Towards Location-Aware Mobile Eye Tracking

Peter Kiefer, Florian Straub, Martin Raubal
Institute of Cartography and Geoinformation
ETH Zurich, Switzerland
{pekiefer, straubf, mraubal}@ethz.ch

Abstract

This paper considers the impact of location as context in mobile eye tracking studies that extend to large-scale spaces, such as pedestrian wayfinding studies. It shows how adding a subject's location to her gaze data enhances the possibilities for data visualization and analysis. Results from an explorative pilot study on mobile map usage with a pedestrian audio guide demonstrate that the combined recording and analysis of gaze and position can help to tackle research questions on human spatial problem solving in a novel way.

CR Categories: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems — Evaluation/methodology; I.2.1 [Artificial Intelligence]: Applications and Expert Systems — Cartography;

Keywords: mobile eye tracking, location-awareness, pedestrian navigation study, outdoor

1 Introduction and Motivation

Context-aware mobile computing (Abowd et al., 1999) strives to offer services on a mobile device that support the user within a specific context, with location being the most prevalent type (location-awareness). This line of research in mobile computing does not focus much on the hardware perspective of mobility, but on the implications of varying location context.

Similar to these discussions in mobile computing we discuss in this paper the impact of location as context in mobile eye-tracking (MET) studies that extend to large-scale spaces (Kuipers and Levitt, 1988). Compared to static eye tracking (SET, also referred to as "remote eye-tracking"), MET methods "provide a good solution for studying perception with the freedom of movement and variable contexts that characterize natural vision" (Franchak et al., 2010, p.21).

Our example in this paper, an outdoor pedestrian wayfinding study, is taken from a spatial cognition research context. While there exist measures to rate people's performance in a wayfinding task, MET data could additionally help to understand *why* mobile decision processes fail in certain situations. Did the participant look at street signs? Did she use a mobile map at critical decision points? Understanding these cognitive processes could help us engineer better mobile services (Raubal, 2009). Other domains in which the location context of an individual plays an important role include car navigation (Sodhi et al., 2002) and indoor wayfinding (Schuchard et al., 2006).

A combined recording, visualization, and analysis of position and gaze seems to be a valuable and novel approach to tackle many of these research questions, but a discussion on the integration of location in MET is currently lacking in the scientific literature.

In this paper, we argue that a hardware-oriented interpretation of the adjective "mobile" in "Mobile Eye Tracking" limits our perspective considerably. We emphasize the mobility of the participant, and less the mobility of the device. Section 2 discusses the implications of combining the participant's location with gaze-overlayed videos. We introduce *location-aware mobile eye tracking* (LA-MET) and describe the conceptual differences to related work in MET. New combined measures on gaze and position are introduced that can be used to analyze LA-MET data. In a first explorative pedestrian navigation study we combine GPS (Global Positioning System) positioning with a gaze-overlayed video (section 3). The study illustrates that location can indeed add valuable information to MET studies. The paper ends with a conclusion and an outlook on the next steps in LA-MET (section 4).

2 Location-Aware Mobile Eye Tracking

With *location-aware mobile eye tracking* (LA-MET) we refer to the combined tracking of an individual's gaze (typically recorded as a gaze-overlayed video) and her 2D or 3D position in a spatial reference system (determined through indoor or outdoor positioning technology).

2.1 Related Work

LA-MET hardware must be mobile, but a head-mounted system is not required. The dashboard-mounted eye tracker used by Jensen et al. (2010) in a real world driving experiment, for instance, is mobile in that sense. Using a head-mounted eye tracker, on the other hand, does not necessarily imply that the experiment can be regarded as LA-MET. Cauchard et al. (2011) use a head-mounted system to evaluate mobile user interfaces, but the participants never change their position. Participants changed their position in several studies, but the position was not tracked (e.g., Franchak et al., 2010; Schuchard et al, 2006; Sodhi et al., 2002). We may assume that Jensen et al. (2010) recorded GPS positions in their car navigation study although they do not explicitly state this.

To our knowledge, MET data and location data have not been combined for analysis. This is especially remarkable as there has been plenty of research on the analysis and interpretation of motion tracks at the intersection of geographic information science (GIScience) and artificial intelligence (see Kiefer, 2011,

---

chapter 3 for an overview), as well as work on mobile activity recognition in the MET community (Bulling, 2010). A first workshop indicates the growing interest in pervasive eye tracking (Bulling et al., 2011), yet none of the papers addresses location and spatio-temporal data analysis.

In GIScience and cartography, eye tracking studies have gained interest for the evaluation of how users interact with a geographic information system (Çöltekin et al., 2010), or how they perceive a map (Opach and Nossum, 2011). Although these studies are related to geoinformation processing, they are lacking a mobile context. Their aim consists in analyzing *how* people work with geographic information, but not *where*.

2.2 User Position and Gaze Position

The unique advantage of LA-MET is that it allows the inclusion of two positions for analyses, the user's position and the gaze position (the location the user looks at). The availability of the user's position enhances the possibilities for analyzing eye tracking data notably. Consider, for instance, a one-dimensional temporal plot tracing the moments in which a mobile map is gazed upon, i.e., an area of interest (AOI) sequence chart with one AOI. Given this gaze data, we can perform analyses in the temporal dimension, such as *duration of map usage*, *map usage frequency* etc., as used in SET studies. However, for tasks that require locomotion it is often more important to know *where* an AOI gaze happened than at what time. Task durations may also differ significantly between participants due to different route choices, making a direct comparison in the temporal dimension difficult.

With the addition of the user's position a variety of spatial analyses methods can be applied. The most basic data structure is the *AOI-annotated locomotion track*, which is a sequence of user positions over time, each annotated with the AOI (if any) of the respective gaze position. Figure 1b shows the AOI-annotated locomotion tracks of five participants performing the same pedestrian navigation task. Red indicates the AOI "paper map". Further geo visualization methods can reveal regularities in the data even more clearly, such as the point density visualization in Fig. 2 (same data as Fig. 1b). We discuss these data in section 3.3. Another example of a new measure is the *average AOI locomotion speed*, defined as the average locomotion speed over all moments with the participant's gaze at a given AOI (also exemplified in section 3.3). Sequence analyses, as another group of measures, can also profit from the user's position data. AOI strings known from SET (e.g., ABCAB) become *geo-referenced AOI strings*, combining AOIs and regions of stay, e.g. (map, BellevueNorth) (street-sign, BellevueNorth) (noAOI, BellevueCenter). Sequences such as these provide a powerful framework for representing the outcome of spatial problem solving processes, such as pedestrian navigation.

We can characterize LA-MET approaches with respect to the coordinate system of their gaze position: type 1 LA-MET approaches determine the gaze position only for a virtual coordinate system overlayed on a mobile object (e.g., a map). Type 2 LA-MET approaches additionally determine the gaze position in 3D coordinates of the environment, which can then be mapped to a real-world object (e.g., a street sign). Manual annotation of gaze positions for type 2 is labor-intensive, while an automated computation requires head tracking and/or image processing. A discussion of these technologies (and their restrictions regarding accuracy etc., especially outdoors) is beyond the scope of this paper. However, all measures introduced in this section so far can be used for both types of LA-MET.

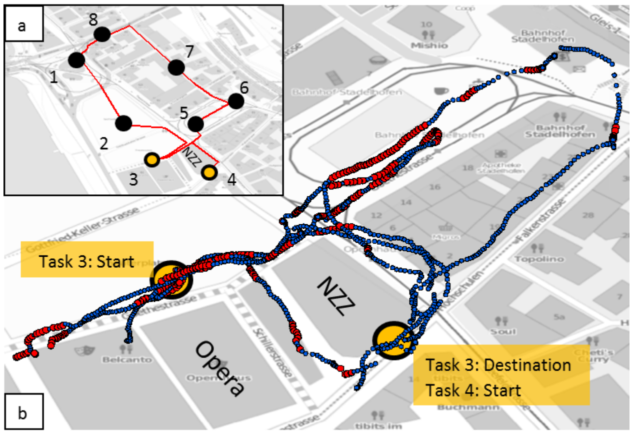

Figure 1 *a) Overview of the study area with optimal path. b) Detailed view of task 3 area with AOI-annotated GPS tracks for five participants (AOI "paper map")*

2.3 Location-Aware Attentive Interfaces

Another dimension used to describe LA-MET approaches is determined by whether the combined gaze and location data are recorded for later analyses or directly processed with the purpose of giving the user immediate feedback. The first option was implicitly assumed in the previous sections and will be exemplified in section 3.

The second option leads to attentive interfaces (Vertegaal, 2002). By combining location and eye tracking data mobile applications can be built that react to the user's (assumed) cognitive processes. Consider, for instance, a mobile tourist audio guide telling you to "look at the cathedral to the left", then checking whether you are spotting the correct building, and if necessary provide a corrective audio command. In terms of section 2.2, this example would be categorized as type 2 LA-MET because it utilizes both user and gaze location. A type 1 location-aware attentive interface would react to the user's gaze only when she looks at a mobile device, such as a tablet PC showing the user's position on a map.

3 The Zurich Audio Guide: An Explorative Pilot Study

We performed an explorative pilot in which LA-MET was used to evaluate audio instructions for human navigation. The goal was to identify issues that would arise in a LA-MET study. In particular we wanted to explore the usefulness of the location context, and determine if the measures provide new information for analyzing human problem solving behavior.

In section 3.1 we describe the set-up of the pilot study, section 3.2. reports on the collected data. Section 3.3 shows selected data and analyses.

3.1 Study Set-Up

We used parts of the podcast “Geld und Geist”[1], an official audio guide of the city of Zurich, Switzerland (available only in German). The original podcast consists of 76 audio tracks that provide historical information on the most important sights. Navigational instructions included in these tracks lead a tourist through Zurich. We picked nine tracks from the beginning, which together form a round trip. Figure 1a shows the route when all navigation instructions are followed correctly. All background information was eliminated, leaving only the navigational instructions. The resulting nine audio tracks had an average duration of 20 seconds each. Three types of navigational instructions were provided: cardinal (compass) directions, landmarks (Lovelace et al., 1999), and references to previously visited places. An example (task 3, translated from German, refer to Fig. 1b):

“Our next destination is the old NZZ building at the intersection ‘Theaterstrasse’/’Falkenstrasse’. The building is a direct neighbor of the Opera. The entrance is close to the tram station ‘Opernhaus’ and facing South towards the ‘Seefeld’ quarter.”

Participants were instructed to navigate to the place described in each audio track. They were given a paper map showing a Google Maps^{TM} screenshot of the same rectangular area as the map in Fig. 1a. The first audio track was a trial task in which participants got acquainted to the hardware and the reactions of passersby. The recording started with the second audio track (task 1). Each participant was accompanied by two researchers, one responsible for keeping an eye on traffic safety, the other for taking notes. Neither of them provided any orientation cues. As soon as a participant thought she had reached the right place, she informed the accompanying researchers who then took notes concerning success or failure. Participants could also give up on a task. In any case, participants were then lead to the pre-defined starting position of the next task.

3.2 Collected Data

We had five participants aged 20, 25, 26, 38, and 67; three were male, two female. Three were university students, one translator, one former teacher. Each of them was presented all eight tasks. The data collected between two tasks was excluded from the analysis.

We recorded for each participant a gaze-overlaid video and a GPS motion track. The gaze was recorded with the Ergoneers Dikablis Cable system[2], a head-mounted video-based monocular eye tracker with 50 Hz recording frequency. Data was stored on a notebook which participants carried in a backpack. GPS position was tracked at a frequency of one latitude/longitude coordinate per second using a standard Android smartphone (Samsung Galaxy S II) attached to the backpack. Collected gaze data was post-processed manually by using the Dikablis tools to handle parts of the video where pupil detection failed due to lightness conditions. This was possible as the Dikablis system records the videos of both, the field and the pupil camera.

Figure 2 *Point density highlighting the spatial positions of users while looking at the mobile paper map.*

Task 3	P1	P2	P3	P4	P5
Task Completed Correctly	No	No	No	**Yes**	No
Average AOI locomotion speed (m/s)	0.51	0.62	0.71	**0.10**	0.55
Task Duration (s)	535	329	433	**216**	201

Table 1 *Average AOI locomotion speed and task performance.*

Visual markers at the edges of the map were used to compute gaze behavior with respect to an AOI “paper map”. A calibration for the distance of the paper map was performed to compensate for the parallax error. The different recording frequencies of gaze and position were accounted for by mapping 50 frames of the gaze video to one GPS position. Each GPS position was annotated with “true” if the user’s gaze hit the “paper map” AOI in the majority of the 50 gaze frames. The fact that information on map gazes taking less than one second may be lost was not a problem for this study because we were not interested in such short gazes.

Thus, the post-processed LA-MET data consisted of triples (latitude, longitude, AOI?) with AOI? $\in$ {true, false}, i.e., AOI-annotated locomotion tracks (visualized in Fig. 1b).

3.3 Results

As the study was an explorative pilot, the following results are rather anecdotal. We only present results for task 3 in which participants had to walk from the entrance of the opera to the NZZ (“Neue Züricher Zeitung”, a Swiss newspaper) building entrance (see Fig. 1b). Figure 2 shows a heat map of the five tracks where red colors indicate positions where the accumulated time of map usage over all participants was high. The figure illustrates that, besides the starting position, three more positions show intensive map use. A possible hypothesis for this behavior may be that ‘audio instructions that are not based on turn-by-turn descriptions cause an increased necessity for additional map usage at cross roads’.

As a second measure we computed the average AOI locomotion speed (with respect to the AOI “paper map”), as introduced in section 2.2. The general idea behind this measure is that there may be people that prefer reading the map while standing, whereas others keep moving while using the map. Values for the

---

[1] In English: “money and intellect”
http://www.stadt-zuerich.ch/vbz/de/index/frei

http://www.stadt-zuerich.ch/vbz/de/index/frei
zeit_events/vbz_podcasts/geld_und_geist.html

[2] http://www.ergoneers.com/

average AOI locomotion speed in our small study ranged from 0.10 m/s to 0.71 m/s with an average of 0.50 m/s (refer to Table 1). Participant P4 obviously differs from the other participants: he reduced his speed, or was even standing still while reading the map, whereas others kept on walking. For each participant we assessed the task correctly performed if they chose the route described in the audio track while at the same time reaching the correct end position. All routes that did not approximately follow the red route from Figure 1a were considered incorrect, more specifically, all routes that did not surround the NZZ building by turning right into "Seefeldstrasse", and then again right into "Falkenstrasse". The results in Table 1 indicate that reducing speed while using the map is the better strategy. Interestingly, although P4 reduced the speed while using the map, he had at the same time the second lowest duration until task completion (215 seconds, see Table 1). It is also notable that all participants used approximately one third of their time on map usage. A possible hypothesis using the average AOI locomotion speed may correlate task performance and preferred speed while using the map.

4 Conclusions and Future Work

The paper has argued for a combined recording and analysis of gaze and (geo-spatial) user position in studies that extend to large-scale spaces. Location-aware mobile eye-tracking helps to analyze not only what an individual is gazing at, but also *from where* she is gazing. A pilot pedestrian wayfinding study has been used to demonstrate how this approach can be used to tackle research questions on spatial problem solving in a novel way.

For future work, we intend to perform a complete study for the Zurich audio guide, tackling some of the questions and hypotheses that were brought up here. One challenge will be to cope with the GPS inaccuracy occurring in urban areas (and present in our data, see Fig. 1b). We will also include data from participants' self-estimation of their sense of direction, such as the Santa Barbara Sense of Direction Scale (Hegarty et al., 2002)[3]. Automatically computing the 3D gaze position for type 2 LAMET (see section 2.2) will also be considered as a next step to add other AOIs than "paper map". Additional measures for LAMET need to be developed. Gaze data could be used to recognize activities (Bulling, 2010), such as "visually searching", and combined with locomotion-based activities, such as "running". Sequence based analyses could include such information. On a higher semantic level, the behavior data could be interpreted as intentions (Kiefer, 2011), leading to *attentive intention-aware services*.

References

ABOWD, G., DEY, A., BROWN, P., DAVIES, N., SMITH, M., & STEGGLES, P. 1999. Towards a Better Understanding of Context and Context-Awareness. In H.-W. Gellersen (Ed.), *Handheld and Ubiquitous Computing* (Vol. 1707, pp. 304-307): Springer Berlin / Heidelberg.

BULLING, A. 2010. *Eye Movement Analysis for Context Inference and Cognitive-Awareness: Wearable Sensing and Activity Recognition Using Electrooculography.* (PhD Thesis), ETH Zürich, Zürich. (19082)

BULLING, A., DUCHOWSKI, A. T., & MAJARANTA, P. 2011. *Proceedings of the 1st international workshop on pervasive eye tracking & mobile eye-based interaction*. Beijing, China: ACM.

CAUCHARD, J. R., LÖCHTEFELD, M., IRANI, P., SCHOENING, J., KRÜGER, A., FRASER, M., & SUBRAMANIAN, S. 2011. *Visual Separation in Mobile Multi-Display Environments.* 24nd annual ACM symposium on User interface software and technology (UIST 2011).

ÇÖLTEKIN, A., FABRIKANT, S. I., & LACAYO, M. 2010. Exploring the efficiency of users' visual analytics strategies based on sequence analysis of eye movement recordings. *International Journal of Geographical Information Systems, 24*(10), 1559--1575.

FRANCHAK, J. M., KRETCH, K. S., SOSKA, K. C., BABCOCK, J. S., & ADOLPH, K. E. 2010. *Head-mounted eye-tracking of infants' natural interactions: a new method.* Symposium on Eye-Tracking Research and Applications (ETRA 2010), Austin, Texas.

HEGARTY, M., RICHARDSON, A. E., MONTELLO, D. R., LOVELACE, K., & SUBBIAH, I. 2002. Development of a self-report measure of environmental spatial ability. *Intelligence, 30*(5), 425--447. doi: 10.1016/S0160-2896(02)00116-2

JENSEN, B. S., SKOV, M. B., & THIRURAVICHANDRAN, N. 2010. *Studying driver attention and behaviour for three configurations of GPS navigation in real traffic driving.* 28th international conference on Human factors in computing systems (CHI 10), Atlanta, Georgia, USA.

KIEFER, P. 2011. *Mobile Intention Recognition.* (PhD Thesis), Springer, New York.

KUIPERS, B., & LEVITT, T. 1988. Navigation and Mapping in Large-Scale Space. *AI Magazine, 9*(2), 25--43.

LOVELACE, K., HEGARTY, M., & MONTELLO, D. 1999. Elements of Good Route Directions in Familiar and Unfamiliar Environments. In C. Freksa & D. Mark (Eds.), *Spatial Information Theory. Cognitive and Computational Foundations of Geographic Information Science* (Vol. 1661, pp. 65--82). Berlin / Heidelberg: Springer.

OPACH, T., & NOSSUM, A. 2011. *Evaluating the Usabilty of Cartographic Animations with Eyemovement Analysis.* 25th International Cartographic Conference, Paris.

RAUBAL, M. 2009. Cognitive Engineering for Geographic Information Science. *Geography Compass, 3*(3), 1087--1104.

SCHUCHARD, R. A., CONNELL, B. R., & GRIFFITHS, P. 2006. *An environmental investigation of wayfinding in a nursing home.* Symposium on Eye tracking research & applications (ETRA 2006), San Diego, California.

SODHI, M., REIMER, B., COHEN, J. L., VASTENBURG, E., KAARS, R., & KIRSCHENBAUM, S. 2002. *On-road driver eye movement tracking using head-mounted devices.* Symposium on Eye tracking research & applications (ETRA 2002), New Orleans, Louisiana.

VERTEGAAL, R. 2002. *Designing attentive interfaces.* Symposium on Eye tracking research & applications (ETRA 2002), New Orleans, Louisiana.

---

[3] A self-estimation was actually collected in the current Zurich audio guide study, but the data was not used in the analysis due to the small number of participants.