Towards Identifying User Intentions in Exploratory Search using Gaze and Pupil Tracking

Thomas Low
Otto von Guericke University
Magdeburg, Germany
thomas.low@ovgu.de

Nikola Bubalo
General Psychology
Ulm University, Germany
nikola.bubalo@uni-ulm.de

Tatiana Gossen,
Michael Kotzyba
Otto von Guericke University
Magdeburg, Germany
firstname.lastname@ovgu.de

André Brechmann
Leibniz-Institute for Neurobiology
Magdeburg, Germany
brechmann@lin-magdeburg.de

Anke Huckauf
General Psychology
Ulm University, Germany
anke.huckauf@uni-ulm.de

Andreas Nürnberger
Otto von Guericke University
Magdeburg, Germany
andreas.nuernberger@ovgu.de

ABSTRACT

Exploration in large multimedia collections is challenging because the user often navigates into misleading directions or information areas. The vision of our project is to develop an assistive technology that is able to support the individual user and enhance the efficiency of an ongoing exploratory search. Such a technical search aid should be able to find out about the user’s current interests and goals. Respective parameters can be found in the central and in the peripheral nervous system as well as in overt behavior. Therefore, we aim at using eye movements, pupillometry and EEG to assess respective information. Here, we describe the set-up and the first results of a preliminary user study investigating the effects of searching an image collection on eye movements and pupil dilations. First data show that numbers of fixation, fixation durations as well as pupil dilations differ systematically when looking at a subsequently selected target as compared with not selected items. These results support our vision that further research additionally investigating EEG can in fact result in better predicting the searchers goals and next choices.

Keywords
Exploratory Search; Gaze Tracking; Multimedia Retrieval; Pupilometry;

1. INTRODUCTION

Multimedia systems offer a lot of information. However, retrieval of a piece of information can be challenging and/or exhausting and hence exploration in large image collections is not an easy search task. It is a highly dynamic process of a user interacting with an image retrieval system that requires learning about the collection structure and/or content. In addition there is no clear stop criterion at the beginning of the search. The process ends when the user decides that their information need is satisfied [4].

For a user searching a huge multimedia collection, technical assistive systems could aid in improving search efficiency. Users often have no exactly defined target in mind. Instead, they search within a certain field of interest. Here, reducing the amount of possible next actions - ideally only one interaction - would be helpful. In order to suggest such next steps during a search process, the assisting function must be informed about the interests and goals of the searcher. This requires investigating the details of the users’ search behavior.

This paper describes the projects vision and first steps and results into the development of such an assistive system, which is investigating user behaviors from a psychological and a neurobiological perspective, and modeling covert and overt reactions for predicting user intentions from a computer science perspective. Before detailing our approach in Section 3, Section 2 provides an overview of suggested approaches using methods such as eye tracking, pupillometry, or EEG. Section 4 describes our setup used to examine gaze parameters during search, and first results concerning gaze and pupil tracking are reported. In the final section, our work is summarized and compared to the related work.

2. RELATED WORK

One important question in exploratory search is which observable variables can be used to anticipate user intentions during the search process and, therefore, help to guide them towards a successful goal fulfillment. Because of the implicit nature of eye control, using gaze data is a promising approach as implicit relevance feedback [11, 12].

Li et al. [8] used fixation duration to identify a users’ implicit relevance feedback and with that, improve the performance of content-based image retrieval systems. Their results indicate that eye tracking based relevance feedback improves image retrieval performance.

Golenia et al. [3] combined EEG and eye tracking data in order to disambiguate image search results that are obtained
using ambiguous textual queries. They were able to make an eye tracking and EEG based prediction that is above chance level.

Nevertheless, there seems to be still one important aspect of information missing. Given that the research described above already utilizes gaze behavior and EEG, one might suggest that information about the activation in the peripheral nervous system might additionally contribute to predicting a user’s intention. Therefore, pupil dilation might be an important and available additional source of information. In fact, pupil size differs not only depending on brightness but also depending on cognitive or affective load [13]. Therefore, the pupil size has been suggested as further information channel for gaze-based interaction [1, 2].

Summarizing the research above, the combination of methods such as eye tracking or EEG is a promising approach to improve the search performance and hence the user experience. The optimal way would be to simultaneously observe behavior, central cognitive processes as well as peripheral arousal.

Following Li et al. [8], we decided that gaze might serve as a fruitful tool investigating the users’ interests. As a central physiological variable, EEG will be added to our measures to directly observe the user’s monitoring process during exploratory search [14]. Here error-related negativity (ERN) potentials will be informative about wrong decisions of subjects and feedback related negativity (FRN) will reflect whether the users expectation on the feedback is met. Moreover, spectral analysis of EEG data may be used to track attentional demand or the users’ engagement during the search process [6, 10].

The goal of the current work is to identify which eye tracking parameters are suited for identifying what kind of images a user is interested in while performing an exploratory search. For a well-defined search task in web retrieval Gwizdka and Zhang [5] reported that eye tracking parameters such as fixation durations and pupil diameter differed significantly between observed websites depending on whether they were relevant or irrelevant to the users search objective. We thus aim at enlarging their data to a different scenario of exploratory search.

3. METHOD

We conducted a lab-based experiment in which users were required to navigate through a collection of abstract images in order to find a specified target image. For the study we developed a tool called NEMP (Neighborhood Explorer using MDS and Procrustes Analysis) which is a web-based tool for navigation in large image collections [9]. NEMP projects a set of images to a two-dimensional map and groups them based on their similarities in a high-dimensional feature space through multidimensional scaling (MDS) [7]. That means that images, which are more similar to each other are placed closer together. The target image is displayed in the top left corner of the window (see Fig. 1). For users to navigate towards the target image in multiple steps it is not reasonable to show all of the images of the collection at once: It would be difficult for users to analyze such a large amount of information. Therefore, only $k$ images are shown at once. The users can navigate through the images in the collection by selecting one of the images they found to be closest to the target in its feature constellation. This selection triggers a generation of a new map containing the previously selec-

Figure 1: NEMP screenshot. 30 polygons are arranged in the map-like interface. The left part of the interface serves to control the user study, i.e. select the target image and the initial view.
required to fill out a short questionnaire inquiring about their subjective user experience with NEMP.

4. DATA ANALYSIS AND RESULTS

In mean, our users searched for 10.8 maps before they found the target. The standard deviation (SD) was 10.23 maps. The mean time spend on a map was 6.66 seconds with a SD of 6 seconds. Considering possible learning effects caused by increasing familiarity we analyzed these variables also separately for each trial. The average number of maps visited during a trial initially increased by 28.77% from the first trial (9.73, SE = 2.21) to the second trial (12.53, SE = 3.92) but decreased over the following trials to 91.1% of the initial page visits (see Fig. 2). Multiple comparisons between the trials revealed no significant differences in frequencies (p < 0.05) for all comparisons. The average time spend on a map differed significantly between the first trial (8.7 seconds, SD = 8.1) and the remaining trials (p < 0.05) which were below 7.3 seconds in average page duration (see Fig. 3).

As a next step, we analyzed the fixations on the polygon stimulus which was chosen on each map as the next polygon of interest and compared their frequency and durations with those fixations placed on the polygons which were not selected. The former will be referred to as the target polygon while the latter will be referred to as distractors. On average users fixated the target polygon of a page 3.1 times, with a SD of 1.4. These fixations lasted for 231 milliseconds on average (SD = 58.44). The accumulated number of fixations on distractors averaged to 19.1 (SD = 10.2) which means that a large number of the 29 distractors was not fixated at all. This results in 0.66 fixations per distractor. Fixation durations on distractors were on average 233 milliseconds. As ANOVAs with repeated measures show, users fixated significantly more often (F(1,14) = 96.03, p < 0.001, η² = 0.87) but not longer (F(1,13) = 0.48, p = 0.5, η² = 0.04) onto the respective target than on the individual distractors. All other main and interaction effects were not of significance for both the number of fixations (Trials: F(4,56) = 0.6, p = 0.66; Stimuli * Trials: F(4,56) = 0.46, p = 0.77) and the fixation durations (Trials: F(4,52) = 1.62, p = 0.18; Stimuli * Trials: F(4,52) = 0.29, p = 0.88). Separately for the trials, the average number of fixations on the target stays significantly higher (p < 0.001) compared to those on distractors (see Fig. 4 and 5).

Further, pupil sizes were collected during fixations on targets and distractors. There were significant differences in pupil diameter between looking at the target polygon and distractor polygons (F(1,13) = 8.67, p = 0.011). More specifically, average pupil diameters were larger for the target polygons than for the distractors at the third (p = 0.007) and fifth trial (p = 0.015). Assuming that activation decreases with increasing familiarity one would expect pupil size to decrease with increasing number of trials. This can be observed for both the targets and distractors on a descriptive level. However, looking at the dynamics, pupil diameters were significantly larger (p < 0.05) in the fourth and fifth trial compared to the second trial (see Fig. 6).
5. SUMMARY AND CONCLUSION

In order to identify useful parameters for intention detection in exploratory search tasks, the number of fixations, their durations, and the pupil sizes during fixations were compared between stimuli which were subsequently selected by users (referred to as targets) and those which were not selected (referred to as distractors). Supporting results from well-defined searches [5], we observed differences in all parameters but the fixation duration: In mean, targets were more frequently fixated than distractors. During fixations, pupils were larger suggesting more mental effort. The dynamics for these five trials showed differences from the very first search relative to all others: In the first search, users looked longer onto one map. This resulted in fact in less maps they had to inspect. Hence, although this strategy was shown to be effective, users reduced the amount of time spent on one map, by the cost of needing more maps. That is, obviously, they did not regard this strategy as efficient and thus, changed it. Interestingly, in this first trial, fixation duration on distractors was longer than on the target. Hence, one might assume that users at the beginning thought more about potential targets, whereas in the subsequent trials they decided earlier. Pupil sizes during fixations also demonstrated to be sensitive to intentions: Pupils during fixating a target were larger than during distractor fixations. This difference was of comparable size in all trials. Hence, one might suggest that pupil size increases denote some decision processes. Nevertheless, on should not over-interpret pupil sizes since the absolute sizes were relatively small. Hence, floor effects cannot be ruled out. Taken together, the results show that finding sensitive parameters in exploratory search is possible. Moreover, combining information about fixations and pupillometry seems to be a promising approach for the development of technical assistance in exploratory search. In future research we plan to further analyze correlations between individual features of images and eye tracking data.

6. ACKNOWLEDGMENTS

This work was supported by the Transregional Collaborative Research Centre SFB/TRR 62 “A Companion-Technology for Cognitive Technical Systems” funded by the German Research Foundation.

7. REFERENCES