35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

A wearable device for measuring eye dynamics in real-world conditions

Simon Knopp, *Student Member, IEEE*, Philip Bones, *Senior Member, IEEE*, Stephen Weddell, *Member, IEEE*, Carrie Innes, and Richard Jones, *Senior Member, IEEE*

***Abstract*— Drowsiness and lapses of responsiveness have the potential to cause fatalities in many occupations. One subsystem of a prototype device which aims to detect these lapses as they occur is described.**

A head-mounted camera measures several features of the eye that are known to correlate with drowsiness. The system was tested with eight combinations of eye colour, ambient lighting, and eye glasses to simulate typical real-world input conditions. A task was completed for each set of conditions to simulate a range of eye movement—saccades, tracking, and eye closure.

Our image processing software correctly classified 99.3% of video frames as open/closed/partly closed, and the error rate was not affected by the combinations of input conditions. Most errors occurred during eyelid movement. The accuracy of the pupil localisation was also not influenced by input conditions, with the possible exception of one subject's glasses.

I. Introduction

In many occupations, lapses of responsiveness can have severe, or even fatal, consequences both for employees and for those around them. This is especially true for jobs that require long periods of sustained attention on a monotonous or repetitive task. For instance, commercial vehicle drivers, pilots, air-traffic controllers, and some medical professionals risk causing fatalities if their task-oriented attention lapses, even briefly. A device capable of detecting these lapses and intervening quickly has the potential to save lives.

The camera hardware and image processing software described in this paper form one subsystem of a head-mounted multi-modal device for measuring drowsiness and detecting lapses of responsiveness. In addition to the camera module there will be several channels of EEG to measure brain activity and inertial sensors to measure head movement. By integrating features from these three modalities it should be possible to detect lapses more quickly and reliably than by using EEG [1] or video [2] alone. Fig. 1 shows a concept rendering of this device with the camera positioned on an adjustable arm below one eye.

S. J. Knopp is a PhD student in the Department of Electrical & Computer Engineering, University of Canterbury, and the New Zealand Brain Research Institute, Christchurch, New Zealand. simon.knopp@pg.canterbury.ac.nz

P. J. Bones and S. J. Weddell are with the Department of Electrical & Computer Engineering, University of Canterbury, Christchurch, New Zealand. {phil.bones,steve.weddell}@canterbury.ac.nz

C. R. H. Innes is with the Department of Medical Physics and Bioengineering, Christchurch Hospital, and the New Zealand Brain Research Institute, Christchurch, New Zealand. carrie.innes@nzbri.org

R. D. Jones is with the Department of Medical Physics and Bioengineering, Christchurch Hospital, the Department of Electrical & Computer Engineering, University of Canterbury, and the New Zealand Brain Research Institute, Christchurch, New Zealand. richard.jones@nzbri.org

A. Requirements

This device is 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. For the camera subsystem this means that the camera must be small and positioned so that it will not obstruct their vision.

To differentiate between blinks and drowsy eye closure it is necessary to measure the duration of eye closure. Drowsiness causes the duration of transient eye closures to increase from a mean of 0.3 ms when alert to a mean of 144 ms, and also to become more variable [3]. The target frame rate for this device is 60 fps, giving a range of up to ∼9 frames of eye closure to use as a measure of drowsiness.

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 capture usable video across this dynamic range.

The device must not restrict the motion of the wearer in any way, so there must not be cables tethering the wearer to any off-body equipment.

In the absence of any commercially available camera modules meeting all of these requirements, we decided to build our own.

II. Hardware

The camera module (Fig. 2) uses an OmniVision OV7735[1] image sensor and is connected to a Gumstix Overo Fire[2] computer-on-module over a 10-bit parallel interface. The

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

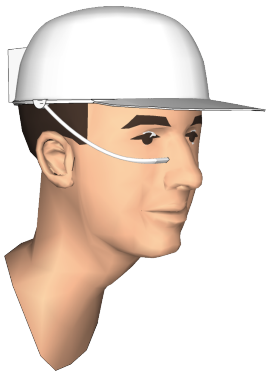


Fig. 1. Concept rendering of the lapse detection device. The camera is positioned below one eye.

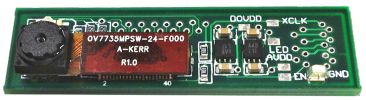


Fig. 2. Camera module developed for the lapse detection device, measuring 12×43 mm.

978-1-4577-0216-7/13/$26.00 ©2013 IEEE

OV7735 is capable of capturing 60 fps at a resolution of 640×480. It has an on-chip image signal processor with support for, among other things, cropping, scaling, and automatic gain and exposure control. The Overo Fire is based around the OMAP3530 system-on-chip which includes an ARM Cortex-A8 processor, a C64x+ DSP core, and a parallel camera interface.

The Gumstix captures video from the camera, applies DSP-accelerated H.264 compression, and streams it over Wi-Fi to a laptop for processing. This split arrangement enables rapid development of image processing algorithms on a PC without the memory and clock speed constraints of an embedded processor.

The camera module includes a near-infrared (NIR) LED to illuminate the eye. Under NIR illumination the iris appears lighter than under visible light [4] which has the desirable effect of increasing the contrast between the pupil and the iris. Using infrared also enables the camera to capture video in the dark where visible illumination would otherwise interfere with the wearer's vision.

III. Image processing

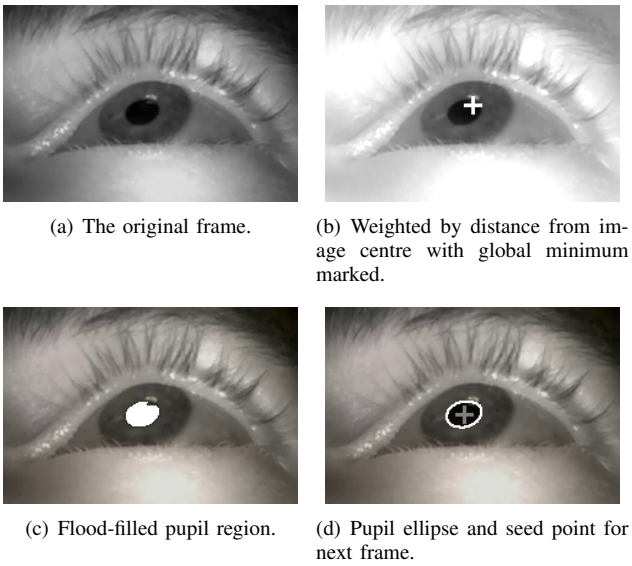
At the lowest level, the image processing software aims to measure the pupil position and diameter and identify whether the eye is open or closed. From these measurements it is possible to derive several measures that are known to correlate with drowsiness. For example, the duration of eye closure is known to increase with drowsiness [3], [5]. The pupil position (in combination with the measured head position) can be used to estimate gaze direction, which in turn can indicate diverted attention. There is also evidence that patterns of changes in pupil diameter are correlated with drowsiness [6].

A. Initialisation

The pupil localisation process is based on flood-filling about a dark seed point. This is a common process for locating connected components which we have applied to pupil localisation [7]. In the first frame of video, the seed point is defined to be a dark point near the centre of the 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 frame is added to its intensity. The seed point is then defined as the global minimum of the image (Fig. 3(b)).

B. Pupil shape

The shape of the pupil is defined by flood-filling about the seed point (Fig. 3(c)). This process starts at a point and recursively fills all connected pixels whose values are within a certain threshold. The flood-fill threshold can be configured to compare each pixel either to its neighbours or to the seed point. In this case the threshold is taken relative to the seed point so that any blurring of the pupil–iris boundary resulting from a slightly out of focus image will not affect the extent of the flood.



(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.

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

The pupil is considered to be partly covered under two conditions: if the widest point of the pupil is close to the top of the pupil, or if the width of the pupil is more 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. 3(d)). This reduces the description of the pupil to five parameters (coordinates of the centre, lengths of major and minor axes, rotation). The centre of the ellipse is then used as the seed point for the next frame.

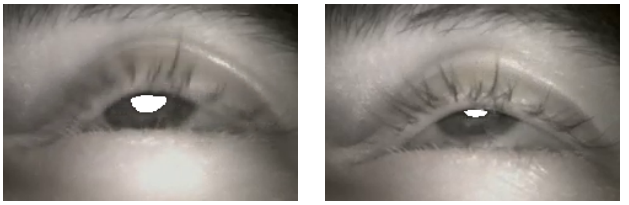
C. Filtering

This process is prone to incorrectly labelling dark eyelashes as the pupil during eye closure. To avoid this, large discontinuities in pupil position and diameter are filtered out. That is, if in one frame the pupil diameter changes by more than four pixels or the position changes by more than the pupil's radius, the "pupil" from that frame is rejected and the one from the previous frame carried forward.

The process from Section III-B is then repeated.

D. Robustness to focus

Many existing pupil localisation algorithms rely on a sharp boundary between the pupil and the iris. For example,



(a) A partly covered pupil during a blink.

(b) A proportionally wide pupil region.

Fig. 4. Criteria for classifying the eye as partly closed.

the "starburst" algorithm [8] looks for points at which the gradient exceeds a given threshold, and the algorithm of Świrski et al. [9] 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, gradient-based methods result in noticeable "jitter" in the pupil boundary when used on video from this device. In contrast, our flood-fill-based method searches for the points at which the intensity differs from that of the seed point by some threshold, regardless of the gradient. Our approach seems to be less affected by the focus of the image.

IV. Experiment

An experiment was carried out as part of the development process for the camera subsystem. As such, it is not an analysis of the performance of a completed system but rather is intended to direct the next stages of development.

The experiment aimed to determine whether the hardware meets the requirements in Section I-A and whether the software can identify eye features as accurately as a human. The chosen input conditions represent a range of situations typical of real-world usage. A total of eight input conditions were tested, representing every combination of three binary variables: *eye colour* – one subject with light irises, one with dark; *ambient lighting* – office lighting and a darkened room; *glasses* – with and without prescription eye glasses. From previous experience, these three variables were expected to have the most effect on the system's performance.

A task was completed for each set of conditions to emulate a typical range of eye movements—saccades, tracking, and eye closure. The subject was seated in front of a computer screen and instructed to watch a dot as it moved around the screen. Initially the dot jumped to each corner of the screen at 1 s intervals. This was followed by a random 2D tracking task similar to that of Poudel et al. [10] for 3.5 s. At that point a beep sounded, instructing the subject to close their eyes. After 1.5 s another beep sounded, the subject opened their eyes and resumed tracking for the remaining 3 s.

Every frame of video from the eight 12 s sessions (5760 frames) was manually annotated with the position and diameter of the pupil. Each frame was also assigned one of four categories: "corrupt", if the data was visibly corrupted, "open" if the pupil was completely visible, "partly closed" if the pupil was partly covered by an eyelid, or "closed" if the pupil was not visible at all.

V. Results

A. Hardware

1) Video corruption: Prior to transmission from the Gumstix, the video is compressed to H.264. Because of the structure of this format, if a frame is lost during transmission over the Wi-Fi network, the subsequent frames will be corrupt for up to 0.5 s.

The only frames which were manually classified as "corrupt" occurred at the beginning of each video, lasting for an average of 14 frames. This was expected since there is currently no synchronisation between the Gumstix and the laptop, meaning the laptop can miss the first few frames of the video stream. Corruption of ∼0.25 s of video at the beginning of the stream has no effect on the functionality of this device.

2) Frame rate: The time at which each frame of video arrived at the laptop was logged during recording. The median frame interval for each session was close to the expected value of 16.67 ms (60 fps). The distributions around this value varied, probably due to differing network conditions. Since the video is captured at a known and fixed frame rate, the actual rate at which frames reach the laptop is not important as long as no frames are dropped, as discussed in Section V-A.1. Increased latency of the order observed in this experiment will have a negligible impact on the speed at which lapse events can be detected and relayed to the user.

B. Software

Box plots in this section have boxes indicating the interquartile range, a line indicating the median, and whiskers extending to the most extreme data point within 1.5× the interquartile range. For sizes given in pixels, 1 px ≈ 0.15 mm. Subject 1 had light-coloured irises, subject 2 dark.

1) Pupil position: The error in the pupil position is defined as the Euclidean distance between the manually annotated centre of the pupil and the centre of the pupil detected by the software. For six of the eight sets of input conditions the median pupil position error was less than 0.7 pixels (Fig. 5). The remaining two were both for subject 1 with glasses. This could be because subject 1's glasses affect the accuracy of the pupil detection algorithm or because the camera was poorly positioned for those measurements—further investigation is needed to determine the cause. Regardless, for the application of detecting diverted attention by approximating gaze, an error of 3.5 px is acceptable.

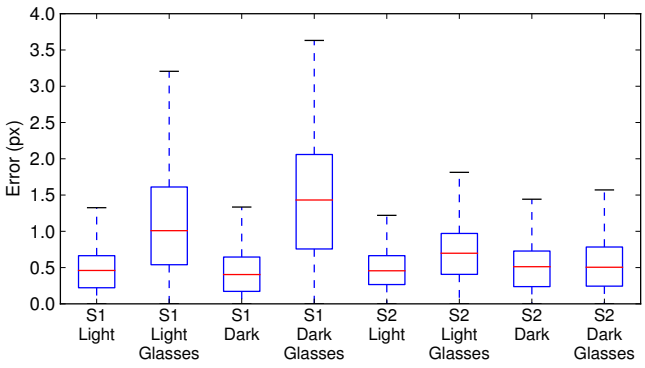


Fig. 5. Error in the position of the automatically detected pupil.

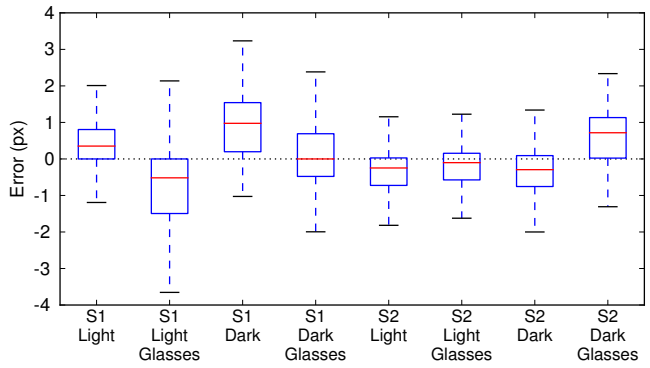


Fig. 6. Error in the diameter of the automatically detected pupil.

2) Pupil diameter: Fig. 6 shows the differences between the pupil diameter measured by the software and that measured manually for each set of input conditions. The errors in the pupil diameter measurements are large enough to cause a problem when measuring fluctuations in diameter. While the simulated lighting conditions used in this experiment are typical of some real-world situations, if the device is to be used in a wide variety of situations then it will also need to contend with rapidly changing lighting, such as when driving past trees. Rapidly changing lighting induces fluctuations in the pupil diameter quite separate from any induced by drowsiness, so fluctuations in pupil diameter are unlikely to be useful for measuring drowsiness.

3) Eye closure: The accuracy with which the software detects eye closure is the most important performance metric for this device. Table I compares the category that each frame was assigned manually (Section IV) against the category assigned by the image processing software (Section III-B). If the software produced the same results as a human, all off-diagonal entries would be zero. Of the 5706 uncorrupted frames, the software categorised 38 incorrectly (0.7%). Most of these errors, though, occurred during eye closure. 33 of these 38 errors occurred when the software classified a "partly closed" frame as either "open" or "closed". That is, during a blink the software tends to keep classifying frames as "open" for longer, and then jumps straight to "closed". In practical terms, if the duration of eye closure is being used as a drowsiness indicator, categorising a "partly closed" frame as "open" has little effect because the time that the eyes are detected as "closed" remains the same. The errors were evenly distributed across the combinations of input conditions—no condition or combination of conditions was found to increase the error rate.

TABLE I

COMPARISON OF MANUAL CLASSIFICATION TO SOFTWARE OUTPUT.

		Manual		
		Closed	Partial	Open
Auto	Closed	626	7	0
	Partial	0	26	4
	Open	1	26	5016

VI. CONCLUSIONS & FUTURE WORK

We have developed a small head-mounted camera module that can capture video at 60 fps and transmit it wirelessly to a PC. Both hardware and software have been shown to work reliably under a variety of input conditions typical of real-world usage. Differences in eye colour, ambient lighting, and the presence of glasses had no effect on the accuracy with which the pupil was located, with the possible exception of one subject's glasses.

Most importantly, the open/closed eye state was classified incorrectly for only 0.7% of the frames, however most of these errors occurred as the eye was closing. While the current error level is acceptable for measuring the duration of eye closure, it would need to be reduced if measuring the speed at which the eyes open and close proves to be highly predictive of future lapses.

Adding an eyelid localisation step to the algorithm could improve the detection of the partially closed eye state and enable measurement of the speed of eyelid movement. While there is evidence that eyelid speed is a useful metric [11], it is likely to be a difficult image processing task because of the interference of eyelashes.

REFERENCES

[1] C. Berka, D. J. Levendowski, M. M. Cvetinovic, M. M. Petrovic, G. Davis, M. N. Lumicao, V. T. Zivkovic, M. V. Popovic, and R. E. Olmstead, "Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset," *Int. J. Human-Computer Interaction*, vol. 17, no. 2, pp. 151–170, 2004.

[2] M. Sigari, "Driver hypo-vigilance detection based on eyelid behavior," in *Proc. 7th Int. Conf. Advances in Pattern Recognition*. IEEE, 2009, pp. 426–429.

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

[4] J. Daugman, "How iris recognition works," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 21–30, Jan. 2004.

[5] 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," National Highway Traffic Safety Administration (USA), Tech. Rep. DOT HS 808 762, 1998.

[6] J. Nishiyama, K. Tanida, M. Kusumi, and Y. Hirata, "The pupil as a possible premonitor of drowsiness," in *Proc. 29th Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2007, pp. 1586–1589.

[7] S. J. Knopp, P. J. Bones, S. J. Weddell, C. R. H. Innes, and R. D. Jones, "A miniature head-mounted camera for measuring eye closure," in *Proc. 27th Conf. Image & Vision Computing New Zealand*, ser. IVCNZ '12. ACM, 2012, pp. 313–318.

[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 *Proc. IEEE Vision for Human–Computer Interaction Workshop at CVPR*, June 2005, pp. 1–8.

[9] L. Świrski, A. Bulling, and N. Dodgson, "Robust real-time pupil tracking in highly off-axis images," in *Proc. 2012 Symposium on Eye Tracking Research & Applications*, ser. ETRA '12. ACM, 2012, pp. 173–176.

[10] G. R. Poudel, R. D. Jones, and C. R. H. Innes, "A 2-D pursuit tracking task for behavioural detection of lapses," *Austral. Phys. Eng. Sci. Med.*, vol. 31, no. 4, pp. 528–529, 2008.

[11] M. Johns, A. Tucker, R. Chapman, K. Crowley, and N. Michael, "Monitoring eye and eyelid movements by infrared reflectance oculography to measure drowsiness in drivers," *Somnologie*, vol. 11, no. 4, pp. 234–242, 2007.