Object recognition and person detection for mobile eye-tracking research. A case study with real-life customer journeys.

Stijn De Beugher, Geert Brône, and Toon Goedemé

1 Introduction

The recent development of user-friendly plug-and-play mobile eye-tracking technology has paved the way for research into visual behavior and real-life user experience in natural environments, such as public spaces, commercial environments or interpersonal communicative settings. The challenge for this new type of pervasive eye-tracking is the processing of data generated by the systems used in real-world environments [7]. Recently, several solutions to the analysis challenge have been proposed (see [2] for an overview). The best-known technique is the use of markers (infrared or natural) to predefine potential areas of interest (AOI), generating a two-dimensional plane within which eye gaze data can be collected for longer stretches of time and generalized across subjects. This paper presents an alternative to the AOI-based methods, building on recent studies combining object recognition algorithms with eye-tracking data [1] and [7].

2 Approach

By combining state-of-the-art object recognition [6] and person- [4] and face- [8] detection techniques for image processing (see figure 1), our system allows for a robust largely automatic analysis of relevant objects without the need for predefined areas of analysis or prior training. To process an eye-tracker experiment, one needs

Stijn De Beugher
EAVISE, ESAT - KU Leuven , Belgium e-mail: stijn.debeugher@thomasmore.be

Geert Brône
MIDI Research Group - KU Leuven, Belgium e-mail: geert.brone@arts.kuleuven.be

Toon Goedemé
EAVISE, ESAT - KU Leuven, Belgium e-mail: tgoedeme@esat.kuleuven.be

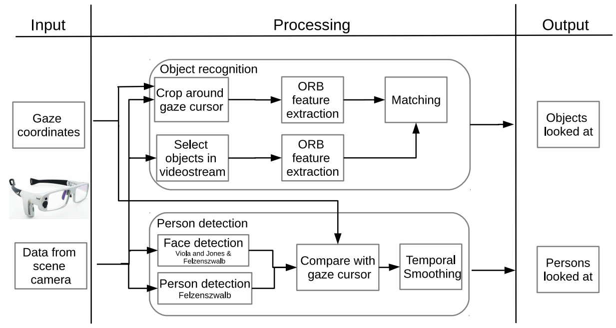

Fig. 1 High-level overview approach

Output is generated in the form of both a visual summary of the eye-tracking data and statistical information.

to select the objects of interest, by simply clicking on the objects while replaying the video of the eye-tracking experiment. The major advantage of this approach is that once the objects of interest are marked, they can be reused to process several eye-track recordings.

The next step of our automatic approach is the calculation of all fixations and fixation durations for the selected objects in the video. This is realized by searching for ORB feature [6] correspondences between the objects of interest and a region around the gaze cursor in each video frame, as shown by the green border in the left part of figure 2. At the same time, the system calculates how often and for how long one looked at another person, during this stage we make a distinction between specifically looking at a face, e.g. during talking, and looking at someone from a larger distance. In order to improve the detection-rate of faces and bodies, we applied two novel approaches. Firstly, we obtained an occlusion robust human torso detector by training a deformable part model [5] with only the top 60% of images from the VOC2009 [3] database, our model is illustrated at the right part of figure 2. The second novelty is a temporal smoothing system in which we use the gaze cursor as a tracker. This system assumes that a valid detection should stand for at least a certain time, thus preventing false detections and it allows us to solve gaps between detection sequences and therefore overcome missing detecitons.

The final step is the visualisation of our detection results. We chose to display the results on a timeline in terms of detected objects to give a chronological overview of the complete eye-tracking experiment, as shown in figure 3. The output of the detection results can be tuned through a set of parameters such as detection threshold, minimum fixation duration or the maximum gap between visual fixations.

3 Experiment

For this study, we conducted a real-life experiment to test the overall performance of our detection scheme for processing mobile eye-tracking data. In order to collect

Fig. 2 Illustration of our matching techniques. Left part of the image illustrates the object correspondences. Right part illustrates a person detection and our model for torso and upperbody detection.

representative data in a natural user environment, we selected the typical customer journey of a visitor to a museum, starting from the ticket counter all the way to the gift shop. Fourteen participants (7 male-7 female) were recorded while they visited a special exhibition at Museum M in Leuven (Belgium). Recordings were made with Tobii Glasses and Arrington Gig-E60 mobile systems and lasted between 35 and 65 minutes.

4 Results

Figure 3 shows the visual output of the detection system. Each fixation on a pre-trained object or person with a duration of at least a certain time (tunable via threshold, for example 150 ms) is displayed as a separate thumbnail of the relevant object of interest, human face or torso. In this figure we have selected 12 objects of interest, including specific works of arts, an elevator, the ticket counter, an Ipod, a route map, human torso or face, etc. In our GUI we visualize the objects of interest at the upper horizontal bar. The lower bar of images represents the actual visual fixations in the entire eye-tracker experiment. For each fixation the start and stop time is shown.

Such a visualisation may be a valuable tool to gain insights into user experience. Since we provide a summary of a complete eye-tracker experiment with respect to the viewing behaviour towards specific objects in a set of thumbnails, analyzing a customer journey is simplified. Our tool makes it possible to answer customer journey related questions such as: ”Did the participants use the elevator to enter the exhibition?”, ”How long did it take before they entered the exhibition?”, ”Did the participants notice the walking guides at the start of the exhibition?”, ”Did they notice there was an Ipod in the exhibition?”, ”How did they navigate through the different works of art (order, time spent looking at the different works, etc.)?”. Since it is possible to reuse the marked objects of interest it is possible to compare recordings of several visitors and produce more generalized statistics, as shown in table 1. These results correspond to the questions of the post-questionnaire.

At the workshop, we will discuss in more detail the results of the experiment, both in terms of precision recall curves for the techniques that were used, as well as computation time and general usability of the system.

Table 1 Questions to be answered in the context of the museum visit.

Question	Visitor 1	Visitor 2	Visitor 3	Visitor 4
Time at cash deck?	1m22s	42s	49s	20s
Looked at work of art in entrance hall?	NO	NO	NO	YES
Make use of elevator or stairs?	Elevator	Stairs	Elevator	Stairs
Time to get to exhibition?	1m43s	5m3s	1m21s	3m17s
Looked at walking guides at start of exhibition?	NO	YES	YES	NO
Looked at Ipod?	NO	NO	YES	YES
Total time at the exhibition?	28m58s	51m13s	35m3s	37m27s
Make use of elevator or stairs to get back from the exhibition?	Elevator	Stairs	Elevator	Stairs

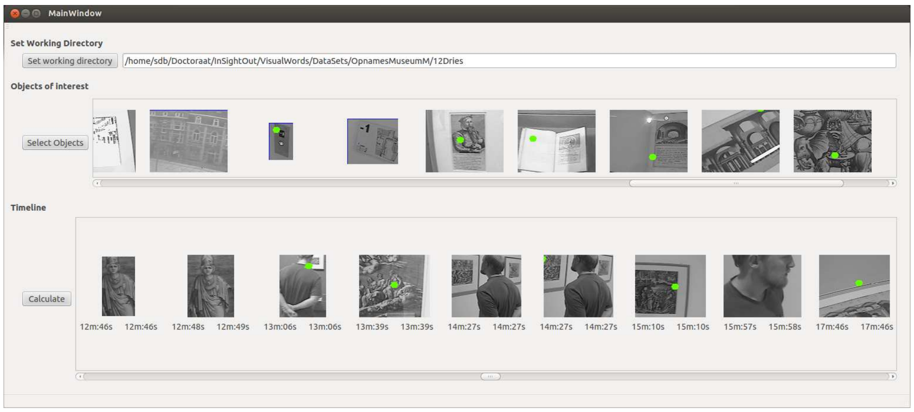

Fig. 3 Visualisation of the detection results. Top: manually selected objects. Bottom: timeline with objects and persons looked at.

References

1. S. De Beugher, G. Brône,T. Goedemé (2012). Automatic analysis of eye-tracking data using object detection algorithms. In Proc. of ACM Conference on Ubiquitous Computing pp. 677-680, Pittsburgh, USA.
2. K. M. Evans, R. A. Jacobs, J. A. Tarduno, J. B. Pelz (2012). Collecting and Analyzing Eye-tracking Data in Outdoor Environments. Journal of Eye Movement Research 5(2)6 pp. 1-19.
3. M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, A. Zisserman. The PASCAL Visual Object Classes Challenge 2009 (VOC2009) Results.
4. P. F. Felzenszwalb, R. B. Girshick, D. McAllester (2010). Cascade object detection with deformable part models. In Proc. of CVPR pp. 2241-2248, San Fransisco, USA.
5. P.F. Felzenszwalb, R. B. Girshick, D. McAllester. Discriminatively Trained Deformable Part Models, Release 4.
6. E. Rublee, V. Rabaud, K. Konolige, G. Bradski (2011). ORB: An efficient alternative to SIFT or SURF. In Proc. of ICCV 2564-2571, Barcelona, Spain.
7. T. Toyama, T. Kieninger, F. Shafait, A. Dengel (2012). Gaze guided object recognition using a head-mounted eye tracker. In Proc. of ETRA, pp. 91-98, California, USA.
8. P. Viola, M. Jones (2001). Rapid object detection using a boosted cascade of simple features. In Proc CVPR pp. 511-518, Kauai, USA.