An experimental setup and protocol for offline performance assessment of eye/gaze tracking strategies

Mateus M. C. Cruz, Ronaldo F. Zampolo
Signal Processing Laboratory – LaPS
Federal University of Pará – UFPA
Belém-PA, Brazil
mateusmotacc@gmail.com, zampolo@ufpa.br

Matthieu P. Silva, Patrick Le Callet
LS2N, UMR CNRS 6004
University of Nantes
Nantes, France
{matthieu.perreiradasilva, patrick.lecallet}@univ-nantes.fr

***Abstract*—Several applications (crowdsourcing, human-computer interface, and web design, just to name a few) can largely benefit from reliable low-cost eye/gaze tracking solutions with either reduced or no-calibration procedure. Although very accurate eye/gaze trackers are currently affordable to common users, special hardware is still needed and calibration procedure remains apparent. This paper presents an experimental setup and protocol aimed at providing ground truth and supporting data to allow offline performance evaluation of different eye/gaze tracking schemes. This work is part of an ongoing research on the use of visual attention models in conjunction with eye/gaze tracking strategies for performance improvement as well as reduction of calibration phase in low-cost solutions. The proposed experimental setup hardware is composed of off-the-shelf equipment (eye/gaze tracker, conventional webcam, and a desktop computer). In turn, the control software is developed in PsychoPy/Python and made publicly available. Data acquisition components and conceived experimental protocol are discussed in details.**

***Keywords*—Experimental setup, experimental protocol, visual attention models, low-cost eye/gaze tracking.**

I. Introduction

Eye/gaze tracking (ET) aims at either monitoring eye movements or estimating one's gaze direction. Such a technology has been applied to a wide range of areas, including marketing, gaming, virtual reality, psychophysics, web design, and interactive interfaces [1]. Two specific applications are of special interest to this ongoing project: crowdsourced subjective experiments using webcam-based eye-tracking, and human-computer interfaces (HCI). Both of them can largely benefit from the availability of reliable low-cost ET solutions with reduced (ideally, transparent) calibration procedure. Such features would favour a larger base of users, remote acquisition of higher quality data, and also higher levels of user comfort.

At present, fairly accurate eye/gaze trackers are affordable to common users [2], [3]. Two drawbacks of current solutions, however, deserve further comments, once they limit in some sense a broader spread of eye/gaze trackers: specific equipment is generally required, like dedicated cameras and infra-red (IR) illumination apparatus; and the existence of an apparent calibration procedure to which the user must participate actively. The latter reduces the spontaneity of the user, by reminding him/her that his/her eyes movements are being recorded. In addition, there are other situations, for instance in experiments with infants or HCI for handicapped people, where user allegiance is not even reasonable.

Those aspects have been motivating the research community to develop low-cost ET strategies in which specific equipment and IR illumination are replaced by conventional webcams and visible light, respectively [4]. Generally, in comparison with commercial solutions, such low-cost alternatives suffer from lower accuracy, precision, and sampling rate as well. Attempts to reduce, or eliminate, explicit calibration phase have also been made by using visual attention models [5] in order to provide additional information about salient spots of a viewed scene, what is supposed to guide and somehow stabilise point of regard (PoR) estimations [6], [7].

In this sense, this paper proposes an experimental setup and protocol aimed at providing ground truth and supporting data to allow offline performance evaluation of different ET schemes. This work is part of an ongoing research on the use of visual attention models in conjunction with eye/gaze tracking strategies (ET+VAM) for performance improvement as well as reduction of calibration phase in low-cost solutions.

The proposed experimental setup hardware is composed of off-the-shelf equipment (commercial eye/gaze tracker, conventional webcam, and a desktop computer). The whole experimental equipment is controlled by a PsychoPy/Python script, whose source code is publicly available. In other words, the referred script is responsible for the entire experimental protocol implementation, including exhibition of test images, control of webcam and ET device, generation of all data necessary to recreate a particular test and for offline synchronisation.

The remainder of the paper is organised as follows. Section II justifies the proposed experimental setup by comparing it with the straightforward approach where one wants to assess the performance of different ET devices in parallel. In Section III, the elements that compose the data acquisition system are discussed. Section IV addresses the experimental protocol conceived, and in Section V the main conclusions

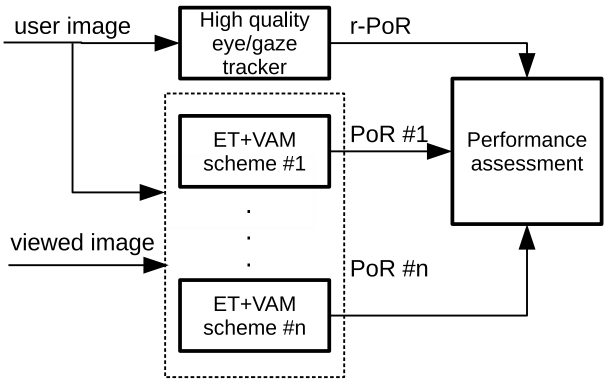

Fig. 1. Parallel PoR acquisition arrangement to assess the performance of different ET+VAM schemes.

of the paper are pointed out. A list of references ends the document.

II. Purpose of the proposed experimental setup

The experimental setup proposed in this work was conceived to solve a practical hurdle found during an investigation on the use of visual attention models in conjunction with low-cost ET strategies. The original idea of that investigation is to improve the performance of low-cost eye/gaze trackers by means of additional information concerning salient regions of viewed scenes provided by visual attention models.

However, to assess the performance of any association between ET schemes and visual attention models, ground-truth data are required. Specifically, such ground-truth data should come from a high-quality tracking device, whose PoR estimations would provide a reference to compare the performance of different ET+VAM strategies.

The key point here is that in order to compare the performance of different ET+VAM systems, the PoR estimations of all candidate systems (the reference data included) must be derived from the very same event, i. e. all systems must work in parallel (Fig. 1), what is challenging to be implemented online due to eventual physical space limitations, synchronisation issues, and adequate management of the experimental setup.

Thus, to avoid preceding difficulties, an offline approach might be preferred. In such an approach, the reference PoR (ground-truth data) is recorded along with all required ET input data as well as information about stimuli schedule. As ET inputs, we have user face images and geometrical features of the experimental setup (viewing distance, screen dimensions, user camera position, and so on.). In turn, the stimuli schedule consists of the list of exhibited images and their corresponding timestamps. Moreover, ground-truth PoR, ET input data and stimuli schedule information must have a unique time base to allow *a posteriori* synchronisation.

The data collected by such a system is stored in a database, from which they can be retrieved for offline performance evaluation of particular ET+VAM schemes once at a time (Fig. 2).

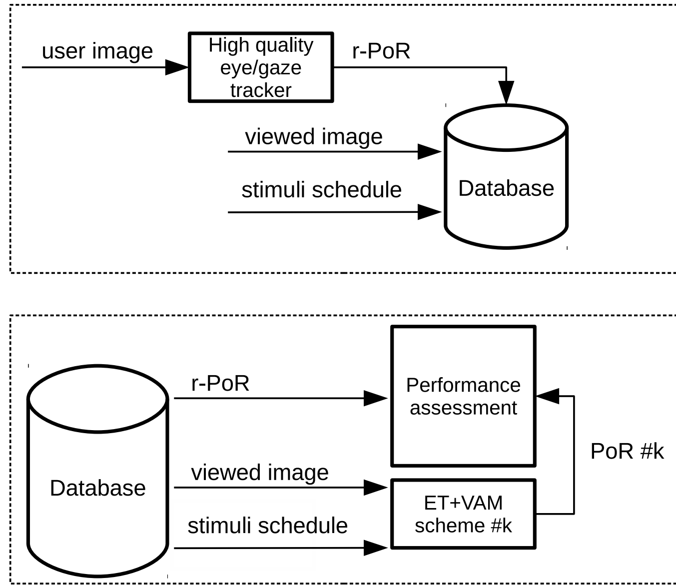

Fig. 2. Offline approach: PoR for a particular ET+VAM scheme (PoR #k) is estimated from data retrieved from a database.

III. Data acquisition components

The proposed data collection system is composed of the following components: user image acquisition; eye/gaze tracking; stimuli exhibition; and experiment controller.

The user image acquisition component comprises a webcam and corresponding control functions implemented by using the Open Source Computer Vision Library (OpenCV) [8]. This means that any webcam supported by OpenCV can be used. In our implementation, we work with a Logitech C920 pro HD, whose frame resolution and rate are set to 960×720 pixels and 30 frames per second (fps), respectively. The purpose of this component is to capture images of the user face during the experimental procedure from which one can perform offline estimation of gaze direction.

The eye/gaze tracker used in this setup is a Tobii EyeX [2]. Controlling and integration of this equipment into the whole experimental system are made by the data acquisition library in C/C++, downloaded from Tobii web page. Due to its accuracy, the points of regard (PoR) retrieved from Tobii EyeX will be considered as ground-truth data whenever needed. In our experiments, data rate varies slightly around 60 Hz.

The exhibition system consists of a conventional LCD flat screen of 19 inches (LCD HP Compaq LA1905wg), on which images or videos can be shown to the user. Such visual stimuli are intended to elicit visual attentional response of the subject. Note that no matter what monitor is actually used, the viewing distance should be chosen accordingly. Guidelines for several types of screens and resolutions can be found in [9], [10].

Controlling, integration and synchronisation of components are performed with PsychoPy [11], [12], an open-source application written in Python, aimed for running experiments in neuroscience, psychology and psychophysics.

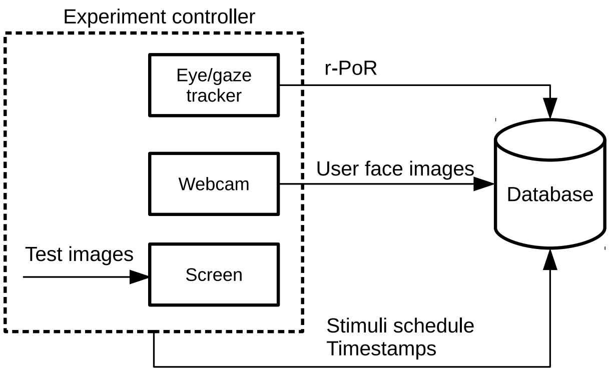

Fig. 3. Data acquisition components.

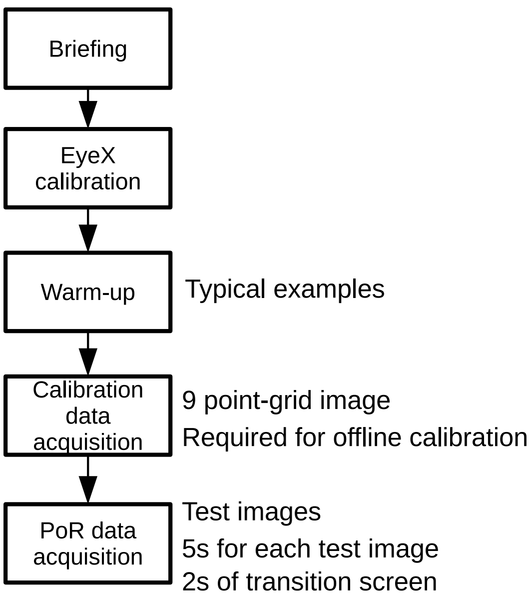

Fig. 4. Block diagram of the experimental protocol.

At the end of a data acquisition session, the yielded output comprises: user face images, reference PoRs, list of exhibited images, and time stamps for every item listed before. Such an output provides input signals (face user images) for testing ET approaches whose performance can be assessed by comparison against a ground-truth (in this case the recorded EyeX PoRs). In addition, other studies can be conducted such as identification of eye movements (saccades, targeting and fixation), calculation of heat maps, determination of scanpaths, and so on.

IV. Experimental protocol

The conceived experimental protocol is divided into five stages: a) briefing, b) eye/gaze tracker (EyeX) calibration, c) warm-up, d) data acquisition for offline calibration, and e) PoR data acquisition (Fig. 4).

The briefing is the very first step of this experiment. During this phase, information about the objectives of the experiment as well as instructions regarding the task to be performed are read to every subject. The text containing such information must be clear, simple and straight. Some time at the end of the reading is provided for making questions if any doubt remains.

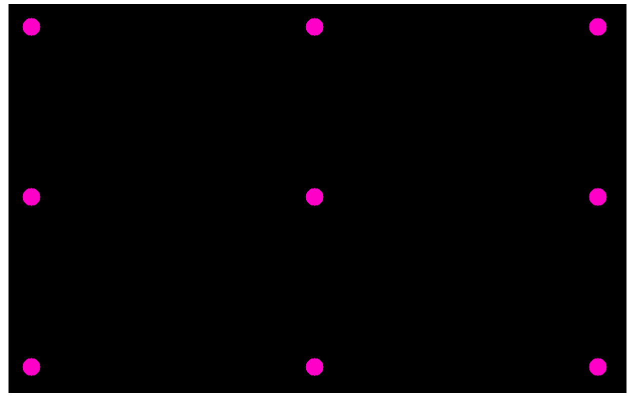

Fig. 5. 3×3 dot grid used to obtain data for offline calibration.

Fig. 6. Experimental interface.

In the second stage, the researcher runs the EyeX calibration software provided by Tobii. This step aims at optimising the eye/gaze tracking device regarding the characteristics of a specific subject.

The third stage, called warm-up, intends to familiarise the subject with the experimental interface as well as with typical stimuli to be shown in the PoR data acquisition phase.

Next, some data is collected in order to allow offline calibration of eye/gaze trackers. This phase consists in showing a 3×3 dot grid, each dot at a time for five seconds in a random sequence, until all the 9 dots are visualised by the subject (Fig. 5).

In the PoR data acquisition step, a set of 14 test images (from Toyama Image Data Base [13], [14]) is randomly shown to the subject. Each test image is kept on screen for 5 seconds, after which a grey transition screen is presented during 2 seconds. Such a design is based on *International Telecommunication Union* (ITU) recommendations (like the one in [15]) to avoid bias in results due to visualisation of particular sequences of images. The purpose of this stage is to get PoR data along with their related timestamps.

The experiment duration is about 3.5 minutes, excluding briefing phase.

V. Conclusion

The paper presents an experimental setup aimed at providing ground-truth points of regard as well as eye/gaze tracker input data, which can be later retrieved from a database for offline performance assessment of different ET+VAM (eye/gaze tracking in association with visual attention model) schemes. Details are given regarding its components, proposed protocol and usage. The PsychoPy source code that controls the whole experiment is available at https://github.com/Mateusmota/etplusvam. Currently, the experimental protocol concerning actual PoR acquisition is in final preparation to be submitted to the ethics committee of authors' university for analysis and consent.

Acknowledgment

The authors would like to thank the Scientific Initiation Scholarship programme (PIBIC) of the Federal University of Pará (UFPA), and CNPq (Science without Borders programme) for the partial financial support of this work.

References

[1] A. T. Duchowski, *Eye Tracking Methodology: Theory and Practice*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2007.

[2] Tobii AB. Tobii EyeX. [Online]. Available: http://tobiigaming.com/product/tobii-eyex/

[3] Eye Tribe Team. The Eye Tribe. [Online]. Available: https://s3.eu-central-1.amazonaws.com/theeyetribe.com/theeyetribe.com/index.html

[4] O. Ferhat and F. Vilarino, "Low Cost Eye Tracking: The Current Panorama," *Computational Intelligence and Neuroscience*, vol. 2016, no. iii, 2016.

[5] A. Borji and L. Itti, "State-of-the-srt in Visual Attention Modeling," *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 35, no. 1, pp. 185–207, 2013. [Online]. Available: http://dl.acm.org/citation.cfm?id=2412386.2412937

[6] Y. Sugano, Y. Matsushita, and Y. Sato, "Calibration-free gaze sensing using saliency maps," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, jun 2010, pp. 2667–2674. [Online]. Available: http://ieeexplore.ieee.org/document/5539984/

[7] Jixu Chen and Qiang Ji, "A Probabilistic Approach to Online Eye Gaze Tracking Without Explicit Personal Calibration," *IEEE Transactions on Image Processing*, vol. 24, no. 3, pp. 1076–1086, mar 2015. [Online]. Available: http://ieeexplore.ieee.org/document/6990593/

[8] OpenCV team. Open source computer vision library. [Online]. Available: http://opencv.org/

[9] P. Le Callet and M. Barkowsky, "On viewing distance and visual quality assessment in the age of Ultra High Definition TV," *VQEG eLetter*, vol. 1, no. 1, pp. 25–30, Mar. 2014. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01150427

[10] ITU-R Recommendation BT.500-13, "Methodology for the subjective assessment of the quality of television pictures." Geneva, Switzerland: International Telecommunication Union, 2012.

[11] J. Peirce, "Generating stimuli for neuroscience using PsychoPy," *Frontiers in Neuroinformatics*, vol. 2, p. 10, 2009. [Online]. Available: http://journal.frontiersin.org/article/10.3389/neuro.11.010.2008

[12] J. W. Peirce, "Psychopy—psychophysics software in Python," *Journal of Neuroscience Methods*, vol. 162, no. 1–2, pp. 8 – 13, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165027006005772

[13] S. Tourancheau, F. Autrusseau, Z. M. Sazzad, Parvez, and Y. Horita, "Impact of the subjective dataset on the performance of image quality metrics," in *IEEE International Conference on Image Processing 2008. ICIP 2008.*, San Diego, CA, United States, Oct. 2008, pp. –. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00321663

[14] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "Which Semi-Local Visual Masking Model For Wavelet Based Image Quality Metric?" in *IEEE International Conference on Image Processing*, San Diego, United States, Oct. 2008, pp. 1180–1183. [Online]. Available: https://hal.archives-ouvertes.fr/hal-00343502

[15] ITU-T, "Objective perceptual assessment of video quality: Full reference television," *ITU-T Telecommunication Standardization Bureau*, 2004.