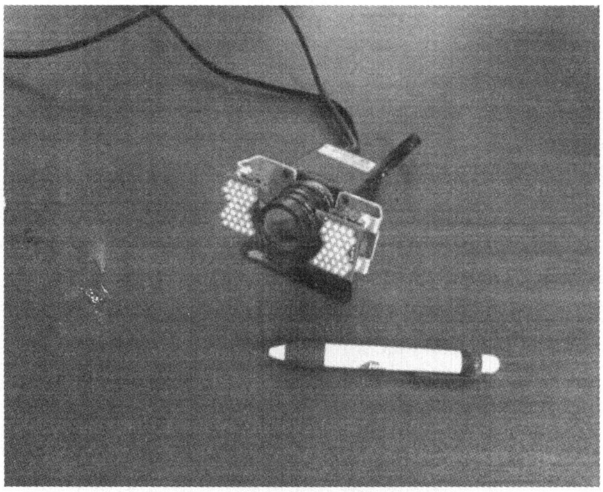

Fig. 1. The AntiSleep mono video camera unit with two IR-illuminators. The unit can be mounted e.g. in the dashboard or at the A-pillar.

The large variety of possible natural illumination conditions in cars can pose significant problems to video based systems. To minimize the influence of different illumination conditions on the image acquisition process, the AntiSleep system uses two IR-illuminators together with an IR-filter. The illuminators and the filter are tuned to frequencies with minimal interference from outdoor light. The fact that the system mainly uses its own illumination makes it functional in complete darkness, and highly increases the robustness to e.g. flickering sunlight and sharp shadows.

B. Image analysis

The images from the camera system are in large processed according to the system overview shown in figure 2. The system starts to operate in an initialization phase, which gradually transforms into a normal tracking phase, as the system acquires knowledge about the driver.

In the automotive scenario we have well-known constraints on the drivers head position and orientation which we make use of in the initial face detection step. We start with a generic 3D head model including the relative 3D distances between generic facial points such as eye features, nostrils and mouth corners. Projecting the head model into the image gives spatial constraints on the image positions of the generic facial features.
The facial features are detected and localized in a coarse-to-fine manner; first a fast, coarse-scale localization of the face, then localization of individual medium and fine-scale features. The generic facial feature points, e.g. eye features, nostrils and mouth corners, are identified and localized using interest point descriptors, inspired by local Gaussian derivatives [4], SIFT [5] and Gabor-jets [6] at certain scales. These descriptors all capture the local image appearance in a neighborhood of a point. The detection scales of the finer-scale feature descriptors are determined from the detection of the coarser-scale features. The probability distribution of each facial feature point descriptor at an appropriate scale is learned from a large set of facial training images, covering a wide range of facial appearances from subjects of different ages, genders and ethnic groups.

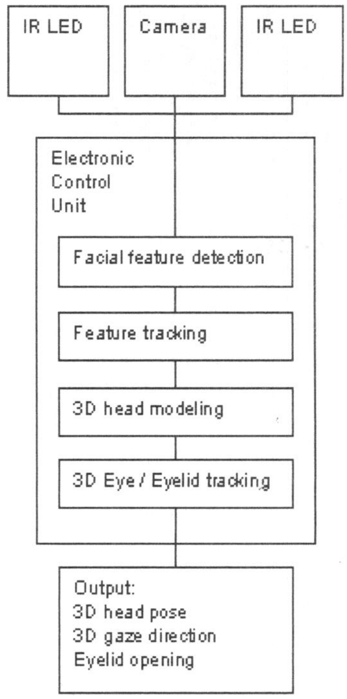

Fig. 2. System overview showing the principal image analysis components.

From the image positions of the facial features and the generic head model, an initial head pose can be estimated.

The detected facial features are tracked as the driver's head moves over time and, using structure-from-motion algorithms [7], their initially generic 3D positions are successively refined. In this way, the initial generic 3D head model is adapted to the current driver and the head pose estimates are improved.
During tracking, the driver-specific appearance of each generic feature is learned for different views. This information is used to stabilize and speed up tracking and recovery from tracking failures. Fast and stable recovery from tracking failures is naturally of vital importance for this kind of measurement system.
In order to detect and add driver-specific facial features, such as e.g. birthmarks, to the head model, detection and tracking of corners and blobs [8] can be performed in the facial area. These facial features candidates can then be accepted or rejected based on rigid-body motion constraints. The ability to incorporate these extra subject-specific features is important as more feature points normally allows for better head pose estimates. The extra features can also be used to handle large head rotations where the initial generic facial features are

occluded.

Based on the estimated 3D positions of the eye features, a 3D model of the eyeball and eyelids is projected into the image. A refined search for the iris is done through a fast ellipse finding algorithm, which uses a modified version of the Hough transform [9]. The possible eyelid positions are matched to edge data to determine the eyelid state. The model of the eyelids makes it possible to estimate the eyelid opening in 3D, see figure 3 and 4.

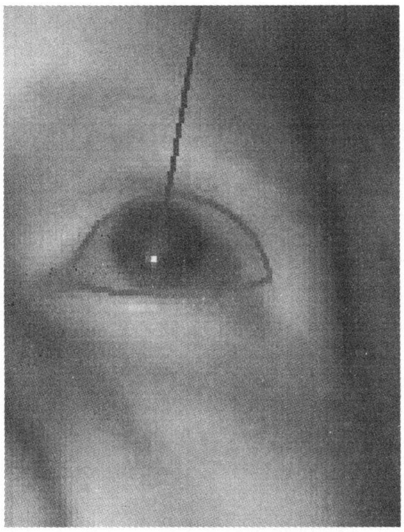

Fig. 3. Zoomed-in detail of the screen output. Parts of the eyelid model are superimposed on the eye region, the red ray is the projection of the estimated gaze direction.

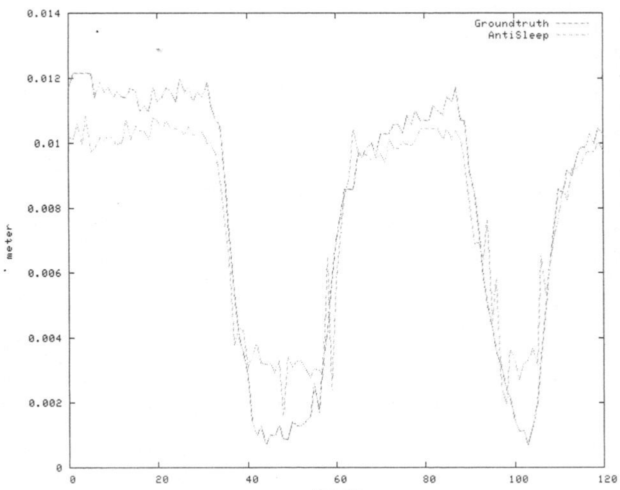

Fig. 4. Comparison of estimated 3D eyelid opening and ground truth from manual annotation.

For drivers wearing eyeglasses, the IR-illumination could in some cases result in unwanted reflections in the glass, which might disturb the eye and eyelid tracking as image information is lost in neighborhoods of the reflections. In these cases, the problem is solved using a (patented) technique where the two IR-illuminators are active in different frames. The position of the IR-reflection in the image will therefore move from frame to frame, and lost image information in one frame can be recovered from the previous frame.

The constraints imposed by a future DSP implementation has been taken into consideration, e.g:
- Small memory footprint, both for program code and for data;
- High memory locality in order to stay within bus bandwidth and cache size limits;
- Integer/fixed-point processing due to limited floating-point capability;
- Only minimal underlying OS support expected.

C. Output

The output of the system is, for each frame, the driver's 3D head position and orientation, the gaze direction in 3D and the eyelid opening of each eye. The system delivers output data at frame rate, i.e. 60 Hz, on a standard, low-end laptop-PC. Estimated quality values are provided for each output based on confidence signals from the involved algorithms. Figure 5 shows an example of the output to the screen. The output data can be visualized in 2D- or 3D-views of a scene model provided by the user. For analysis purposes, gaze zones can be defined in the scene model. The output is prepared for integration with third-party fatigue and inattention analysis tools.

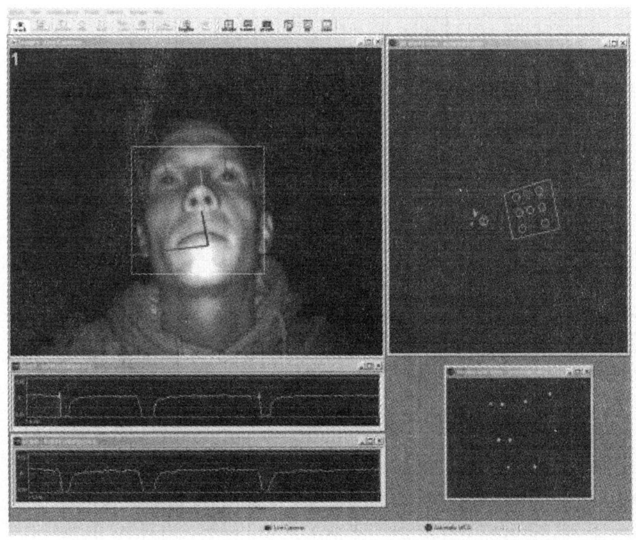

Fig. 5. Screen output showing the tracked head, eyelid open-ing graphs, the 3D head pose in a world model and estimated 3D positions of a subset of facial features.

D. Future work

Without any driver-specific gaze calibration, gaze directions to the four major automotive gaze-zones; on road/main instrument/ left mirror/right mirror, can still be resolved. However, if we want to move beyond this resolution, e.g. to detect gazes to individual instruments, gaze calibration is necessary. As no interaction with the driver can be required, such calibration could be done automatically using statistics of gazes to well-known in-cabin hot spots, gathered during normal driving tasks.

E. Related mono-camera systems

In the last years, there have been quite many research projects studying mono-camera systems for head and/or gaze tracking purposes, we will here briefly mention a few of them. A rather early project [10] relies on skin color and optical flow for head and eye tracking. An interesting and promising approach to simultaneous on-line learning of the driver's appearance and 3D face shape is presented in [11]. In [12], it is studied how gaze tracking can be performed in the same framework. Many of the largest car manufacturers are involved in R&D projects along the directions outlined in this paper. A large Japanese car manufacturer recently announced a mono-camera system, measuring head orientation for driver inattention detection, that is to be offered in Japan in the spring of 2006 [13].

III. SUMMARY AND DISCUSSION

Visual cues of driver fatigue and inattention can nowadays be measured by computer vision based systems. While accurate 3D measurements of head pose and gaze are easier to acquire using multi-camera systems, such systems are not suitable for automotive installations where the cost and size of each unit has to be kept to a minimum. Recent research results from academia and industry show promising results in measuring the visual cues using mono camera systems. We have in this paper outlined the ideas and technical principles behind Smart Eye AntiSleep, which is a mono camera system measuring 3D head pose, gaze direction and eyelid closure, specially designed for automotive installations.

In the mid 1980´s the first automotive airbags were installed in serial production. Since then the use of passive safety devices such as airbags and seat belts has exploded, and car related deaths has decreased at the same time. The new generation of automotive safety systems is active safety devices. Instead of minimizing the damage if there is an accident, they work to prevent the accident from taking place. The first such systems in production have been based on sensors looking out, assessing the situation around the car. AntiSleep is an example of a concept where the driver is in focus. The actual behavior of the driver is measured and can, together with signals from a variety of other sensors, be analyzed in order to correctly assess the probability of an imminent accident. The use of active safety systems promises safer transportation in the near future.

REFERENCES

[1] V. L. Neale, T. A. Dingus, S. G. Klauer, J. Suweeks, M. Goodman, "An overview of the 100-car naturalistic study and findings", Paper Number 05-0400, National Highway Traffic Safety Administrationin, Washington D.C..

[2] S. Nordbakke, "Driver fatigue and falling asleep – experience, knowledge and action among private drivers and professional drivers", Report no 706/2004, Norwegian Public Roads Administration, Oslo

[3] "Smart Eye AntiSleep – monitoring fatigue and inattention cues in automotive applications", Smart Eye AB, 2005, available at http://www.smarteye.se

[4] V. Vogelhuber and C. Schmid, "Face detection based on generic local descriptors and spatial constraints", *Proc. Int. Conf. on Pattern Recognition (ICPR)*, Barcelona, 2000, vol 1, no. 1, pp 5084.

[5] D. G. Lowe, "Object recognition from local scale-invariant features", *Proc Int Conf on Computer Vision (ICCV)*, Corfu, 1999, pp 1150

[6] L. Wiskott, J-M Fellous, N. Kruger, C. von der Malsburg, "Face recognition by elastic bunch graph matching", *IEEE Pattern Analysis and Machine Intelligence*, vol 19 pp 775-779, 1997

[7] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, Cambridge University Press, 2000

[8] L. Bretzner and T. Lindeberg, "Feature tracking with automatic selection of spatial scales", *Computer Vision and Image Understanding*, vol 71(3), pp 385, 1998

[9] P.V.C. Hough, "Machine Analysis of Bubble Chamber Pictures", *Int. Conf. on High Energy Accelerators and Instrumentation*, CERN, 1959

[10] P. Smith, M. Shah, N. da Vitoria Lobo, "Monitoring head/eye motion for driver alertness with one camera", *Proc. Int. Conf. on Pattern Recognition (ICPR)*, 2000

[11] J. Xiao, S.Baker, I. Matthews, T. Kanade, "Real-time combined 2D+3D active appearance models", *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2004

[12] T. Ishikawa, S. Baker, I. Matthews, T. Kanade, "Passive Driver Gaze Tracking with Active Appearance Models", *Proc. World Congress in Intelligent Transortation Systems*, 2004

[13] "Toyota enhances pre-crash safety system with driver-monitoring function", News release, Toyota Motor Corp., Sep, 2005