Automatic analysis of in-the-wild mobile eye-tracking experiments

Stijn De Beugher[1], Geert Brône[2], and Toon Goedemé[1]

[1] EAVISE, ESAT - KU Leuven, Belgium
[2] MIDI Research Group - KU Leuven, Belgium

Abstract. We discuss a novel method for the analysis of mobile eye-tracking data in natural environments. Mobile eye-trackers generate large amounts of data, making manual analysis very time-consuming. Available solutions, such as marker-based analysis minimize the manual labour, however they require experimental control, making real-life experiments practically infeasible. Here, we discuss a novel method for the processing of mobile eye-tracking data based on the integration of computer vision techniques. Using such an approach allows us to automatically detect specific objects, faces and human bodies in images captured by a mobile eye-tracker. By mapping the gaze data on top of those detections, we get insights in the visual behaviour. In this paper, we briefly describe the applied image processing techniques. We also present a thorough comparison between the manual analysis and our automatic analysis in both speed and accuracy on a challenging, large-scale real-life experiment.

Keywords: mobile Eye-tracking, object detection, person detection, annotation

1 Introduction

The development of mobile eye-tracking systems has opened up the paradigm of eye-tracking to a wide variety of research disciplines and commercial applications. Whereas traditionally, the analysis of eye gaze patterns was largely confined to controlled lab-based conditions due to technological restrictions (i.c. obtrusive hardware restricting the flexibility of use and potential research questions), mobile systems allow for eye-tracking in the wild, without a necessarily predefined set of research conditions. Because of this increased flexibility, research into visual behaviour and real-life user experience now extends to natural environments such as commercial environments or to interpersonal communicative settings including research on human-human interaction.

A mobile eye-tracker is a head-mounted wearable device combining two types of cameras. The scene camera is looking forward and captures the field of view, while the eye-camera(s) on the other hand capture the eye-movements, from which the gaze direction is estimated. Output of such an eye-tracker consists of the images captured by the scene camera with the gaze-locations laid on top of them.

One of the key challenges for this type of eye-tracking is the processing of data generated by the systems. By abandoning the traditional well-controlled lab-based conditions, the data stream generated by the eye-trackers becomes highly complex, both in terms of the objects and scenes that are encountered, and the gaze data that need to be analyzed and interpreted. How can researchers avoid the painstaking task of manually coding large amounts of data, which is extremely time-consuming, without losing the full potential of mobile eye-tracking systems?

In this work we discuss an alternative to the traditional marker based analysis, building on recent studies combining several image processing techniques with eye-tracking data [1], [2] and [3]. By mapping gaze data on objects and object classes to be recognized in the scene video data, a number of restrictions of AOI-based approaches no longer hold, including the need to work with predefined static areas. This extended abstract summarizes our previous work [4] except for the person re-identification which is unpublished work.

2 Approach

We apply three different techniques to analyse the eye-tracking data. The first part of this section discusses the implementation of an object detection technique to detect how often and for how long a particular object was viewed. The second part handles the implementation of techniques to count how often and for how long a face or a person was viewed. Finally the third part discusses a method for person re-identification to identify at which particular person the subject looked at.

2.1 Object recognition

This part of our approach focuses on how we process eye-tracker data to generate basic statistics for specific objects to be detected. This is done in several steps:(i) The input images of the forward looking camera are cropped around the coordinates of the gaze cursor.(ii) In the next step the user selects objects of interest in the datastream by simply selecting them while the video is playing. (iii)In the third step our algorithm searches for correspondences between each cropped frame and object of interest frame using ORB features [5]. Then a score is calculated to each pair based on the number of matches. (iv)In a final step consecutive similar frames are clustered into a "visual fixation", which is defined as a series of images in which the same object was viewed with a minimal duration time.

2.2 Person detection

In the second part of our approach we focus on the detection of faces and human bodies as it is of great importance in for example human-human interaction analysis. By mapping the gaze date on top of face and person detections, we are able to automatically count how often and how long a person looked to another person or more specifically looked to another face. We use an own trained human upper body model in a Deformable Part Model [6] based approach for the detection of human upper bodies and a Haar-based model [7] for the detection of human faces. In order to further improve the detection rate we proposed another novelty, i.e. a temporal smoothing approach based on the gaze cursor to avoid many false positives and false negatives. Furthermore each detection is tracked using a Kalman filter.

2.3 Person re-identification

In the context of human-human interaction analysis one is traditionally interested in how often a person looks to another person. In a triad, were three persons were equipped with a mobile eye-tracker during a natural conversation, one would also like to know to which person the participant is looking. Figure 1 shows sample frames of the scene camera of one of the participants.

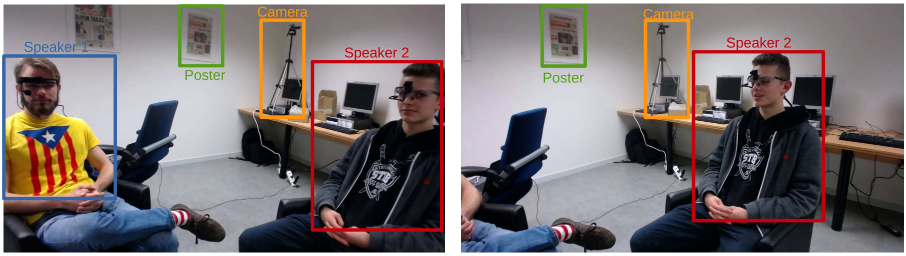
We expanded our person detection algorithm with a person re-identification step. In the example case of figure 1, it is clear that we could distinguish both persons based on the color of their clothes, to be extracted using a histogram comparison. Within one of the bounding box generated by the upper body detection, a reference histogram of each participant is calculated.

In a next step, we apply our person detector as explained above. For each frame where there is overlap between a person detection and the gaze cursor, we calculate a histogram of the detection window. By comparing this new histogram to the previously developed histograms, our system is able to identify at which person the participant was looking at. The results of this approach is briefly discussed in the next section.

3 Results

In order to validate our approach, we used our approach to automatically annotate a real-life mobile eye-tracking experiment that was conducted as part of a human-human interaction experiment. Here we automatically process the eye-tracker data of one of the three participants. This recording has a duration of 14 minutes and consists of 20568 frames. The purpose of the analysis was to investigate the visual behaviour towards four items: speaker wearing the yellow sweater, speaker wearing dark sweater, poster on the wall, one of the external recording cameras. The goal is to segment the entire recording according to when and how long our participant looked at the given items. In figure 1(a), the four object types are shown.

For validation, we asked a third person to manually assign a label to each of the automatic generated segments. Finally we compared our automatic analysis with the manual labelling to find out the level of agreement. This comparison



(a) (b)

Fig. 1: Different objects and persons that were automatically labelled. Since the camera is wearable, the scene is captured under different viewpoints.

revealed that the level of agreement is very high (97,2%), making our automatic analysis applicable for the analysis of mobile eye-tracking experiments.

Next to the accuracy, there is also a significant improvement in analysis time. The automatic analysis of the entire video of 14 minutes of video material took approximately 27 minutes of computing time. The manual allocation of labellings to the segments in ELAN, which is only a part of the entire labelling job, took about 60 minutes of manual input.

4 Conclusions

In this extend abstract we presented an approach for automatic eye-tracker data processing based on object recognition, face and person detection and re-identification. Using the analysis of a long lasting challenging eye-tracker experiment, we prove the applicability of our approach both in terms of outstanding accuracy and a significant reduction in processing time.

References

1. De Beugher, S., Ichiche, Y., et al.: Automatic analysis of eye-tracking data using object detection algorithms. In: Proc. of UbiComp. (2012) 677–680
2. Toyama, T., Kieninger, T., et al.: Gaze guided object recognition using a head-mounted eye tracker. In: Proc. of the ETRA. (2012) 91–98
3. Yun, K., Peng, Y., Samaras, D., et al.: Studying relationships between human gaze, description, and computer vision. In: Proc. of CVPR. (2013) 739–746
4. De Beugher, S., Brône, G., Goedemé, T.: Automatic analysis of in-the-wild mobile eye-tracking experiments using object, face and person detection. In: Proc. of VISAPP, Lisbon, Portugal (2014) 625–633
5. Rublee, E., Rabaud, V., et al.: Orb: An efficient alternative to sift or surf. In: Proc. of ICCV. (2011) 2564–2571
6. Felzenszwalb, P.F., Girshick, R.B., McAllester, D.: Cascade object detection with deformable part models. In: Proc. of CVPR. (2010) 2241–2248
7. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proc. of CVPR. (2001) I–511–I–518 vol.1