COGAIN Symposium
Wuppertal, August 21st 2017.

Gaze interaction using low-resolution images at 5 FPS

Carlos E. L. Elmadjian
University of São Paulo, Brazil

Antonio Diaz-Tula
University of São Paulo, Brazil

Fernando O. Aluani
University of São Paulo, Brazil

Carlos H. Morimoto
University of São Paulo, Brazil

With eye trackers gradually becoming personal wearable devices, gaze-based interaction will become a relevant technique for wearable applications. However, it is a common belief that high-resolution images and high frame rates are desirable to achieve the accuracy and precision required for human interaction. Because of the high computational load, a wearable eye tracker would have their batteries quickly drained out. In this paper we investigate how much processing power can be saved by lowering these requirements, and still maintain the performance adequate for human interaction. We have conducted an experiment using a head-mounted Pupil Labs eye tracker. Our results from 10 participants show that accuracy and precision remain below one degree of error for image resolution of 240 lines, and frame rates as low as 5 frames per second (FPS). Using this minimum setup, we estimate that power consumption can be reduced by 90% compared to the eye tracker camera regular settings (480 lines and 30 FPS). We also propose an algorithm that successfully detects reading behavior in real-time at 5 FPS in order to demonstrate the usefulness of gaze data at such low rates.

Keywords: low-power eye tracking, frame rate, image resolution

Introduction

Traditional gaze-based computer applications were developed for desktops (Jacob, 1990; Majaranta & Räihä, 2002). However, with recent advances in head-mounted eye trackers, new opportunities have been created for control and interaction using wearable computers.

Recent head-mounted eye trackers have become sufficiently light and comfortable to be used in natural environments. Current models such as the SMI Eye Tracking Glasses and the Tobii Pro Glasses 2 provide accuracy lower than 1^o and frame rates above 60 Hz. Despite significant evolution in the form factor, real-time eye tracking is still a challenge for mobile and wearable devices, so data is generally stored to be processed later. For wearable gaze-based interaction though, having such a heavy job running all the time would quickly consume all the power stored in the device's battery.

The use of low resolution cameras can reduce the computational power needed to segment the pupil and estimate the point-of-gaze (PoG). Processing less frames per second can also reduce power consumption. Nonetheless, the effects of lowering image resolution and frame rate on PoG estimation and usability of gaze-based interfaces have not been properly investigated. In this paper we present the results of an experiment that shows how the Pupil gaze estimation algorithm (Kassner, Patera, & Bulling, 2014) performs with low resolution images and video frame rates. To demonstrate the feasibility of gaze-based interaction in such conditions, we have developed an algorithm for reading detection in low frame rates that outperforms the state of the art.

Methods

Ten volunteers (undergraduate and graduate students, one female) took part in the experiment. All had normal or corrected-to-normal vision using contact lenses. The study was conducted in a room with regular illumination and no sunlight incidence, using a binocular head-mounted Pupil eye tracker (Kassner et al., 2014). Data was collected using the Pupil Labs software while participants were seated in a fixed chair at about 55 cm from a 22" monitor.

The task consisted of looking at 17 visual targets (concentric rings presented in random order) displayed on the monitor. Nine targets were used for calibration and all 17 targets were used for error computation. The Pupil eye tracker uses a simple feature based gaze estimation technique, but different and more sophisticated approaches can also be found in the

The 2017 COGAIN Symposium: Communication by Gaze Interaction
Wuppertal, Germany. August 19th and 21st, 2017

literature (Hansen & Ji, 2010; Coutinho & Morimoto, 2013; Z. Zhang & Cai, 2014; X. Zhang, Sugano, Fritz, & Bulling, 2015).

Four 2D markers were placed at each corner of the monitor so it could be reliably detected on the scene image. Error was computed as the difference from the scene point estimated by the Pupil algorithm and the projection of the target on the scene image computed with the markers. For each target we recorded 2 seconds of video, which roughly corresponds to 60 frames for the eye and scene cameras. The two eye cameras had a resolution of 480 lines and the scene camera, 720 lines.

Two different conditions were used to collect data: the **baseline** and the **central** conditions. In the baseline, participants used a chin-rest to keep their head steady at about 55 cm from the monitor. For the central condition, participants did not use the chin-rest and could perform natural head movements to look at the targets. Recorded eye videos were down-sampled to 15, 10, and 5 frames per second (FPS), and the resolutions were reduced to 240, 120, and 60 lines. Then an off-line adapted version of the Pupil Labs software was used to process each video combination of FPS × resolution. The scene video remained at 720 lines for all conditions.

Results

Results show that accuracy and precision are not affected using an eye image with 240 lines instead of the camera's full 480 lines. Furthermore, with 240 lines the gaze estimation error remains below 1^o, as it can be seen in Table 1. Another interesting result is that accuracy and precision are not affected when FPS are reduced from 30 Hz to 15, 10, and 5 Hz. Additionally, there was no difference between the central and baseline conditions, which is explained by the fact that the Pupil eye tracker estimates the point-of-gaze by mapping the pupil center to the detected monitor on the scene image, thus being robust to some head movement.

For video resolutions of 120 and 60 lines, mean error in gaze estimation was between 8^o and 25^o for all FPS, probably due to the fact that pupil contours were hard to detect at such resolution, lowering the overall confidence value provided by the Pupil-Labs algorithm. Hence, given the larger error and standard deviation, we did not include these results.

Power consumption was estimated through an experiment using the *psutil* Python library. During this estimation, only one of the processor cores (Intel Core-i7, 1.90 GHz) was activated. The processor was configured to operate statically at its maximum frequency. The Pupil algorithm was evaluated on a set of videos with the same content (eye moving to 17 positions) but with different image resolutions (480 and 240) and FPS (30, 15, 10, and 5). For the 240-line resolution, we found that, compared to 480 lines at 30 FPS, power consumption was reduced from 72% (30 FPS) to 90% (5 FPS).

Table 1
Gaze estimation error for the center and baseline conditions.

	Center		**Baseline**	
FPS/Lines	**480**	**240**	**480**	**240**
30	0.75±0.45	0.78±0.49	0.83±0.7	0.81±0.55
15	0.76±0.46	0.77±0.47	0.84±0.75	0.81±0.55
10	0.75±0.44	0.77±0.47	0.82±0.64	0.80±0.54
5	0.75±0.44	0.77±0.50	0.86±0.83	0.79±0.55

Gaze interaction

With lower frame rates, detection of fixations and saccades might be affected, compromising gaze-based interaction. At 5 Hz, saccades are practically untraceable and fixations require longer periods of data acquisition to be recognized. Nonetheless, we argue that some behaviors with well-established patterns such as reading could still be used as an interaction mechanism in these conditions to mediate user tasks in a wearable context. Thus, we developed an algorithm (Elmadjian, Kurauchi, & Morimoto, 2016) for real-time reading detection and performed an experiment with 9 participants in which our algorithm showed a true positive rate above 90% and accuracy of 86% at 5 Hz. This was possible due to a pattern-matching strategy that considered both spatial differences and temporal displacements in fixations.

Conclusions

In this paper we have presented preliminar results about the effect of using lower image resolution and frame rates on gaze estimation accuracy and precision for the Pupil eye tracker. The experimental results showed that, by using an eye image with 240 lines, both accuracy and precision are similar to the original resolution of 480 lines, with an error below 1^o of visual angle. We have also found that processing only 5 FPS did not affect accuracy and precision. By combining a smaller image resolution and FPS, it might be possible to save up to 90% of processing power in the wearable eye tracker. To demonstrate that data at low rates is still useful for interaction, we have shown that it is possible to detect reading behavior at only 5 Hz, though for some interaction purposes that require low latency, such as eye typing, 10 to 15 Hz might improve user experience.

Acknowledgments

This work is supported by the São Paulo Research Foundation (FAPESP) under Grants No. 2016/10148-3, 2015/26802-1, and 2012/04426-0.

References

Coutinho, F. L., & Morimoto, C. H. (2013, February). Improving head movement tolerance of cross-ratio based eye trackers. *Int. J. Comput. Vision*, *101*(3), 459–481. Retrieved from `http://dx.doi.org/10.1007/s11263-012-0541-8`

Elmadjian, C., Kurauchi, A., & Morimoto, C. H. (2016). Recognizing reading behavior in low-power conditions: a step further towards wearable computing. In *Extended proceedings of ihc'16, brazilian symposium on human factors in computing systems. workshop of eye gaze applications - wega* (pp. 222–226). São Paulo: Brazilian Computer Society. Retrieved from `http://www.lbd.dcc.ufmg.br/curadoria/minger/recentinsertion/ihc/2016/pdf/043.pdf`

Hansen, D. W., & Ji, Q. (2010). In the eye of the beholder: a survey of models for eyes and gaze. *IEEE transactions on pattern analysis and machine intelligence*, *32*(3), 478–500.

Jacob, R. J. (1990). What you look at is what you get: eye movement-based interaction techniques. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 11–18).

Kassner, M., Patera, W., & Bulling, A. (2014). Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 acm international joint conference on pervasive and ubiquitous computing: Adjunct publication* (pp. 1151–1160). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/2638728.2641695`

Majaranta, P., & Räihä, K.-J. (2002). Twenty years of eye typing: Systems and design issues. In *Proceedings of the 2002 symposium on eye tracking research & applications* (pp. 15–22). New York, NY, USA: ACM. Retrieved from `http://doi.acm.org/10.1145/507072.507076`

Zhang, X., Sugano, Y., Fritz, M., & Bulling, A. (2015). Appearance-based gaze estimation in the wild. *CoRR*, *abs/1504.02863*. Retrieved from `http://arxiv.org/abs/1504.02863`

Zhang, Z., & Cai, Q. (2014). Improving cross-ratio-based eye tracking techniques by leveraging the binocular fixation constraint. In *Proceedings of the symposium on eye tracking research and applications* (pp. 267–270). New York, NY, USA: ACM.