

# Towards a more effective method for analyzing mobile eye-tracking data: integrating gaze data with object recognition algorithms

Geert Brône, Bert Oben

Lessius Antwerpen

Dept. of Applied Linguistics

Sint-Andriesstraat 2, Antwerp, Belgium

[geert.brone@lessius.eu](mailto:geert.brone@lessius.eu), [bert.oben@lessius.eu](mailto:bert.oben@lessius.eu)

Kristof Van Beeck, Toon Goedemé

Lessius Mechelen

Dept. ICT - EAVISE

De Nayerlaan 5, St.-Katelijne-Waver, Belgium

[toon.goedeme@lessius.eu](mailto:toon.goedeme@lessius.eu)

## ABSTRACT

In this paper we present the outlines of a new project that aims at developing and implementing effective new methods for analyzing gaze data collected with mobile eye-tracking devices. More specifically, we argue for the integration of object recognition algorithms from vision engineering, such as invariant region matching techniques, in gaze analysis software. We present a series of arguments why an object-based approach may provide a significant surplus, in terms of analytical precision, flexibility, additional application areas and cost efficiency, to the existing systems that use predefined areas of analysis.

In order to test the actual analytical power of object recognition algorithms for the analysis of gaze data recorded *in the wild*, we develop a series of test cases in different real world situations, including shopping behavior, navigation, handling and usability of mobile systems. By setting up these case studies in close collaboration with key players in the relevant fields (retailers, signage consultants, market and user-experience research, and developers of eye-tracking hard- and software), we will be able to sketch an accurate picture of the pros and cons of the proposed method in comparison to current analytical practice.

## Author Keywords

Mobile eye-tracking, automatic gaze analysis, object recognition, real world research

## ACM Classification Keywords

H.5.2. User interfaces / user-centered design, I.4 Image processing and computer vision, I.4.8 scene analysis

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*UbiComp '11*, Sep 17–Sep 21, 2011, Beijing, China.

Copyright 2011 ACM 978-1-60558-843-8/10/09...\$10.00.

## INTRODUCTION

The recent development and commercial availability of mobile eye-tracking devices (such as the SMI IView X™, Tobii Glasses and Mangold Mobile Eye) has generated a broad range of potential applications for studying gaze behavior in a natural environment (e.g. [2], [8]). Hard- and software innovations allow for easy-to-use, unobtrusive tracking solutions beyond the traditional boundaries of lab-controlled conditions. The exponential increase in flexibility in comparison to static (e.g. screen-based) eye-tracking systems, however, also involves a significant additional cost, most notably so in the efficiency of analysis. Whereas data generated on a predefined two-dimensional plane (as in the case of a screen-based system) allow for a robust (semi-)automatic statistical analysis across subjects and conditions, the lack of such a fixed frame (and thus fixed area of analysis) presents a significant challenge for the analyst confronted with a highly complex data stream.

In order to reduce the substantial cost of manual video data annotation of gaze behavior, which may make mobile eye-tracking applications economically unfeasible, a number of solutions have been developed and implemented. One such solution that has been adopted in several commercially available systems (including e.g. Tobii Glasses) is the use of infrared-based markers to predefine one or more potential areas of analysis (AOA) or more fine-grained areas of interest (AOI) (cf. figure 1).



Figure 1. AOA based analysis of gaze data

In this system, physical markers are attached to specific areas in a research site to define a two-dimensional plane,

which allows for automatic mapping of eye-tracking data, fast data aggregation, (semi-)automatic statistical analysis and presentation of results (gaze plots, heat maps, etc.).

### Limitations of AOA marker systems

Although working with AOAs enables researchers to quickly and efficiently tackle specific questions on the visual distribution in predefined planes, the method has a number of limitations:

- (1) potentially relevant areas of analysis need to be defined *before* testing takes place. Any relevant observations outside these fixed zones still need to be processed manually;
- (2) tracking multiple fields or objects with identical/similar features (i.e. object categories such as price tags in a supermarket or traffic signs) requires the attachment of markers to each individual token;
- (3) an AOA is a two-dimensional plane on which gaze behavior can be mapped. Determining the distribution of visual attention between *individual objects* within an AOA still requires significant manual labor;
- (4) AOA systems do not fully do justice to the flexibility of mobile eye-tracking systems, as the objects of interest need to be tied to a *fixed position* in the AOA in order to allow for a semi-automatic analysis. Handling of objects outside of the AOA again requires manual annotation work.
- (5) Large natural test environments such as e.g. entire supermarkets or museums require the installation of a huge number of IR beacons, which is very expensive and therefore not economically feasible.

The related ASL Gazemap method is a commercially available technique designed to counter the above limitations. However, although natural markers are detected with vision techniques, the method is still very AOA-oriented and limited to a small environment.

### PROPOSED TECHNIQUE

In order to overcome most of the limitations of AOA-based eye-movement analysis, we propose a complementary method that takes objects and object categories rather than areas as the basic analytical layer on which gaze data are mapped for analysis. An object recognition algorithm automatically analyses the video stream coming from the eye-tracker's scene camera. The algorithm can be trained to recognize specific objects at the image region around the gaze coordinate. By using this method, the system can automatically build statistics about what objects the test person is looking at. An illustration is given in figure 2.

The potential benefits of such an approach are:

- (1) a target of analysis is not restricted to a predefined and fixed AOA, but rather can be any object (category), stationary or in motion, that enters the visual field;



Figure 2. Object recognition illustration in shopping setting

- (2) potentially relevant objects need not be determined in advance, as gaze data collected by a mobile eye-tracker can serve as input stream for an off-line object recognition module;
- (3) the distribution of visual attention can be analyzed highly accurately and with a minimum amount of manual preparation or analysis (coding and processing), and the results of these analyses can be represented in novel ways. One such format is what we call an *object cloud* (by analogy with word clouds), which shows the objects that 'caught the eye' most in a complex scene.

### Implementation of object recognition algorithms

The object recognition algorithms to be used in this system need to meet a series of requirements. Most importantly, they need to be robust against changing viewpoints, partial occlusions and changing illumination conditions. We propose the application of the family of so-called *invariant region matching techniques* that define interest regions in an image, based on specific features of the image content. These regions are described with descriptor vectors that are invariant to changes in illumination and viewpoint. The use of descriptor vectors allows for a fast processing of potential correspondences between the visual content of multiple pictures. Figure 3 illustrates this process.

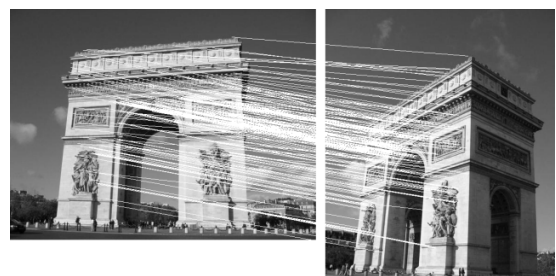


Figure 3. Corresponding regions in pictures with a highly different viewpoint, detected using [1]

Many object recognition algorithms based on this technique have been proposed. A survey is given in [9], while [5] and [6] report comparative experiments. Although the most widespread technique is David Lowe's SIFT algorithm [4], we opt to use Herbert Bay's Speeded Up Robust Features (SURF) [1], which has a substantially faster execution time with a comparable performance as compared to SIFT. If the

objects do not have sufficient texture or show in-class appearance variance, one can choose for robust classification techniques such as Felzenszwalb *et al.* [3].

In order to reduce the calculation load of the algorithm, we propose a system that takes as input for the analysis not the full video stream but rather an established perimeter around the visual focus (gaze cursor). The visual content in and around the focus of attention is then compared to one or more objects that have been trained using example pictures. When the correspondence threshold is reached, the system automatically records the positive recognition and the overlap with the gaze data.

### **Object selection: Training-by-looking-at**

The recognition of correspondences across objects obviously presupposes that algorithms are trained for relevant objects. This is traditionally done by manually collecting training images, marking object boundaries and feeding them into the system. We propose a simplified alternative to that process for the mobile eye-tracker with object recognition by introducing the innovative training step *training-by-looking-at*. Instead of using a separate image collection process, we propose to use the eye-tracking device to train the system: test subjects are asked (before or after recording) to look at one or more relevant objects, and the resulting output is used to calculate invariant regions. These invariant regions then build the basis for recognizing objects in the video stream. We will develop an experimentally validated procedure for this, to ensure that all viewpoints of complex objects are covered without including the background.

### **PROOF-OF-PRINCIPLE TESTING**

One of the key challenges of our project is to proof the relevance and efficiency of an object-based eye-tracking method in different test settings and applications. For that purpose, we have developed a series of proof-of-principle case studies in different real world situations, each with their own challenges and requirements. Among the fields that have an interest and proven track record in applying eye-tracking data are (mobile) user interface design, market research, branding & packaging design, signage consultancy and (traffic) safety, HCI and others.

In order to get an accurate picture of the pros and cons of our proposed method, the case studies are set up in close collaboration with organizations and companies that have (extensive) experience in applying (mobile) eye-tracking for specific research questions. More specifically we have set up joint projects with a total of 11 partners in different fields: retailing (e.g. one of Belgium's largest retailers), consultancy and training, market research, user experience and brand design, the cultural sector (e.g. a major museum), and development (one of the global market leaders in the development of eye-tracking technology).

The case studies are set up in such a way that they cover a broad spectrum of applications linked to the above fields. In

order to get a maximum of input and feedback, the partners are actively involved in setting up, conducting and evaluating the studies. More specifically, as the tests aim at uncovering the potential surplus of an object-based approach in comparison to an IR-based AOA analysis, each of the tests is run in parallel using the different methods.

In this paper, we report on three proof-of-principle studies that are currently being designed and conducted.

### **Case study 1: complex visual settings**

Existing mobile eye-tracking systems using an IR-based AOA analysis are most frequently applied to complex informational scenes with different elements competing for visual attention. The prototypical example are shelves in a supermarket, where positioning, packaging, pricing information etc. may all influence a shopper's distribution of attention, and by extension, his/her choices. Manufacturers, retailers and package designers are interested in singling out the various factors that influence consumer behavior, and eye-tracking is one method to study these factors.

The experiment consists of two conditions, one with predefined AOAs and one with object recognition software. In the first condition, the experimenters (who are, as mentioned, collaborators from the project partners) are given instructions to install IR-markers at several areas in a real supermarket. Each of the conditions figures 10 test subjects, who are first calibrated and then sent into the supermarket with a specific shopping list (featuring product types rather than brands). In the second condition, after the actual shopping instruction the subjects are shown a series of products that also appeared in the predefined AOAs in condition 1. The input of this additional step is used as training material for the object recognition tool in condition 2. After recording, the experimenters are instructed to collect and analyze the data and present the main results in the form of gaze plots, heat maps or object clouds.

After the conclusion of the experiment, both conditions are evaluated on the basis of the above-mentioned parameters, which are partly objective (time efficiency) and partly user-oriented (user-friendliness, flexibility, quality). The results of the comparative study build the basis for a first systematic pro-and-con-analysis of both methods for the specific setting of complex visual scenes with multiple competing elements.

### **Case study 2: changing test conditions**

One of the main challenges for pervasive eye-tracking is maintaining a reliable data recording in changing conditions, including sudden changes in illumination, rapid movement of objects or referent point (e.g. head movements), etc. In order to test the robustness of the system with object-recognition software, we designed a second proof-of-principle study in collaboration with a selection of the above-mentioned project partners.

The case study consists in a navigation task in a large setting with variable conditions (inside and outside, variable illumination conditions inside, etc.). Test subjects are instructed to navigate from point A to point B with the help of signposts. As in the first study, one group of participants is tested using the IR-based system and one group with the object recognition system (with the sign(s) being trained before recording). In comparison to the first experiment, the test is not primarily designed to test time efficiency in setting up both conditions, as it is clear that attaching IR markers to a set of identical signs is more time-consuming and expensive than training this sign for recognition. Rather, the test is designed to compare both systems for robustness in variable conditions and amount of manual post-processing needed. For both systems, the degree of successful sign recognition is calculated in the different measuring situations of the test in order to arrive at a general assessment of their robustness.

### Case study 3: moving objects and scenes

Mobile eye-tracking systems equipped with IR-based AOA-analysis software are designed to be applied to static areas so as to allow for the automatic rendition of internal distributional information within those areas (which parts of the covered field are focused most often, and in which order?). They are not particularly suited for a naturalistic study of gaze behavior in a more dynamic setting, with (fast-)moving scenes and objects, as discussed in [6].

One such setting is the distribution of visual attention within a moving vehicle. When driving in a car, relevant information, ranging from billboards to traffic signs, flashes by quickly. Setting up a naturalistic study for such a context using mobile eye-tracking technology requires a system that can process the multitude of incoming information (cf. study 1) in vastly changing circumstances (cf. study 2). The goal of the third proof-of-principle study is to test the reliability of our object-recognition system for analyses involving moving stimuli (from the perspective of the experience) with changing background conditions. As was the case for the second study, the main focus is not on the time efficiency of our system in comparison to an AOA-approach, with an IR-marker method being practically unfeasible when collecting data for a range of objects. Rather, we primarily aim to test the robustness of both systems in terms of degree of successful recognition.

As in the previous two studies, we have two test conditions (AOA system vs. object recognition system), with 10 test subjects for each condition. The test subjects are taken along a predefined route as a passenger in a car. They are instructed to pay attention to traffic signs and billboards along the way. In the AOA conditions, the experimenters attach IR markers to a selection of these signs. In the object recognition condition, these signs are trained into the system. The processing of recorded data and the resulting

comparison of both conditions runs parallel to the procedure presented for the first study.

### CONCLUSION

Although much of the work discussed in this note is still in progress, the general architecture of the project presented may be of relevance to the burgeoning paradigm of pervasive eye-tracking. Most notably, it was shown that the implementation of robust and flexible object recognition algorithms may significantly improve and simplify the analysis of real-world visual behavior through eye-tracking. By designing a series of proof-of-principle tests in close collaboration with organizations that apply eye-tracking to real-life situations on a regular basis (or have a substantial interest in it), we will provide a detailed, user-based assessment of the pros and cons of an automated object-based eye movement analysis.

### REFERENCES

1. Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L.. SURF: Speeded up robust features. *Computer Vision and Image Understanding* 110, 3 (2008), 346-359.
2. Duchowski, A.T. *Eye Tracking Methodology*. Springer, London, 2007.
3. Felzenszwalb, P., Girshick, R., McAllester, D., Cascade Object Detection with Deformable Part Models, *IEEE (CVPR)*, 2010
4. Lowe, D.. Distinctive image features from scale invariant keypoints. *International Journal of Computer Vision* 60, 2 (2004),91-110.
5. Mikolajczyk, K. and Schmid, C. A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence* 27,10 (2005),1615-1630.
6. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T. and Van Gool, L. A comparison of affine region detectors. *IJCV* 65, 1/2 (2005), 43-72.
7. Papenmeier, F. and Huff, M. DynAOI: A tool for matching eye-movement data with dynamic areas of interest in animations and movies. *Behavior Research Methods* 42, 1 (2010), 179-187.
8. Tegenkvist, A. Case study: It's all in the eyes. *Research Live Magazine*, Feb 2011 [<http://www.research-live.com/magazine/case-study-its-all-in-the-eyes/4004622.article>]
9. Tuytelaars, T. and Mikolajczyk, K. Local Invariant Feature Detectors - Survey. *Trends in Computer Graphics and Vision* 3,1 (2008), 1-110.