A miniature head-mounted camera for measuring eye closure

Simon J. Knopp
Electrical & Computer Eng.
University of Canterbury
NZ Brain Research Institute

Philip J. Bones
Electrical & Computer Eng.
University of Canterbury

Stephen J. Weddell
Electrical & Computer Eng.
University of Canterbury

Carrie R. H. Innes
NZ Brain Research Institute

Richard D. Jones
Electrical & Computer Eng.
University of Canterbury
NZ Brain Research Institute

ABSTRACT

This paper describes a miniature camera module for capturing close-up video of one eye and the image processing steps to locate the pupil and measure eye closure from this video. This camera is one component of a multi-sensory device for measuring drowsiness and detecting complete momentary lapses of responsiveness.

We describe a flood-fill-based algorithm for locating the pupil and shape-based criteria for determining whether the pupil is partly covered by the eyelid. Percentage eye closure (PERCLOS) is implemented as an example of a meaningful measurement that can be derived from this extracted pupil data. Preliminary results show that the algorithm produces output very close to that obtained by manual frame-by-frame classification of the eye video.

Categories and Subject Descriptors

I.4.9 [**Image Processing and Computer Vision**]: Applications; I.4.7 [**Image Processing and Computer Vision**]: Feature Measurement—*Size and shape*; I.4.6 [**Image Processing and Computer Vision**]: Segmentation—*Region growing, partitioning*

1. INTRODUCTION

The camera and image processing techniques described in this paper form one subsystem of a multiple-sensor device for detecting drowsiness and "lapses"—complete transient losses of responsiveness [7]. As shown in the block diagram in Fig. 1a, the device will have a camera module to capture video of one eye, several EEG electrodes to measure brain activity, and inertial sensors to measure head movement. Relevant features will be extracted from these data streams and fed into a classification stage to identify lapses and microsleeps and to quantify drowsiness.

1.1 Motivation

Lapses of responsiveness can have severe, or even fatal, consequences for people in a variety of occupations and for those around them. This is especially true for jobs which require maintaining high levels of attention on monotonous tasks for long periods of time. For instance, commercial vehicle drivers, pilots, air-traffic controllers, and some medical professionals risk causing fatalities if their task-oriented attention lapses, even briefly [12]. Creating a device capable of detecting these lapses and intervening quickly has the potential to save lives.

A previous study [5] found that normally rested people had an average of 79 microsleeps per hour with an average duration of 3.3 s when completing a tracking task [10]. Additionally, 8 of the 20 subjects had sleep episodes lasting for more than 15 s. Although the prevalence of such events may be lower outside of the lab, this project aims to reduce the potential for them to cause harm.

1.2 Requirements

This device is ultimately intended to be worn as a piece of safety equipment for a variety of occupations, so it must be as unobtrusive to the wearer as possible. The camera must therefore be small and positioned in such a way that it does not impair their vision.

During a blink, the eye typically takes $\sim$100 ms to close, and this time increases with drowsiness [13]. To be able to differentiate between blinks and slower drowsy eye closure it is necessary to measure the time it takes to close the eyes. The typical closing time should therefore span several frames in order to measure the speed with reasonable resolution. We have chosen 60 fps as the target frame rate so that the time to close the eyes will typically span 6 frames.

The device must operate under a wide range of lighting conditions, from office lighting to driving in direct sunlight to driving at night. The camera subsystem must be able to capture video across this dynamic range, providing illumination if necessary.

The device should not restrict the motion of the wearer in any way, so there should be no cables tethering the wearer to an off-body device.

In the absence of any commercially available products meeting all of these requirements we decided to build our own camera module.
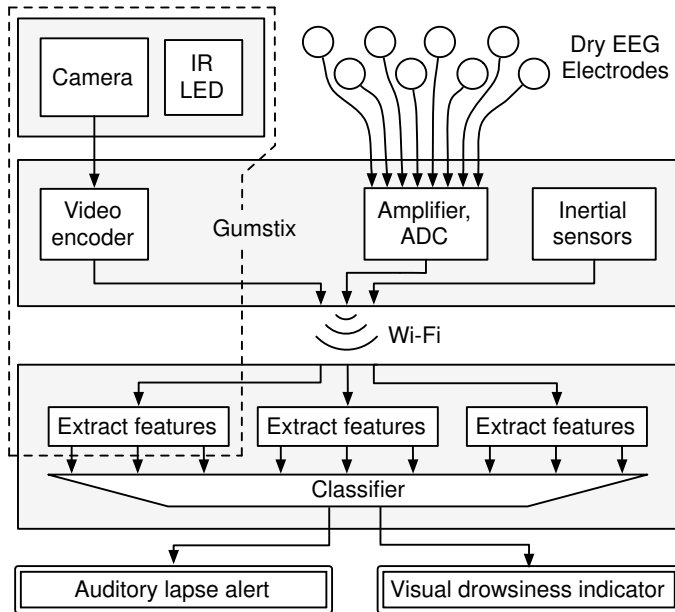
(a) Block diagram. The Gumstix computer-on-module acquires data from the sensors and streams it wirelessly to a PC for processing. The dotted region is the subsystem described in this paper.
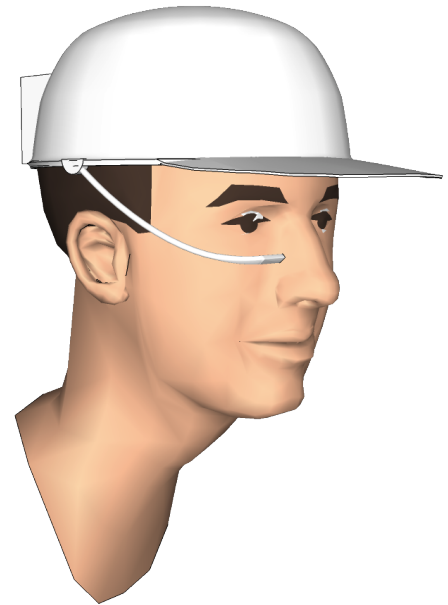
(b) Concept rendering. The camera is positioned below one eye.

Figure 1: Design of the multi-sensor lapse detection device.
© Simon Knopp

In Section 2 we outline some of the technical considerations when developing this camera module, followed by a description of the pupil detection algorithm in Section 3. Section 4 outlines the results of a preliminary experiment to determine the performance of the pupil detector. Finally in Section 5 we give our conclusions and possibilities for future work.

2. HARDWARE CONSIDERATIONS

There are a number of important factors to consider when designing a camera module to meet the requirements of Section 1.2. After describing the system architecture we discuss two of these considerations: the video frame rate and the illumination of the eye.

2.1 System Architecture

The camera module we have developed consists of an Omni-Vision OV7735[1] image sensor connected to a Gumstix Overo Fire[2] computer-on-module over a 10-bit parallel interface. The OV7735 is a colour CMOS sensor capable of capturing 60 fps at a resolution of 640×480 (VGA). It also has an on-board image signal processor with support for, among other things, cropping, scaling, and automatic gain and exposure control. The Gumstix is based around the Texas Instruments OMAP3530 system-on-chip which includes an ARM Cortex-A8 processor, a TMS320C64x+ DSP core, and a parallel camera interface.

The Gumstix captures video from the camera, applies H.264 compression, and streams it over Wi-Fi to a laptop

(Fig. 1a, dotted region). Physically, the completed device will have a form similar to the concept rendering in Fig. 1b, with the camera positioned on an adjustable arm below one eye.

2.2 Frame rate

To speed up development we decided to implement the image processing algorithms on a PC rather than the Gumstix. While this means that the device is then not a stand-alone unit (at least during development), it provides the freedom to experiment with complex image processing routines without being constrained by the speed of the Gumstix. However, in order to stream video to a PC the Gumstix must first compress it. Uncompressed 8-bit greyscale video at 640×480 and 60 fps has a bit rate of 147 Mbit/s—almost three times the maximum throughput of 802.11g Wi-Fi without accounting for network overhead.

Texas Instruments, the manufacturer of the system-on-chip (SoC) used on the Gumstix, provides an H.264 codec optimised to run on the SoC's DSP core. Unfortunately this codec can only achieve an average frame rate of approximately 30 fps at VGA resolution. To reach 60 fps we have to crop and scale the video considerably; lines are cropped from both the top and bottom, leaving 320×480, and the result is downsampled to 160×240. This reduction in the vertical field of view means that more care must be taken when aligning the camera with the eye, and increases the potential for the eye to move out of the frame if the camera mount is bumped.

Instead of blindly cropping the image, it may be possible in the future to locate a region of interest (ROI) on the Gumstix and stream that to the PC. Such a hybrid approach would

---

[1] http://www.ovt.com/
[2] http://www.gumstix.com/

(a) Ambient lighting only, with pupil barely discernable.

(b) Ambient + NIR with improved contrast between pupil and iris.

Figure 2: The effect of NIR illumination on pupil–iris contrast.

allow a simplistic eye tracking algorithm to run on the Gumstix, while retaining the frame rate advantage of streaming small images and the ability to run computationally intense algorithms on the PC. This approach would provide more tolerance in the positioning of the camera relative to the eye.

2.3 Illumination

The camera module has an infrared light-emitting diode (IR LED) with a peak wavelength of 850 nm to illuminate the eye. Under near-infrared (NIR) illumination the iris appears lighter than under visible light [2] which has the desirable effect of increasing the contrast between the pupil and the iris (Fig. 2). NIR illumination also allows us to capture video in the dark where using visible light would otherwise interfere with the wearer's vision.

On the back side of the camera module, facing away from the face, is an ambient light sensor. This sensor has two sensing elements which allow separate measurements of visible and infrared levels. Initially we intended to use the sensor as part of an adaptive illumination system, altering the power of the IR LED according to ambient light levels. After some experimentation, however, we found that with the IR LED at a fixed power output the camera's automatic exposure and gain controls were able to compensate sufficiently to provide images with constant brightness. The ambient light sensor may still be of use for compensating for lighting conditions in software when monitoring pupil diameter.

Eyes under infrared illumination can exhibit either the bright- or dark-pupil effect depending on the distance between the IR source and the optical axis of the camera [6]. The bright-pupil effect occurs when the source is close to the camera and IR enters the pupil, reflects off the retina, and comes back out of the pupil to the camera. The dark-pupil effect occurs when the IR source is positioned away from the camera. In this case the narrow aperture of the pupil prevents any reflected IR from reaching the camera. A dark pupil arrangement is easier to implement since there is no need for ring lights or beam splitters to provide co-axial illumination [1].

The need for separation between the IR LED and the camera, though, conflicts with the requirement for the module to be as small as possible. On the current hardware the two are separated by ∼6 mm. At the typical camera–eye distance of ∼35 mm this is far enough away to induce the dark-pupil effect under office lighting. In the dark, however, without visible light incident on the cheek and forehead to which the camera can adjust its exposure and gain, there is a glow visible at the bottom edge of the pupil. This interferes with the
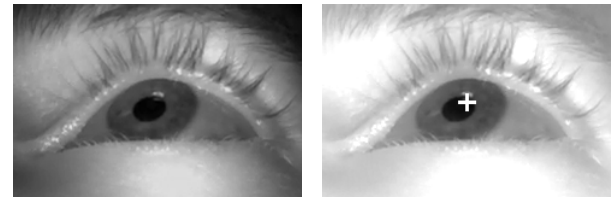
(a) The original frame.

(b) Weighted by distance from image centre with global minimum marked.

(c) Flood-filled pupil region.

(d) Pupil ellipse and seed point for next frame.

Figure 3: Locating the pupil by flood-filling about a dark seed point.

pupil localisation process described in Section 3.1; in subsequent revisions of the camera module hardware we will move the IR LED further from the camera to avoid this.

3. IMAGE PROCESSING

The following section describes the process by which the pupil is located in the video and how that information is used to measure eye closure.

3.1 Pupil localisation

The pupil localisation process is based on flood-filling about a dark seed point.

3.1.1 Initialisation

In the first frame of the captured video stream, the seed point is defined to be a dark point near the centre of the cropped frame. To exclude dark areas from, for instance, where the face curves back away from the camera, the brightness of each pixel is linearly weighted by its distance from the centre of the image. That is, for each pixel the distance to the centre of the cropped frame is added to its intensity. The seed point is then defined as the global minimum of the image (Fig. 3b). If the value of the pixel at the seed point is above a fixed threshold, the frame is discarded and the procedure is repeated on the next frame.

3.1.2 Pupil shape

After locating a suitable seed point, the shape of the pupil is defined by flood-filling about that point (Fig. 3c). This process starts at a point and recursively fills all connected pixels whose values are within a certain relative threshold. The thresholding step could compare each pixel to either its neighbours or the seed point. We compare each pixel to the seed point so that any blurring of the pupil–iris boundary resulting from a slightly out of focus image will not affect how far the flood continues.

The pupil is considered to be partly covered under two conditions: if the widest point of the pupil is within 5 pixels of the top of the pupil, or if the width of the pupil is more
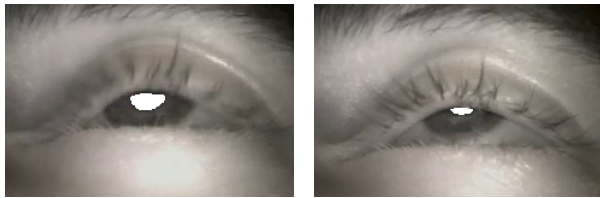
(a) A partly covered pupil during a blink.

(b) A proportionally wide pupil region.

Figure 4: Criteria for classifying the eye as partly closed.

than twice its height. An example of each of these cases is shown in Fig. 4.

If the pupil is not partly covered, the boundary of the pupil is approximated by fitting an ellipse to the boundary of the flood-filled region (Fig. 3d). This reduces the description of the pupil to five parameters (coordinates of the centre, lengths of major and minor axes, rotation). The seed point for the subsequent frame is then defined to be the centre of the ellipse.

3.1.3 Adjustment of seed point

In the form described so far, this algorithm can sometimes get "stuck" on dark areas that are not the pupil. For example, during a blink when the pupil is partly or completely covered the position of the seed point will not be updated. If a dark eyelash happens to pass through that seed point during the blink then the eyelash may be picked up as the pupil, which will result in the seed point tracking the eyelash in subsequent frames. To improve the algorithm's robustness in such situations, we add an adjustment step at this point. Regardless of whether the pupil is visible or covered, the seed point is moved to the darkest point within a 10×10 pixel window of its current position in each frame.

The process is then repeated from Section 3.1.2.

3.1.4 Discussion

Many existing pupil localisation algorithms rely on a sharp boundary between the pupil and the iris. For example, the "starburst" algorithm [8] looks for points at which the gradient exceeds a given threshold, and the algorithm of Świrski et al. [11] relies on a Canny edge detector. Given the short object distance of this camera setup (and therefore shallow depth of field) and the fact that it is prone to being bumped while in use, the captured video can often be slightly out of focus. This blurring reduces the gradient at the pupil–iris boundary. Because of this, in our experience gradient-based methods result in noticeable "jitter" in the pupil boundary.

In contrast, the flood-fill-based method described in this section searches for the points at which the intensity differs from that of the seed point by some threshold. This approach seems to be less affected by the focus of the image.

3.2 Eye closure

A useful measure of eye closure is "PERCLOS", the percentage of eyelid closure [4]. PERCLOS measures the percentage of time that the pupil is at least 80% covered over a one minute window. It has been found to correlate highly with lapses on a discrete psychomotor vigilance task [3] and is therefore used as a measure of fatigue. We use it here as a simple example of useful information that can be extracted

from the pupil.

The following procedure is used to calculate PERCLOS: For each captured frame, the pupil is located as per Section 3.1. If no pupil is visible in the frame, the eye is considered to be closed. If a partly covered pupil is visible (Section 3.1.2) then the eye is considered to be closed only if the height of the visible pupil region is less than 20% of the height of the pupil region in the most recent frame in which the pupil was not partly covered. In all other cases the eye is considered to be open. This binary eye state is appended to a ring buffer of $60\,\mathrm{fps} \times 60\,\mathrm{s} = 3600$ elements. The PERCLOS value is then simply the percentage of entries in the ring buffer which are "closed".

4. PRELIMINARY RESULTS

To get a preliminary estimate of the accuracy of the pupil localisation algorithm, we recorded two 15 minute sessions on a single subject. The tests were carried out at 1:00 pm and midnight after 6 and 17 hours of wakefulness respectively. The subject was seated in front of a computer screen and instructed to visually track a dot moving around the screen for the duration of the session, similar to the 2D tracking task of Poudel et al. [10]. Tracking this moving target simulates the range of eye motion typical of real-world tasks such as scanning the road ahead while driving.

4.1 Frame-by-frame analysis

To measure the accuracy of the pupil detection algorithm, four one-minute periods from each session were manually classified frame-by-frame and compared to the automatically classified values. Each frame was assigned to one of four categories:

I. Open — the entire pupil is visible,

II. Partly closed — $0\% <$ pupil coverage $< 80\%$,

III. Mostly closed — $80\% \leq$ pupil coverage $< 100\%$,

IV. Closed — the pupil is not visible.

These correspond to the open/closed/partly-closed groupings from Section 3.1 with the additional height condition from Section 3.2. Note that categories I & II are "open" and III & IV are "closed" for the purposes of the PERCLOS calculation but they are kept separate here for the sake of a more detailed analysis.

The results of this analysis are presented in Table 1. Each value in the table represents the number of frames that were manually classified as being in that row and automatically classified as being in that column. That is, if the algorithm produced the same results as a human, all off-diagonal entries would be zero. The table includes all frames from the eight manually classified one minute periods.

Table 1: Comparison of the number of frames in each manually vs. automatically classified category.

| | | Automatic | | | |
		I	II	III	IV
Manual	I	27673	375	0	42
	II	190	285	0	6
	III	0	17	2	15
	IV	0	2	0	195
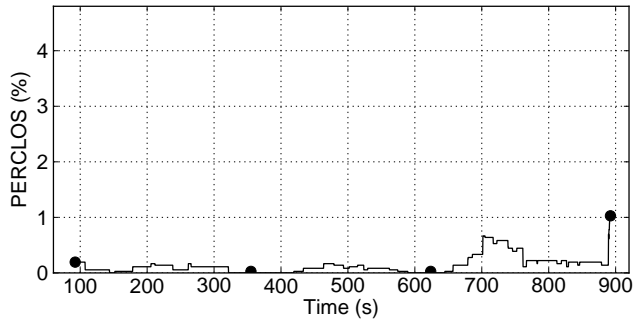
(a) PERCLOS for the 1 pm (alert) session.
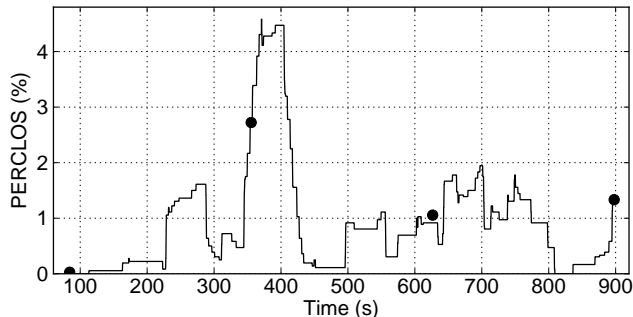
(b) PERCLOS for the 12 am (drowsy) session.

Figure 5: PERCLOS values for the two test sessions with values from the manually classified data overlaid (black dots).

Several of the frames manually classified as category I (open) were automatically classified as category II (partly covered) and vice versa. Some of these errors will be due to the ambiguity around which dark pixels belong to the pupil and which to the dark line between the eyelid and the cornea. Others are due to the glow from the IR LED altering the shape of the detected pupil region as described in Section 2.3. Note that these errors have no effect on the PERCLOS value since categories I and II are both "open".

Many of the differences for frames manually classified as category III (mostly covered) are due to having to estimate whether the visible pupil region is less than 20% of the full pupil height when classifying manually.

The frames manually classified as category I and automatically classified as category IV are those in which the algorithm failed to detect a pupil that was actually visible.

4.2 PERCLOS analysis

The value of PERCLOS over time for the each of the two sessions is plotted in Fig. 5. As expected, PERCLOS was both substantially higher and more variable in the second session when the subject was drowsy.

The PERCLOS values from the manually classified data

$$\frac{\text{III} + \text{IV}}{\text{I} + \text{II} + \text{III} + \text{IV}} \times 100\%$$

are marked with black dots on Figs. 5a and 5b. Despite the differences between manually and automatically classified values shown in Table 1, all PERCLOS values from the manually classified data are very close to the automatically classified values.

5. CONCLUSION & FUTURE WORK

We have developed a miniature infrared camera capable of capturing video at 60 fps and wirelessly streaming it to a PC. We have also demonstrated a pupil localisation algorithm which is capable of producing PERCLOS measurements that align closely with manually measured values.

We have used PERCLOS as an example application of the pupil localisation algorithm, but we intend to extract more useful information from the eye video. By tracking the pupil and eyelids over time it is possible to measure the blink frequency, blink duration, eyelid opening/closing speed, and pupil diameter. Patterns in each of these parameters can be used as indicators of the person's level of alertness. For instance, there is evidence of certain patterns of pupil diameter changes that occur shortly before people report feeling drowsy [9].

By combining the pupil location with data from inertial sensors we can estimate the wearer's gaze direction. From this information we may be able to detect diverted attention lapses in situations where it is known that the wearer should be maintaining attention in a particular direction.

The camera mounting system also needs to be improved. As yet, no effort has been made to work with people wearing glasses or sunglasses; with the camera in its current position, the frames of some glasses would block the view of the eye. We need to investigate the best position from which to see around the edge of the glasses or through the lens while still remaining outside the wearer's field of view. It may even be possible to integrate the camera into the frame of specially designed glasses to form part of the device.

6. REFERENCES

[1] J. S. Babcock and J. B. Pelz. Building a lightweight eyetracking headgear. In *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications*, ETRA '04, pages 109–114. ACM, 2004.

[2] J. Daugman. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):21–30, Jan. 2004.

[3] D. F. Dinges, G. Maislin, J. W. Powell, and M. M. Mallis. Evaluation of techniques for ocular measurement as an index of fatigue and the basis for alertness management. Technical Report DOT HS 808 762, National Highway Traffic Safety Administration (USA), 1998.

[4] L. Hartley, T. Horberry, N. Mabbott, and G. P. Krueger. Review of fatigue detection and prediction technologies. Technical report, National Road Transport Commission (Australia), Sept. 2000.

[5] C. R. H. Innes, G. R. Poudel, T. L. Signal, and R. D. Jones. Behavioural microsleeps in normally-rested people. In *Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4448–4451. IEEE, 2010.

[6] Q. Ji and X. Yang. Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. *Real-Time Imaging*, 8(5):357–377, 2002.

[7] R. D. Jones, G. R. Poudel, C. R. H. Innes, P. R. Davidson, M. T. R. Peiris, A. M. Malla, T. L. Signal, G. J. Carroll, R. Watts, and P. J. Bones. Lapses of responsiveness: Characteristics, detection, and

underlying mechanisms. In *Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1788–1791. IEEE, 2010.

[8] D. Li, D. Winfield, and D. J. Parkhurst. Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. In *Proceedings of the IEEE Vision for Human–Computer Interaction Workshop at CVPR*, pages 1–8, June 2005.

[9] J. Nishiyama, K. Tanida, M. Kusumi, and Y. Hirata. The pupil as a possible premonitor of drowsiness. In *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1586–1589, Aug. 2007.

[10] G. R. Poudel, R. D. Jones, and C. R. H. Innes. A 2-D pursuit tracking task for behavioural detection of lapses. *Australasian Physical & Engineering Sciences in Medicine*, 31(4):528–529, 2008.

[11] L. Świrski, A. Bulling, and N. Dodgson. Robust real-time pupil tracking in highly off-axis images. In *Proceedings of the 2012 Symposium on Eye Tracking Research & Applications*, ETRA '12, pages 173–176. ACM, 2012.

[12] L. Torsvall and T. Åkerstedt. Sleepiness on the job: continuously measured EEG changes in train drivers. *Electroencephalography and Clinical Neurophysiology*, 66(6):502–511, 1987.

[13] A. J. Tucker and M. W. Johns. The duration of eyelid movements during blinks: changes with drowsiness. *Sleep*, 28:A122, 2005.