Starting to get bored: An outdoor eye tracking study of tourists exploring a city panorama

Peter Kiefer[1], Ioannis Giannopoulos[1], Dominik Kremer[2], Christoph Schlieder[2], Martin Raubal[1]

[1]Institute of Cartography and Geoinformation
ETH Zurich
Stefano-Franscini-Platz 5
CH-8093 Zurich, Switzerland
{pekiefer,igiannopoulos,mraubal}@ethz.ch

[2]Chair of Computing in the Cultural Sciences
University of Bamberg
An der Weberei 5
D-96047 Bamberg, Germany
{dominik.kremer,christoph.schlieder}@uni-bamberg.de

Abstract

Predicting the moment when a visual explorer of a place loses interest and starts to get bored is of considerable importance to the design of touristic information services. This paper investigates factors affecting the duration of the visual exploration of a city panorama. We report on an empirical outdoor eye tracking study in the real world with tourists following a free exploration paradigm without a time limit. As main result, the number of areas of interest revisited during a short period was found to be a good predictor for the total exploration duration.

CR Categories: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Input devices and strategies; I.5.4 [Pattern Recognition]: Applications

Keywords: prediction, outdoor study, visual exploration, mobile eye tracking

1 Introduction

Tourism has been characterized as "a leisure activity which presupposes its opposite, namely regulated and organized work" [Urry 2002] (p. 2). Human geographers studying touristic decision processes have paid special attention to a mode of travel, called cultural tourism when the traveler is primarily interested in the lifestyle of the people living in a place and their history [Burns et al. 2010]. Those travelers are free to choose which places to visit and have complete control over their time management.

Despite the freedom of choice for the individual, the distribution of the interest of cultural tourists is far from uniform at the aggregate level. At a typical tourist destination, a few places are visited by masses while most places have only few visitors [Girardin et al. 2009]. A characteristic power law relationship holds, with the most photographed place attracting easily more than double of the photographs compared to the place ranking second [Balomenou and Garrod 2010; Schlieder and Matyas 2009]. From studies of GPS tracked tourists it is also known that the interestingness of places measured by the number of photographs can show a different rank ordering than the interestingness measured by the duration of the visits [Schlieder and Kremer 2011]. However, the spatial and temporal resolution of monitoring methods based on GPS tracking and photographing activity is not sufficient to gain a better quantitative understanding of this process of losing interest in a place.

Being able to predict the moment when a visual explorer of a place loses interest and starts to get bored is of considerable importance to the design of touristic information services. Mobile services such as touristic recommender systems [Ardissono et al. 2012] aim at interacting with their users in the least disruptive way possible. Eye tracking is known to provide a valuable data source for attentive user interfaces [Vertegaal 2003]. A recommendation service which displays its information on a head-mounted display to the tourist should give the user sufficient time to freely explore the panorama before starting to deliver hints such as "Did you notice the castle on the hill?"

In this paper, we investigate the following research question: *can we predict the stop of visual exploration of a city panorama based on a tourist's gaze?* With a short overview on related work we underline the novelty of our research question (section 2). We describe a mobile eye tracking study of tourists visually exploring a city panorama without a time limit (section 3). Based on the 12 participants we analyzed so far, our data indicate that the number of revisits until 57 seconds is a good predictor for the total exploration duration (see section 4 for complete results). The results are discussed in section 5 and we conclude our work in section 6.

2 Related Work

In the social sciences the concept of boredom is often associated with the concept of leisure [Conrad 1997]. Within a certain socio-cultural context, boredom can be framed as a mismatch between experience and expectations [Haldrup 2004]. In congruence, Conrad [1997] reports that repetition and the lack of control of temporal decision-making are possible sources of boredom. Current research on touristic behavior in human geography usually follows a qualitative, interpretative approach [Burns, et al. 2010; Haldrup 2004]. Quantitative data is obtained mainly by surveys. A recent study of visual exploration by tourists mentions eye tracking as a methodical approach, but uses photos taken from urban space for tracking the tourist's gazes [Stoetzer 2010].

Previous prediction models in eye tracking approach the question of *where* somebody will direct her attention. This can, for instance, be modeled with Markov models [Gordon and Moser 2007] or with saliency models [Parkhurst et al. 2002]. In this

Preprint version. Final version will soon appear in ACM DL.

Figure 1 *Study set-up (left). Panorama, as seen through the scene camera (right), overlaid with markers (red) and AOIs (blue).*

paper, we pose a novel question: can we predict when the participant will *stop looking* at the stimulus? Prediction models exploit spatio-temporal regularities in gaze data. Machine learning offers an automated and objective way of finding these regularities [Bednarik et al. 2012]. Here, we learn a very simple prediction model: a linear regression.

Attentive interfaces [Vertegaal 2003] can, in principle, be designed with explicit or implicit interaction [Schmidt 2000]. During explicit interaction, the user intentionally fixates at a certain position with the goal of triggering an interaction. Implicit interaction records the user's gaze during regular interaction and, at some later time, uses this information to adapt to the user's needs [Ajanki et al. 2009; Giannopoulos et al. 2012]. This paper contributes to future implicit gaze-based interaction concepts as it provides new insights into the prediction of when the observer of a panorama will lose interest. This moment of losing interest could be a good moment for offering assistive information.

3 Study Set-Up and Procedure

Our study was conducted in October 2012 on a hillside near the city center of Bamberg, Germany. The terrace garden of the "Michaelsberg" monastery provides a nearly 360° panorama consisting of different parts of the medieval town of Bamberg as well as park elements. Tourists arriving at the vantage point were asked whether they were willing to participate in a touristic eye tracking study. Only persons not wearing glasses or heavy eye make-up were considered. All participants were first time visitors to Bamberg on their first day who had just arrived at the vantage point. In total, 15 persons participated in the experiment. 2 of them had to be dropped due to calibration issues. A third participant cancelled the study after 39 seconds, resulting in 12 valid datasets (6 males, 6 females). All participants were German speaking, from Germany or Switzerland. 7 tourists had an age between 30 and 65, 4 participants were younger than 30, one older than 65.

The used hardware consisted of the Dikablis system by Ergoneers (http://www.ergoneers.com/index_EN.html, head-mounted, monocular, 25 Hz). Data were transferred via cable to a recording laptop, placed on a box beside the participant (see Figure 1, left). A Mexican sunhat reduced the interference of sunlight with the infrared illuminator of the eye tracker. Each participant was placed 2 meters away from a balustrade. We used the calibration procedure provided by the Dikablis software with the participant not seeing the main panorama. The parallax effect did not constitute a problem, as all objects in the city panorama were farther away than 10 meters.

Each participant was then instructed to face the panorama and given the following instruction: "Please, start looking at the city panorama as long as you like. Tell us immediately when you have finished." Visual exploration times were on average 170.58 +/- 112.0 seconds. Figure 1 (right) shows the view from the participant's position, as recorded by the scene camera.

4 Analysis and Results

4.1 Data Processing

For each of the 51,175 frames of the eye video, the pupil detection was checked manually and corrected when necessary. 8 Areas of Interest (AOIs) in the city panorama were identified by a human geographer familiar with Bamberg (see Figure 1, right). These AOIs were defined w.r.t. visual markers. Two types of markers were used: paper markers attached to the balustrade (see Figure 1, left), and so-called "key markers" (the red tetragons in Figure 1, right). A key marker is a manually drawn tetragon around distinctive features in the image. A visual flow algorithm in the Dikablis software helps following the marker for the following frames. The key markers allow for the definition of AOIs around objects which are distant from the paper markers. A point-in-polygon operation then determined for each gaze the AOI hit (at most one AOI per gaze). We assign "NOAOI" to gazes without an AOI hit.

We computed fixations using a minimum fixation duration of 200 milliseconds [Widdel 1984] and a radius of 30 pixels. Each of the resulting fixations was assigned to the AOI which most frequently occurred in the originating gaze sub-sequence. It is not trivial to compute fixations from gaze in mobile eye tracking due to smooth pursuits and the freedom of head movements. Smooth pursuits did not occur in our case because there were no moving objects in the city panorama. Head movement can cause loosing fixations when focusing on an object and moving the head in parallel. This is only relevant for us when gaze is located on an AOI. In most cases, the head movement will be found as one starting fixation in the AOI and one end fixation in the AOI, which is no problem for our further analysis.

4.2 Results: Total Exploration Time

As a first step, we analyzed the correlation between the amount and intensity of visual input processed during the total visual exploration, and the duration. All results are based on our current sample size of n=12 participants. We consider three hypotheses:

Hypothesis 1: *The larger the field of exploration, the longer it takes the tourist to explore the city panorama.*

The field of exploration is defined by the total angle φ that was explored from the participant's location ($0° \leq \varphi \leq 360°$). The rationale behind this hypothesis is that those participants deciding to explore more from the panorama are visually saturated later in time than those participants who restrict themselves to a small portion of the visual field. Our data support this hypothesis: a linear regression revealed a positive correlation between the total exploration field and the total exploration duration (R^2=0.58, $p<0.05$).

Hypothesis 2: *The more AOIs are visited, the longer it takes the tourist to explore the city panorama.*

An AOI visit occurs when the participant has an uninterrupted sequence of one or several fixations within one AOI. The analysis was performed using the absolute number of visited AOIs, i.e., counting each AOI once. Thus, this value ranges between 0 and the maximum number of AOIs (8 in our case). This hypothesis follows the same idea as hypothesis 1, with the difference that visual saturation is this time measured by meaningful AOIs rather than by the angle. Again, our data support this hypothesis: a linear regression of the total number of visited AOIs and the exploration duration revealed a positive correlation (R^2=0.55, $p<0.05$).

Hypothesis 3: *The more AOI revisits occur, the longer it takes the tourist to explore the city panorama.*

An AOI revisit occurred each time the participant visited an AOI that had already been visited before. Hypothesis 3 was found to be the one hypothesis supported strongest by our data: a linear regression of the exploration duration and the number of AOI revisits revealed a strong positive correlation (R^2=0.8, $p<0.05$).

4.3 Results: A Predictive Model

In a second step, we considered whether the three measures used in section 4.2 can be utilized to predict when a tourist will stop exploring a panoramic view. The model should be able to predict the total duration of exploration, given the eye tracking data up to a certain time $t_{predict}$. The time $t_{predict}$ was chosen heuristically as the shortest exploration duration in our data ($t_{predict}$ = 57 sec).

A linear regression of the total exploration duration (dependent variable) and total number of AOI revisits that occurred during the first 57 seconds (independent variable) revealed a negative correlation (R^2=0.34, $p<0.05$), see Figure 2. Participants who performed more AOI revisits during the first 57 seconds of the panorama exploration had a shorter total duration. Interestingly, this correlation is negative, whereas the corresponding correlation over the total time is positive (see section 4.2).

We compared the predictive accuracy of "AOI revisits at 57 seconds" with that of potential other models. A linear regression of the total exploration duration and the field of exploration during the first 57 seconds did not reveal any correlation (R^2=0.006, p=0.8). Similarly, a linear regression of the total exploration duration and total AOIs visited during the first 57 seconds did not yield any correlation (R^2=0.0, p=0.9).

A further analysis revealed that, while "AOI revisits until 57 sec" is found to be a good predictor, "average dwell times until 57 sec" is not (R^2=0.000004, p = 0.9). This is surprising, because these two themselves are correlated (although not significantly): a linear regression of the average dwell times during the first 57 seconds (dependent variable) and total number of AOI revisits that occurred during the first 57 seconds (independent variable) revealed a negative correlation (R^2=0.2957, p = 0.06).

5 Discussion and Future Directions

In the previous section we reported on preliminary results of a tourist eye tracking study indicating that it is possible to build a predictive model for the stop of visual exploration of a city panorama based on the number of AOI revisits up to a certain time $t_{predict}$. It is clear that more work needs to be done before these results can be generalized. Only one place was considered, and the number of participants did not allow us to investigate individual or group differences, such as cultural background, age, and gender, or contextual factors, such as weather and time of the day. This is clearly a central issue for future work.

The duration of the observation period of 57 seconds was chosen heuristically as the shortest observed total exploration time. It is possible that the optimal value for $t_{predict}$ is different for each individual. However, our goal was to develop a model that generalizes over individuals as good as possible, since often knowledge about the individual is not available. We did not analyze whether a prediction would be possible for longer time spans, such as 90 or 120 seconds. The rationale behind this is that a gaze-based assistant should be able to make its predictions as early as possible. Increasing the number of participants will allow us to select $t_{predict}$ on a stronger empirical basis.

From a cognitive science point of view, it may be worthwhile to discuss how the correlation described in section 4.3 can be explained. First of all, it is unclear whether every participant truly stopped the visual exploration due to visual saturation

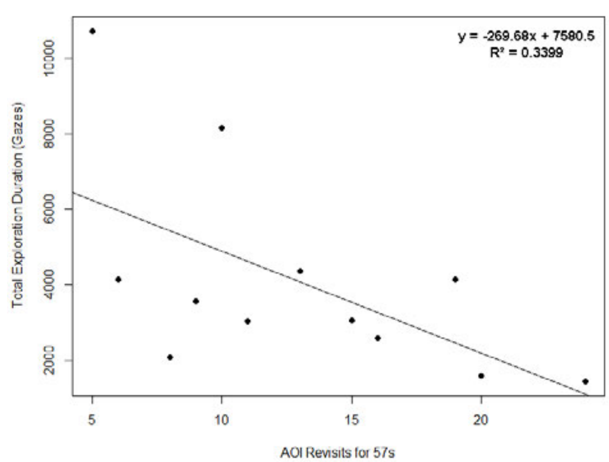

Figure 2 *Linear regression with number of AOI revisits during the first 57 sec as independent, and total exploration duration as dependent variable.*

and/or boredom. All 12 participants were very relaxed, taking their time after the experiment for a chat about their travel plans. Still, we cannot totally exclude the possibility that time pressure, such as felt expectations from travel mates, influenced their behavior. The experimental conditions of wearing an eye tracker in public space might constitute another influence on their behavior. However, it is reasonable to assume that our setting was much more realistic than compared to a laboratory setting. Our participants were actual tourists in exactly the usage context a potential gaze-aware assistant would be designed for. The study was seamlessly integrated into their overall tourist experience. A related methodological question is concerned with controllability. We tried to account for controllability as much as possible. For instance, the experiment was performed on two subsequent days with comparable weather conditions. Still, external factors, such as the exact position of clouds or passersby in the yard could not be totally equalized.

As future work, we would like to investigate whether it is possible to build more accurate predictive models by integrating multiple criteria. We will also consider the prediction of boredom at any time during exploration from significant changes in the gaze measures. Finally, the usability and social acceptance of gaze-based tourist assistants depends on a number of factors not covered in this research so far.

6 Conclusion

In this paper we investigated a novel research question: can we predict the moment of stop in a visual free exploration task? We reported on a real world eye tracking study of 12 tourists exploring a city panorama. Results indicate that the amount of visual input and intensity of visual exploration are strongly correlated with the exploration duration. Most interestingly, a negative correlation between the number of revisits during a short observation period and the total exploration duration was found. Although these results are constrained by the small number of participants, they point towards possible future directions in intelligent gaze-based tourist assistance.

References

AJANKI, A., HARDOON, D., KASKI, S., PUOLAMÄKI, K., and SHAWE-TAYLOR, J., 2009. Can eyes reveal interest? Implicit queries from gaze patterns. *User Modeling and User-Adapted Interaction 19*, 4 (2009/10/01), pp. 307-339.

ARDISSONO, L., KUFLIK, T., and PETRELLI, D., 2012. Personalization in cultural heritage: the road travelled and the one ahead. *User Modeling and User-Adapted Interaction 22*, 1-2, pp. 73-99.

BALOMENOU, N. and GARROD, B., 2010. Using Volunteer-employed Photography: Seeing St David´s Peninsula through the Eyes of Locals and Tourists. In P. Burns, et al. Eds: *Tourism and Visual Culture, Volume 2: Methods and Cases*, CAB International, Wallingford, UK, pp. 111–119.

BEDNARIK, R., VRZAKOVA, H., and HRADIS, M., 2012. What do you want to do next: a novel approach for intent prediction in gaze-based interaction. In *Proceedings of the Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, California 2012), ACM, pp. 83-90.

BURNS, P., PALMER, C., and LESTER, J., 2010. Tourism and Visual Culture CAB International, Wallingford, UK.

CONRAD, P., 1997. It's Boring: Notes on the Meanings of Boredom in Everyday Life. *Qualitative Sociology 20*, 4 (1997/07/01), pp. 465-475.

GIANNOPOULOS, I., KIEFER, P., and RAUBAL, M., 2012. GeoGazemarks: Providing gaze history for the orientation on small display maps. In *Proceedings of the 14th International Conference on Multimodal Interaction (ICMI '12)* (New York, NY, USA2012), ACM, pp. 165-172.

GIRARDIN, F., VACCARI, A., GERBER, A., BIDERMAN, A., and RATTI, C., 2009. Quantifying urban attractiveness from the distribution and density of digital footprints. *International Journal of Spatial Data Infrastructure Research 4*, pp. 175–200.

GORDON, P.C. and MOSER, S., 2007. Insight into analogies: Evidence from eye movements. *Visual Cognition 15*, 1, pp. 20-35.

HALDRUP, M., 2004. Laid-Back Mobilities: Second-Home Holidays in Time and Space. *Tourism Geographies 6*, 4 (2004/11/01), pp. 434-454.

PARKHURST, D., LAW, K., and NIEBUR, E., 2002. Modeling the role of salience in the allocation of overt visual attention. *Vision Research 42*, 1, pp. 107-123.

SCHLIEDER, C. and KREMER, D., 2011. Visiting the Same Place but Seeing Different Things: Place Models of Touristic Behavior. In A. Henrich, et al. Eds: *Visibility in Information Spaces and in Geographic Environments - Post-Proceedings of the KI'11 Workshop*, Fakultät WIAI, University of Bamberg, Bamberg, Germany, pp. 15--21.

SCHLIEDER, C. and MATYAS, C., 2009. Photographing a City: An Analysis of Place Concepts Based on Spatial Choices. *Spatial Cognition & Computation 9*, 3 (2009/08/28), pp. 212-228.

SCHMIDT, A., 2000. Implicit human computer interaction through context. *Personal Technologies 4*, 2-3, pp. 191-199.

STOETZER, S., 2010. Using technology in Visual Analysis of Identificational Strategies. In P. Burns, et al. Eds: *Tourism and Visual Culture. Vol. 1: Theories and Concepts*, pp. 76-89.

URRY, J., 2002. *The tourist gaze*. Sage Publications, Thousand Oaks, CA.

VERTEGAAL, R., 2003. Attentive User Interfaces. *Communications of the ACM 46*, 3, pp. 31-33.

WIDDEL, H., 1984. Operational Problems in Analysing Eye Movements. In G.G. Alastair and J. Frank Eds: *Advances in Psychology*, North-Holland, pp. 21-29.