Universität Bielefeld

CITEC
Cognitive Interaction Technology
Center of Excellence
Bielefeld University

Proceedings of the

SAGA

Solutions for Automatic Gaze-Data Analysis

Workshop 2013
@ CITEC, Bielefeld University
Bielefeld, Germany

SFB 673
Alignment in
Communication

Dr. Thies Pfeiffer, Dr. Kai Essig (Eds.)

Proceedings of the First International Workshop on Solutions for Automatic Gaze-Data Analysis 2013 (SAGA 2013)

Printed on October 23rd, 2013

This is a paper-preprint of the online proceedings of the SAGA Workshop 2013. The online proceedings will be available as OpenAccess under the following URL: http://biecoll.ub.uni-bielefeld.de/ and the DOI: 10.2390/biecoll-saga2013_0

Editors

- Dr. Thies Pfeiffer
- Dr. Kai Essig

Address

Center of Excellence Cognitive Interaction Technology (CITEC)
Bielefeld University
Inspiration 1
33615 Bielefeld
Germany

E-Mail: saga@eyemovementresearch.com
WWW: http://saga.eyemovementresearch.com

Workshop Organizers

- Dr. Kai Essig
- Dr. Thies Pfeiffer
- Jun.-Prof. Dr. Pia Knoeferle
- Prof. Dr. Helge Ritter
- Prof. Dr. Thomas Schack
- Prof. Dr. Werner Schneider

Team of Student Organizers Jasmin Bernotat, Jill Hagenbrock, Kai Harmening, Jacqueline Hemminghaus, Matthias Orlikowski, Jan Pöppel, and Katja Wenske

Acknowledgment

We are thanking the Center of Excellence Cognitive Interaction Technology (CITEC) and the Sonderforschungsbereich 673 (SFB 673) Alignment in Communication for funding and hosting the SAGA 2013 Workshop.

Many thanks to all the participants, who gave live to this workshop with their contributions!

Contents

Foreword

Background

Eye movements have been primarily measured in the domain of computer-based experiments on desktop-computer systems. With progress in technology, the early head-mounted eye tracking systems for desktop computers have been partly replaced by remote units which offer some freedom for the participants and are much less obtrusive. Such a desktop setup is easy to control, provides solid grounds for repeatable experiments and offers absolute knowledge about timecourse and locations of the presented stimuli. These are ideal conditions for rigid experimental designs where large numbers of repeatable items are tested on reasonably large numbers of participants. In some domains, the number of participants easily exceeds 100 and more. For such experiments, the time required for data cleaning, preprocessing, and annotating events of interest is a practical issue that has to be well considered. Desktop-based eye-tracking solutions directly provide information about the recorded gaze in coordinates of the computer screen and, with the help of an adequate software to implement the experiments, provide the fixated areas of interest in no time. Many systems also offer a programming interface, that allow researchers to bring eye tracking to new domains. The available solutions thus already have the potential to significantly reduce the time needed for preprocessing and annotation, letting the researcher focus on the analysis of the recorded data. Increasing the number of participants in a desktop-based experiment is thus primarily a question of time needed for the recording, but does not have a significant impact on the effort needed for the analysis.

With technical progress, eye-tracking systems have become smaller and, finally, mobile. Today, eye movements can be recorded on-the-go during shopping in large malls, when playing soccer games, or when driving cars, both in indoor and outdoor conditions.

The analysis of such mobile eye-tracking studies, however, requires much more effort. First of all, the participants are typically moving around and turn their heads in a much larger area than covered with the previous desktop-based experiments. That means that the spatial area of the stimulus domain is typically much larger. If the participant is not only turning but also relocating, the stimulus may no longer be reduced to a 2D plane as in many computer-based experiments, but its extension and position in 3D has to be considered.

In addition to that, with head-mounted mobile eye-tracking units, the gaze information is no longer provided in terms of coordinates on the stimulus, but in coordinates relative to a video recorded by a scene camera. This means that the link between the fixated pixel on the video and the stimulus beneath that pixel has to be drawn by other means and in most of the cases this requires manual annotations by humans. The movements of the participants also introduce a second issue: in the computer-based setup, every participant perceives the same stimulus material in exactly the same timing and from a reasonably similar perspective. This simplifies the aggregation of the collected data. In mobile experiments with a freely moving perspective, this is in general no longer true. Not any two participants will ever see the same stimulus from the same perspective at the same time. And we have not yet talked about dynamic stimuli that are moving as well, such as other cars in the driving situation, other shoppers in the mall, or the other players in a match of soccer.

These and other issues lead to a significant increase in pre-processing and annotation time required when migrating from desktop-based to mobile experiments. In addition to the increased effort required to analyze the individual recordings, the number of participants will also have to be increased in many cases, simply because mobile experiments are not as highly controllable as their desktop-based counterparts. Many questions may thus not be appropriately addressed with reasonable effort. Research may thus not be able to make fully advantage of the recent technologies for mobile eye-tracking, simply because of the effort required to prepare the collected data for analysis.

Motivation for the SAGA Workshop 2013

The SAGA Workshop contributes to the current international discussion, which was intensely present e.g. at this years European Conference on Eye Movements 2013: Should we migrate from desktop-based experiments in the laboratory into the "wild" of the real world? What kind of research would benefit from this migration? And if so, how can we support this migration and address the issues described in the previous section?

In recent years, several studies have substantiated the question of whether findings from desktop-based experiments under laboratory conditions transfer to the real world, e.g., when addressing social interaction or implicit knowledge as in sports or other bodily activities. These are essential issues that are in particular in-line with the research at the Center of Excellence for Cognitive Interaction Technology (CITEC) and the Sonderforschungsbereich SFB 673 Alignment in Communication at Bielefeld University, for example when cognitive interaction technology is developed and evaluated in the laboratory which will finally be transferred to robotic platforms that are to operate in real-world households.

At the same time, we observe an increase in technical systems that make use of eye movements for interaction purposes, either explicitly using gaze-gestures or by observing eye movement patterns to infer current activities (reading, walking, etc.) or stages in decision processes. Providing the missing link between gaze and objects of interest in real-time would significantly change the way we are able to interact with machines in the future.

With the SAGA Workshop 2013 we aim at bringing together basic research, applied research, and industry partners to exchange thoughts, identify reservations, riscs, and challenges, explore new ways, and, ideally, shape a joint vision on how to step out into the wild.

Thies Pfeiffer & Kai Essig

Part I

Session 01: Gaze Analysis in Basic Research

DOI: 10.2390/biecoll-saga2013_1

In Pfeiffer and Essig (2013): Proceedings of the First International Workshop on Solutions for Automatic Gaze-Data Analysis 2013 (SAGA 2013). DOI: 10.2390/biecoll-saga2013_0

A Pilot Study on Eye-tracking in 3D Search Tasks

Ekaterina Potapova, Valsamis Ntouskos, Astrid Weiss, Michael Zillich, Markus Vincze and Fiora Pirri

1 Introduction

Eye-tracking is an important step in the evaluation of computational visual attention models for comparison to human visual perception. There exist two types of visual attention: bottom-up and top-down attention. Bottom-up attention is driven by object properties, while top-down attention takes place when there is a specific search task at hand. There exist several datasets for bottom-up attention, less experiments were made to collect eye-tracking for specific search tasks, such as driving a car by Borji *et al.* [1]. In this work we are interested in collecting statistically valuable eye-tracking data for visual search tasks that are typical in daily life and of great interest to researchers in robotics and HRI.

The majority of existing datasets is concerned with the analysis of static images or video sequences. However, it has been shown that the inclusion of depth perception changes human attention behavior [2] and several attempts have been made to collect eye-tracking data while free-viewing stereo image pairs [3], artificially created 3D scenes [4], and even real 3D environments [7]. Therefore, we want to create a dataset in which people perform tasks in real 3D environments. Collecting such experimental 3D eye-tracking data will strongly benefit the community, and will show the influence of depth and motion on human perception.

Ekaterina Potapova, Astrid Weiss, Michael Zillich, Markus Vincze
ACIN, Vienna University of Technology, e-mail: {potapova,weiss,zillich,vincze}@acin.tuwien.ac.at

Valsamis Ntouskos, Fiora Pirri
ALCOR, University of Rome, e-mail: {ntouskos,pirri}@dis.uniroma1.it

The research leading to these results has received funding from the Austrian Science Foundation (FWF) under grant agreements No. TRP 139-N23, InSitu and No. TRP 623-N23, V4HRC, and the European Communitys Seventh Framework Programme FP7/2007-2013 under grant agreement No. 600623, STRANDS

We aim at producing a pilot dataset with 3D eye-tracking data from three participants to check if it is feasible to create a benchmark for saliency algorithms and computational attention models aimed at solving top-down search tasks.

2 Task Description

In the following we will present two exemplary tasks participants were asked to perform in our pilot study: a free-viewing and a counting task. We collected eye-tracking data in real cluttered environments using the Gaze Machine [5].

2.1 Experimental Setup

As a setup for our experiments we chose a cluttered scene, as it offers a variety of possibilities for bottom-up and top-down search tasks. The selected scene represented a pile of toys as it can be found in every child's room. Toys are stacked together and occluding each other. Therefore, 3D data is essential, because depth can potentially help to separate items. Our experimental setup consisted of a set of toys, randomly placed on the table. A white screen was placed in front of the participant and lifted for a specific amount of time, ensuring a precisely defined duration of viewing the scene. All questions to participants and their answers were documented. Additionally, experiments were video taped.

2.2 Free-viewing Task

In this task we studied bottom-up attention. In order to keep participants motivated the goal was to remember as many details as possible during a limited period of 30 seconds. After the completion of the experiment participants were asked several questions, to check their performance, such as 'Where can this type of scene be typically observed?', 'How many toys are in the scene?', 'Could you please name some toys you have seen?'. Points of regard were mapped onto image sequences for each participant separately to create preliminary fixation distance maps. We expected participants to be attracted by regions being in contrast with the surrounding, similar to [6].

2.3 Counting Task

In this task we studied top-down attention. We asked participants to search for cars in the scene and count them loudly when they find one. Search patterns were recorded. The experiment was stopped when a participant reported that all cars found. With such data it becomes possible to study human behaviour in terms of evaluation of directed top-down visual attention. We expected participants to sequentially search the scene from the most attractive region, i.e. saliency region, to the least attractive.

3 Participants

As a pilot study we invited three participants to check the feasibility of our methodological approach. All three were male students at the University of Rome, with age from 27 to 29. All of them had normal or corrected to normal vision.

4 Data Processing

To create a benchmark that can be used by other researchers, a 3D reconstruction model of the scene was created at first. Points of regard were localized and mapped onto this reconstruction. However, because of lighting conditions and image blurriness due to participants movements, it was not possible to localize points of regard on the unified model. Therefore, it was not possible to create hit maps in 3D. These issues will be addressed in the followed-up main study. Additionally, points of regard were mapped to one of two images in a stereo pair. Fixation distance maps were created for each frame for each participant. No averaged fixation distance maps were created at this points.

5 Preliminary Results

No hit maps were created in the pilot study, therefore it is not possible to quantitatively compare eye-tracking data with existing models of visual attention. Preliminary results show that attentional models designed to work on static images or videos fail to predict fixations when they are applied in real world environments.

The results of the free-viewing experiment were qualitatively compared to contrast-based attentional model described by Itti *et al.* [6]. Comparison showed that while contrast-based model predicts fixations on objects popping out from the scene based on color, human fixations were mostly distributed in areas with multiple tiny objects. One interesting question here is, whether this behaviour was influenced by the given task to motivate participants or by the nature of human attention. Potentially, it can show that humans are attracted by textured regions. This aspect should be investigated in more details.

Fixation data from counting experiment showed that humans tend to return to previous fixations. Currently existing models of human visual attention do not assume the possibility to go back and forth between already attended location. Potentially, this finding can indicate that humans tend to verify their visual experience by examining same regions with interesting objects several times.

In general, the pilot study greatly helped to adjust the setting for the upcoming studies. Data, obtained from the pilot study can lead to designing an attentional system comparable to human performance, when depth and orientation in space are taken into account.

6 Conclusion and Future Work

In this paper we addressed the issue of creating a benchmark for visual attention models based on eye-tracking data in realistic 3D environments. As an example a setup with cluttered toys was created. Participants performed two different tasks: free-viewing and counting. Pilot study showed that classical contrast-based models fail to predict human fixations in real 3D environments. Next steps will be to create a unified 3D reconstructed model of the scene and map fixation points onto this model. Our future work will also include the design of a general protocol to perform more experiments on visual search and algorithms to process eye-tracking data.

References

1. Borji A., Sihite D., Itti L. (2012). Probabilistic Learning of Task-Specific Visual Attention. In: Computer Vision and Pattern Recognition, 2012.
2. Wexler M., Ouarti N. (2008). Depth Affects Where We Look. In: Current Biology 18, 2008.
3. Lang C., Nguyen T., Katti H., Yadati K., Kankanhalli M., Yan S. (2012). Depth Matters: Influence of Depth Cues on Visual Saliency. In: European Conference on Computer Vision, 2012.
4. Vishwanath D., Kowler E. (2004). Saccadic localization in the presence of cues to three-dimensional shape. In: Journal of Vision 4, 2004.
5. Pirri F., Pizzoli M., Rudi A. (2011). A general method for the point of regard estimation in 3D space. In: Computer Vision and Pattern Recognition, 2011.
6. Itti L., Koch C., Niebur E. (1998). A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(11), 1998.
7. Santner K., Fritz G., Paletta L., Mayer H. (2013). Visual Recovery of Saliency Maps from Human Attention in 3D Environments. In: IEEE International Conference on Robotics and Automation, 2013.

DOI: 10.2390/biecoll-saga2013_2

In Pfeiffer and Essig (2013): Proceedings of the First International Workshop on Solutions for Automatic Gaze-Data Analysis 2013 (SAGA 2013). DOI: 10.2390/biecoll-saga2013_0

Mental representation and mental practice: The influence of imagery rehearsal on representation structure, gaze behavior and performance

Cornelia Frank, Kai Essig and Thomas Schack

1 Introduction

Up to now, research has elicited differences in mental representations of complex action between experts and novices [1]. That is, while mental representations of experts are organized hierarchically and structured in a functional way, representations of novices are less hierarchically organized and less structured, such that they match poorly the functional and biomechanical demands of the task. Recently, it has been demonstrated that novices' representation structures of complex action functionally adapt as a result of physical practice during skill acquisition [2]. More recently, it has been suggested that mental practice adds to this cognitive adaptation process [3]. Specifically, after mental and physical practice, participants showed quite elaborate representation structures. In contrast, participants practicing physically only revealed less elaborate representation structures. Thus, representation structures develop differently depending on the type of practice. Moreover, mental practice seems to add to the functional adaptation of representations during skill acquisition.

In a next step, we were interested in gaining a more detailed understanding of the perceptual-cognitive background of performance changes during skill acquisition. Therefore, in the present study the question was addressed whether changes in mental representation structure of the putt by way of mental practice go along with both changes in gaze behavior while putting and changes in putting performance during early skill acquisition. More specifically, our goal was to investigate the effect of mental practice on motor performance and to further explore the cognitive-perceptual background of motor control and error learning.

Cornelia Frank, mail: cornelia.frank@uni-bielefeld.de

Kai Essig, mail: kai.essig@uni-bielefeld.de

Thomas Schack, mail: thomas.schack@uni-bielefeld.de

Neurocognition and Action – Biomechanics Research Group/

Cognitive Interaction Technology Center of Excellence (CITEC) Bielefeld University

2 Experiment

The study consisted of a pre-test, an acquisition phase of three days, and a post-test. Novice golfers (N = 30) were assigned to one of three conditions according to their pre-test putting performance: (1) combined mental and physical practice, (2) physical practice, and (3) no practice. The combined practice group practiced the putt in golf both physically and mentally (i.e., repeatedly imagined the putting movement and the ball stopping on the target). The physical practice group practiced the golf putt physically only during acquisition phase, while the control group did not practice at all. Participants' representation structures of the putt, their gaze behavior and their putting performance were assessed prior to and after acquisition phase. In addition, imagery ability was measured. Furthermore, a post-experimental questionnaire was administered after each imagery session to serve as a manipulation check.

More specifically, putting performance was assessed by tracking and capturing the final position of the ball after each putt via motion capture system. In addition, gaze behavior while putting was measured using a mobile eye-tracking system with scene and eye camera attached to a helmet. In order to examine mental representation structure, structural dimensional analysis of mental representations (SDA-M) was employed [4]. In short, SDA-M is used to obtain psychometric data on mental representation structure of a complex movement (here: the golf putt) in long-term memory. In other words, with this method it is possible to learn about distances and groupings of basic action concepts (i.e., mental representation structure). Specifically, a split procedure serves to estimate distances between the basic action concepts (BACs) of a predetermined set of concepts. For the golf putting movement, a set of 16 BACs had been identified [2]. Accordingly, this set of BACs was used for the present study. As described elsewhere in more detail [2, 4], the split procedure is performed in front of a computer with the screen displaying the BACs of the complex movement (here: the golf putt). One selected BAC is permanently displayed on the screen (anchor concept) while the rest of the BACs (n = N-1; here: 15) are presented successively in randomized order. Participants are asked to decide, one after another, whether a given BAC is related to the anchor concept or not during movement execution. Once a given list of BACs is finished, the next BAC serves as an anchor concept and the procedure continues. The split procedure ends after each BAC has been compared to the remaining BACs in the list. Following this procedure, a hierarchical cluster analysis then serves to outline the groupings of the set of BACs (i.e., mental representation structure).

3 Results

In line with previous work, preliminary findings of the present study revealed that representation structures changed over the course of practice. Moreover, representation structures of the combined mental and physical practice group were more elaborate after acquisition phase than those of the physical practice group and the control group. With respect to gaze behavior, preliminary analyses of participants' gaze prior to the onset of the movement [5] indicated changes in fixation durations of the final fixation before initiation of the motor action (i.e., the putting movement) as a result of practice. Specifically, both practice groups revealed longer fixation durations of the final fixation prior to movement onset after three days of practice compared to pre-test. Moreover, differences between the groups were obvious after acquisition phase. That is, for post-test, longest fixation durations have been found for the combined mental and physical practice group, followed by the physical practice group, with the control group revealing the shortest fixation durations prior to movement initiation.

4 Discussion and Future Work

Preliminary results of the present study support findings from previous research such that mental practice adds to the cognitive adaptation process during motor skill acquisition. Importantly, with respect to gaze behavior prior to the onset of the movement, data indicate differences between the practice conditions. It seems to be the case that combined mental and physical practice is associated with longer fixation durations, and therefore longer information processing, prior to movement initiation compared to physical practice only. Hence, our preliminary results suggest that the influence of mental practice and physical practice on gaze behavior is different prior to movement onset. These preliminary findings point toward the idea that mental practice may not only add to the cognitive adaptation, but also to the perceptual adaptation during motor skill acquisition. Further details on analyses and findings regarding gaze behavior will be presented and discussed during the workshop.

ACKNOWLEDGMENTS: Cornelia Frank gratefully acknowledges the financial support from the German Research Foundation grant DFG EXC 277 "Cognitive Interaction Technology" (CITEC).

References

1. Schack, T., & Mechsner, F. (2006). Representation of motor skills in human long-term memory. *Neuroscience Letters, 391*, 77-81.
2. Frank, C., Land, W. M., & Schack, T. (2013). Mental representation and learning: The influence of practice on the development of mental representation structure in complex action. *Psychology of Sport and Exercise, 14,* 353-361.

3. Frank, C., Land, W. M., & Schack, T. (2013). Behind the curtain: The influence of mental practice on the development of mental representation structure in early skill acquisition. *Annual meeting of the Research on Imagery and Observation (RIO) group in Dublin, Ireland.*
4. Schack, T. (2012). Measuring mental representations. In: Tenenbaum, G., Eklund, R. C., Kamata, A. (Eds.) *Handbook of measurement in sport and exercise psychology* (pp. 203-214). Human Kinetics, Champaign.
5. Vickers, J. N. (2009). Advances in coupling perception and action: The quiet eye as a bidirectional link between gaze, attention, and action. In M. Raab, J. G. Johnson, & H. R. Heekeren (Eds.), *Mind and motion: The bidirectional link between thought and action. Progress in Brain Research, 174*, 279-288.

DOI: 10.2390/biecoll-saga2013_3

In Pfeiffer and Essig (2013): Proceedings of the First International Workshop on Solutions for Automatic Gaze-Data Analysis 2013 (SAGA 2013). DOI: 10.2390/biecoll-saga2013_0

Mobile Eyetracking for Decision Analysis at the Point-of-Sale: Requirements from the Perspectives of Marketing Research and Human-Computer Interaction

Martin Meißner, Jella Pfeiffer, and Thies Pfeiffer

1 Introduction

In a typical grocery-shopping trip consumers are overwhelmed not only by the number of products and brands in the store, but also by other possible distractions like advertisements, other consumers or smartphones. In this environment, attention is the key source for investigating the decision processes of customers. Recent mobile eyetracking systems have opened the gate to a better understanding of in-store attention. We present perspectives from the two disciplines marketing research and human-computer interaction and refine methodical and technological requirements for attention analysis at the point-of-sale (POS).

2 The Marketing-Research Perspective

For those who want to sell their products to the potential customers, knowledge about the customers' decision processes is crucial. Most of the research findings cited below come from experiments which have been set up in lab-like situations. For marketing researchers, however, experiments in more realistic decision environments are important to solve the following central research questions.

How can we bring more attention to our product in the store? A couple of studies have been conducted to determine whether attention to products in a store is top-down (endogenous), which means that consumers direct their attention to products according to their preferences, or rather bottom-up (exogenous), which means that preferences are being formed based on attention (Theeuwes, 2010; Orquin and Mueller Loose, 2013). Identifying in which decision contexts attention is driven by endogenous and exogenous factors has been recognized as an important research topic (see, e.g. Atalay et al., 2012). Research has shown that visual saliency is an exogenous factor that makes products more likely to be chosen (Milosavljevic et al., 2012). That makes tools which assess the saliency of products in competitive settings a valuable resource for marketing practitioners. Centrality is a second important exogenous factor. Recent findings suggest that central options in

Jella Pfeiffer, Chair of Information Systems and Business Administration
Johannes Gutenberg-University Mainz, e-mail: jella.pfeiffer@uni-mainz.de
Thies Pfeiffer, e-mail: Thies.Pfeiffer@uni-bielefeld.de
Cognitive Interaction Technology Center of Excellence, Bielefeld University
Martin Meißner, e-mail: mmeissner@wiwi.uni-bielefeld.de
Chair of Business Administration and Marketing, Bielefeld University

a supermarket shelf receive an increased amount of attention which also increases the likelihood that a central-positioned product is being chosen (Atalay et al., 2012).

What can we learn from the attentional processes of consumers about their decision processes? Research has shown that the finally chosen product receives more attention than the non-chosen products (Pieters and Warlop, 1999). Building on these findings, Shimojo et al. (2003) gave evidence that the increased attention on the finally chosen product results from the fact that attention shifts to this option in the very last seconds of the decision process. This effect has been called a gaze cascade effect and has been replicated in a couple of different decision environments (see, e.g., Shi et al., 2013), but not at the POS.

Researchers have also developed metrics to divide the attentional process into different stages. Russo and Leclerc (1994), for example, suggest using refixations to define three stages of the decision process which they named the orientation, evaluation and verification stage. An important aim of research in this field therefore is to develop ways to define decision stages and associated information needs.

How do previous attentional processes influence later attentional processes in POS-decisions? Most of the previously described studies have been tested in single decisions. Grocery-shopping trips, however, most of the times include dozens of purchase decisions which may influence one another. The influence of attentional processes in earlier decisions on later ones has been investigated in only a small number of studies (see, e.g., Janiszewski et al., 2013).

3 The Human-Computer-Interaction Perspective

Imagine consumers wearing an attentive mobile interactive cognitive assistant (AMICA), e.g. realized as “intelligent goggles”, at the point-of-sale that gives advice based on the context-sensitive data collected by an integrated eyetracking device. The AMICA will detect the interaction context by localization and computer vision techniques and activate apps tailored to the specific situation, here a shopping decision. Instead of a command-based interface, the attentive system monitors the ongoing cognitive processes of the wearer in particular by observing eye movements and establishing a semantic link between the fixations and the objects of the environment. Just as an observant sales-person, the AMICA will stay in the background, monitoring the ongoing decision process and only provide help when it detects uncertainty in the user’s gaze patterns, or if it is directly asked to do so.

Regarding general behavior models of such decision processes, HCI perspective and Marketing Research meet in their interest in mobile eyetracking systems. For example, findings on decision stages or the gaze cascade effect (see Section 2) allow predicting the stage of the information acquisition process. In each of the stages, consumers have different information needs. Early in the decision process, explanation-facilities about the use of the recommender system help building trust and lead to a higher chance of adoption (Wang and Benbasat, 2005; Wang and Benbasat, 2007). Furthermore, consumers might need additional information about the product attributes being relevant in the purchase decision depending on the consumers' product knowledge. In the next stage, promising products must be identified by the system which the consumer will compare in more detail along several product attributes in the later stages (Häubl and Trifts, 2000).

However, an AMICA needs to identify the context and the objects of interest in real-time, to detect products and attributes of interest. The detection needs to be highly robust and be able to cope with dynamic environments, because errors will not average out with an increased number of trials. It also requires a high level of personalization, both to adapt to the peculiarities of the individual user's cognitive processes and to the user's preferences to provide sound recommendations.

4 Requirement Analysis

Based on the research questions outlined above, we identify the following information and technical requirements for mobile eyetracking at the POS:

It is not surprising that knowledge about typical features such as *fixations* and *dwell-times* are a common requirement. *Transitions between fixations* are equally important and thus *saccadic distances* and *saccadic speed between fixations* provide highly supplemental information. In all cases, the *semantic link between* overt *visual attention and the object of interest* is of uttermost importance. Information about visible *product categories* and *individual products in focus* as well as *product features* such as price, brand name and *specific features,* e.g., nutritional information are required. Finally, the *choice of the customer,* as the outcome of a decision process, needs to be detected, too.

Marketing Research is in particular interested in the *topological distribution of attention*, e.g. over a shelf, to answer questions regarding the *saliency of products* in a shelf or shelf set-up and centrality of a product for product placement. These aspects are not so important for HCI, but robustness of the systems could nevertheless benefit if topological information is available. As Marketing Research is interested in *aggregating and visualizing gaze data from many customers,* a relatively *stable topology of the products* and a *large number of respondents* for statistical tests are required.

HCI, in contrast, focusses on the individual and the interaction does not stop at the shelf: Typically customers take products out of the shelf to inspect them, to haptically test them, and to carry them around. Approaches that *support dynamic and noisy scenarios* are thus a must in the envisioned HCI context. *Low latencies* for the classification of gaze data and the fixated objects of interest are important to support a timely reaction of the system. This is challenged by the requirement of a *robust classification process* to allow for a smooth human-computer interaction without many interactive reparations and backtracking operations.

5 Summary

We have approached the analysis of the decision processes at the POS from two perspectives. By tendency, Marketing Research is often associated with the interests of salesperson, but this is not necessarily the case. While Marketing Research has a stronger focus on the description of behavior in general, HCI focusses stronger on individual preferences. In the end, both approaches may cross-fertilize, as one benefits from the methods and results aimed at by the other discipline, e.g. HCI could benefit from an initial general behavior model that can be personalized over time and Marketing Research could benefit from unobtrusive and flexible methods for measuring decision processes as they happen at the POS and not in clean laboratory environments. The technology developed in the HCI part (AMICA) is applicable to other domains, as long as the domain specific aspects of the cognitive models and the computer vision parts are adapted.

ACKNOWLEDGMENTS: This project is funded by the Internal University Research Funding of the Johannes Gutenberg-Universität Mainz and by the Bielefelder Nachwuchsfonds of Bielefeld University.

References

1. Atalay AS, Onur Budur H, Rasolofoarison D (2012). Shining in the Center: Central Gaze Cascade Effect on Product Choice. Journal of Consumer Research 39(4):848-866.
2. Häubl G, Trifts V (2000). Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids. Marketing Science 19(1):4–21.
3. Janiszewski C, Kuo A, Tavassoli NT (2013). The Influence of Selective Attention and Inattention to Products on Subsequent Choice. Journal of Consumer Research, in print.
4. Milosavljevic M, Navalpakkam V, Koch C, Rangel A (2012). Relative Visual Saliency Differences Induce Sizable Bias in Consumer Choice. Journal of Consumer Psychology 22(1):67-74.
5. Orquin, JL, Mueller Loose S (2013). Attention and Choice: A Review on Eye Movements in Decision Making. Acta Psychologica 144 (1):190-206
6. Pieters, R, Warlop L (1999). Visual Attention during Brand Choice: The Impact of Time Pressure and Task Motivation. International Journal of Research in Marketing 16 (1):1-16
7. Russo, JE, Leclerc F (1994). An Eye-fixation Analysis of Choice Processes for Consumer Nondurables. Journal of Consumer Research 21 (2):274-290
8. Shi W, Wedel M, Pieters R (2013). Information Acquisition during Online Decision-Making: A Model-Based Exploration Using Eye-Tracking Data. Management Science in print.
9. Shimojo S, Simion C, Shimojo E, Scheier C (2003). Gaze Bias Both Reflects and Influences Preference. Nature Neuroscience 6(12):1317-1322.
10. Theeuwes J (2010). Top-down and Bottom-up Control of Visual Selection. Acta Psychologica 135(2):77-99.
11. Wang W, Benbasat I (2005). Trust in and adoption of online recommendation agents. Journal of the Association for Information Systems 6(3):72–101.
12. Wang W, Benbasat I (2007). Recommendation Agents for Electronic Commerce: Effects of Explanation Facilities on Trusting Beliefs. Journal of Management Information Systems 23(4):217–246.

Part II

Session 02: Methods for Gaze Analysis

DOI: 10.2390/biecoll-saga2013_4

In Pfeiffer and Essig (2013): Proceedings of the First International Workshop on Solutions for Automatic Gaze-Data Analysis 2013 (SAGA 2013). DOI: 10.2390/biecoll-saga2013_0

Experiences from eye tracking chat-communication using the INKA-SUITE

Prof. Dr. Andrea Kienle, Christian Schlösser, Philipp Schlieker-Steens

1 Introduction

Eye tracking is a common way for analyzing the usability of applications (e.g., Duchowski, 2007), stationary single user systems (e.g., Shneiderman & Plaisant, 2010) and websites (e.g., Nielsen & Loranger, 2006). Existing eye tracking systems are limited to static content like websites. In these applications the content is placed at the same position for all participants and can be easily compared. If this area of interest (AOI) is not static, but changes its size, shape or position, the evaluation becomes more complex and requires an amount of manual analysis. This is especially the case in cooperative applications like chats. The difficulty is, that you cannot predict the time of change, because it is caused by user input (e.g., new chat message, scrolling) - the dynamics of the AOI cannot be foretold. Existing eye tracking software is therefore not suitable in situations of changing AOIs. Therefore a new approach is needed to overcome these problems.
The paper is structured by introducing you to the INKA-SUITE itself, followed by the description of a first experiment with its results. At the end there is an outlook.

2 INKA-SUITE

The INKA-SUITE is a platform that integrates the application and the analysis of eye tracking dynamic AOIs (e.g., Kienle et al., 2013). The base of the INKA-SUITE is the connectivity to Tobii eye trackers (http://www.tobii.com). With a direct connection between the eye tracker gaze data stream and GUI, the INKA-SUITE identifies AOIs at runtime, regardless of size, shape and position. Every gaze data record is complemented by the underlying GUI element, identified by names, ID's and/or references to other database tables such as chat messages or users. A subsequent manual work is eliminated and the evaluation can be started immediately. This profound cross-linking between analyses related functions and user software cannot be adapted by existing eye tracking analysis software.

Prof. Dr. Andrea Kienle, Christian Schlösser, Philipp Schlieker-Steens
University of Applied Sciences and Arts Dortmund, Emil-Figge-Str. 42, 44227 Dortmund (Germany),
Email: andrea.kienle@fh-dortmund.de, christian.schloesser@fh-dortmund.de, philipp.schlieker-steens@fh-dortmund.de

The components of the INKA-SUITE are:

(a) A **server** to control and manage clients as well as providing templates for the clients. Using templates enables researchers to conduct A/B tests.

(b) A **client application** for user interaction, which is connected to an eye tracker and presents the chosen template.

(c) An **analysis component** for managing projects and analyzing the collected data. Because fixations are represented by GUI elements, further information, such as fixation length and underlying AOI, can be retrieved through tooltips.

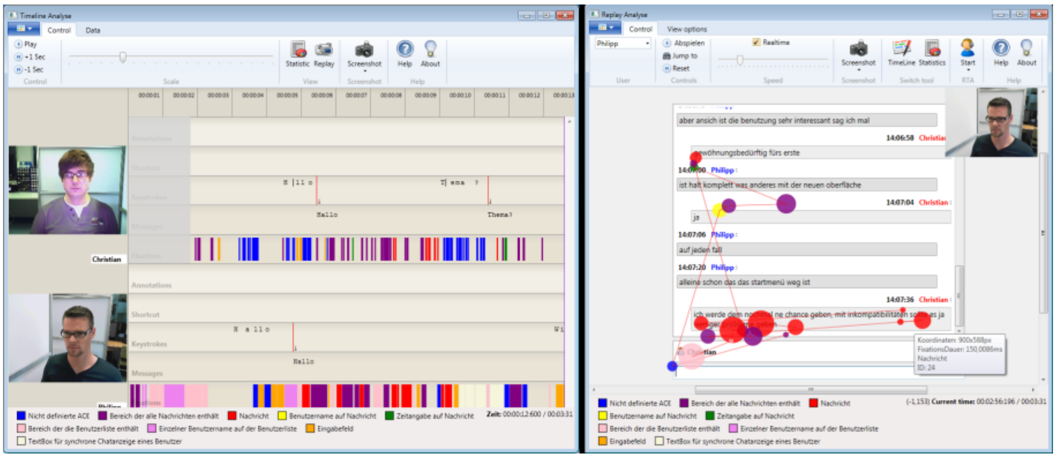

Fig. 1: INKA-SUITE Timeline (left) and Replay (right) as part of the analysis component

The analysis component processes the recorded data and is divided into three parts:

(a) In the **statistic tool**, general information about the chat session and also about each user is presented. For example these are duration of chat, keystrokes per minute, time to first message and messages per minute.

(b) Within the **timeline tool,** the entire chat session is presented on a timeline for each user (see Figure 1, left). This ensures the comparability between users. Shortcuts, keystrokes, chat messages and fixations of the users are listed separately for each user. Using annotations, important areas can be marked. The displayed output is fully dynamic and can be adapted to the current problem. The presentation of the collected data in form of the timeline, is oriented at the table structure of Beißwenger (2007) and the presentation of fixations by Stellmach et al. (2010).

(c) The **replay tool** shows a replay of the chat from the user's point of view (see Figure 1, right). This tool is chronology replaying the chat, similar to a screen recording, but with options to select and deselect the data that is output at runtime, also showing fixations and saccades as a scanpath (e.g., Mealha et al., 2012).

3 Experiment

To test the INKA-SUITE, an experiment in cooperation with the Institute of German language and literature of the TU Dortmund was accomplished. In this study the usability of the INKA-SUITE should be tested at first. Secondly questions from the linguists with respect to the influence of different user interfaces in chats for the structure of discussions should be answered.

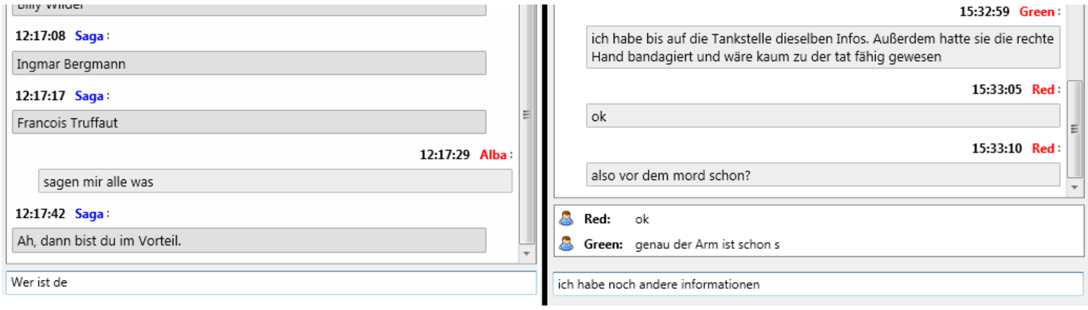

Fig. 2: Two templates: Standard Chat (left) and Talk (right)

By using the templating feature two templates were created (see Figure 2). One template (Figure 2 left) represents a Standard-Chat-Environment (SCE) (e.g., Beißwenger, 2003) with an input field and a chat protocol. Template two (Figure 2 right) is designed as a so called Talk layout similar to the Unix Talk. It contains the same features as the SCE but in addition a synchronous user list, which shows all connected chat partners and a live view from their text input fields.

Day 1 - Chat	**Groups of two**	**Groups of three**
Movie jury - scenario	2	1
Murderer - scenario		2
Day 2 - Talk	**Groups of two**	**Groups of three**
Movie jury - scenario	2	1
Murderer - scenario		2

Tab. 1: Study organization

The table above offers an overview about the experiment organization. In the study, 26 participants were recorded while chatting in two different scenarios. To solve the problems given by the scenarios, the participants have to exchange their information which differs to each other.
The using of the INKA-SUITE while collecting and analyzing the data provides no problems. All ten chat-communications were completely recorded and supply a wide database for the analysis by the linguists.

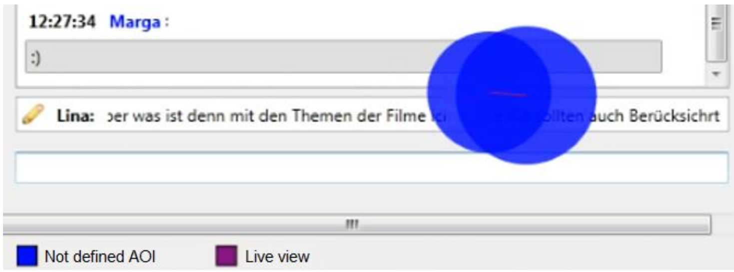

Fig. 3: Study result – eye tracking inaccuracy

The study revealed that the inaccuracy of the eye trackers is a problem (see Figure 3). Some gaze plots do not land on the AOI which is fixated. Therefore a manual effort is needed in some cases. A solution could be repositioning and larger scaling of the AOIs. The analysis by the linguists is still ongoing.

4 Outlook

The INKA-SUITE is basis for further research like automatic pattern recognition (e.g., revisions), contextualized communication and extension of simultaneous chats. A planned project called Chat++, enabled by real-time AOI identification, could therefore support features like reading awareness, eye tracking-based referencing and activity and context-awareness.

References

1. Beißwenger, M. (2003): Sprachhandlungskoordination im Chat. In: Zeitschrift für germanistische Linguistik 31 (2), S. 198–231.
2. Beißwenger, M. (2007): Sprachhandlungskoordination in der Chat-Kommunikation: Walter de Gruyter (26).
3. Duchowski, Andrew (2007): Eye Tracking Methodology. Theory and Practice. Second. London: Springer-Verlag London Limited. Online verfügbar unter http://dx.doi.org/10.1007/978-1-84628-609-4.
4. Kienle, A.; Schlieker-Steens, P.; Schlösser C. (2013): INKA-SUITE: An Integrated Test-Environment for Analyzing Chat Communication. In: CSCL 2013 - Conference Proceedings Volume II, Conference Proceedings Volume II, S. 285–286.
5. Mealha, O.; Veloso, A.; Almeida, S.; Rodrigues, R.; Roque, L.; Marques, R.; Manteigueiro, C. (2012): Eye tracking data representation and visualization: on information and communication studies at CETAC. MEDIA.
6. Nielsen, Jakob; Loranger, Hoa (2006): Prioritizing web usability: New Riders.
7. Shneiderman, Ben; Plaisant, Catherine (2010): Designing the user interface. Strategies for effective human-computer interaction. 5. Aufl. Boston: Addison-Wesley.
8. Stellmach, Sophie; Nacke, Lennart; Dachselt, Raimund (2010): Advanced gaze visualizations for three-dimensional virtual environments. In: Proceedings of the 2010 symposium on eye-tracking research & Applications. ACM, S. 109–112.

DOI: 10.2390/biecoll-saga2013_5

In Pfeiffer and Essig (2013): Proceedings of the First International Workshop on Solutions for Automatic Gaze-Data Analysis 2013 (SAGA 2013). DOI: 10.2390/biecoll-saga2013_0

Using spatial statistics to investigate within-trial correlations of human gaze positions

Hans A. Trukenbrod, Simon Barthelmé, Felix Wichmann, and Ralf Engbert

1 Introduction

The distribution of fixation locations can be interpreted as an intensity (density) function of an underlying spatial point process (Barthelmé, Trukenbrod, Engbert, & Wichmann, in press). In point process theory (Illian, Penttinen, Stoyan, & Stoyan, 2008), we analyze the point-to-point interactions to infer possible generating mechanisms. The pair correlation function (PCF) provides a mathematical measure of the statistical interaction of neighboring points (Law, Illian, Burslem, Gratzer, Gunatilleke, & Gunatilleke, 2009). Here we demonstrate that the inhomogeneous PCF can be used to analyze sequences of fixation locations recorded from human observers. The resulting PCF removes first-order heterogeneity induced by systematic variation of saliency within a given scene from second-order spatial statistics. Our results indicate significant spatial clustering at short length scales.

Hans A. Trukenbrod
University of Potsdam, Department of Psychology, Karl-Liebknecht-Str. 24/25, 14476 Potsdam, Germany, e-mail: Hans.Trukenbrod@uni-potsdam.de

Simon Barthelmé
University of Geneva, Boulevard du Pont-dArve 40 1205, Genéve, Switzerland, e-mail: Simon.Barthelme@unige.ch

Felix Wichmann
University of Tübingen, Neural Information Processing Group, Faculty of Science, Sand 6, 72076 Tübingen, Germany, e-mail: Felix.Wichmann@uni-tuebingen.de

Ralf Engbert
University of Potsdam, Department of Psychology, Karl-Liebknecht-Str. 24/25, 14476 Potsdam, Germany, e-mail: Ralf.Engbert@uni-potsdam.de

2 Experiment

The present study explores the possibility to apply the *pair correlation function* (PCF), which provides a description of the distribution of distances between points, to the spatial analysis of fixation locations generated during single scanpaths. Our results are based on a reanalysis of scene-viewing data previously published by Mergenthaler and Engbert (2009).

Participants: We tested 24 students of the University of Potsdam (mean age: 23 years) with normal or corrected-to-normal vision. All participants received credit points for participation.

Stimulus material: Images consisted of twelve randomly selected, colored, natural landscape photographs presented across the entire screen (CRT display; Iiyama Vision Mater Pro 514; frame rate 100 Hz; resolution: 1024 $\times$ 768 pixels).

Task and procedure: Participants were instructed to position their head on a chinrest in front of a computer screen (viewing distance: 50 cm). Eye movements were recorded using a Eyelink II video-based eye-tracker (SR-Research, Osgoode/ON, Canada) with a sampling rate of 500 Hz. Each image was presented for 10 s. Participants were asked to freely explore each image. Between images, participants were engaged in a secondary fixation task for about 20 s.

Statistical analysis: Pair correlation functions (PCFs) were calculated in two steps. A detailed description of the underlying procedure can be found in Law et al. (2009).

First, we numerically estimated the intensity (or spatial density) of all fixations on a given image (across all participants). The intensity function was computed applying a 2D kernel density estimator (Baddeley & Turner, 2005; R Development Core Team, 2013). The estimated intensity $\hat{\lambda}(x,y)$ gives the density of all fixations, where dependence on position (x,y) indicates inhomogeneity.

Second, due to spatial inhomogeneity in the spatial distributions (see step one) we estimated the *inhomogeneous pair correlation function*. The inhomogeneous PCF $g_{inhom}(r)$ is a function of the distance r between points. The estimated inhomogenous PCF is a quantitative measure of spatial correlations between fixation locations, where the inhomogeneous spatial density $\hat{\lambda}(x,y)$ is taken into account.

The interpretation $g_{inhom}(r)$ is as follows: For distances r with $g_{inhom}(r) \approx 1$ fixation locations are statistically uncorrelated, while values of $g_{inhom}(r) > 1$ reveal spatial clustering at distance r, and values of $g_{inhom}(r) < 1$ reveal inhibition at distance r.

3 Results

Inhomogeneous pair correlation functions (PCFs) of the experimental data are shown in Figure 1. Gray lines depict PCFs of individual trials and reveal high variability between trials. The average PCF across all trials is plotted in red. The mean

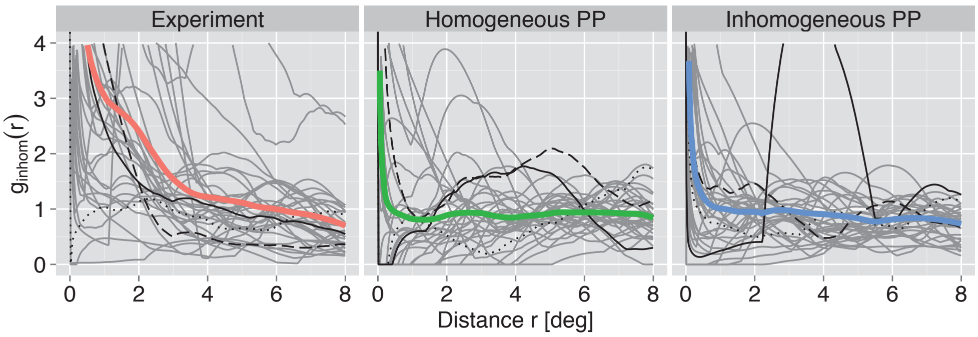

Fig. 1 Inhomogenous pair correlation function (PCF) for the experimental fixation sequences and surrogate fixation sequences. Surrogate fixation sequences were generated using a homogeneous and inhomogeneous Poisson process, respectively. Gray lines show PCFs of individual trials, colored lines the average PCF across all trials.

PCF indicates spatial clustering of fixations at small distances $r < 3^\circ$, while fixations at larger distances can be explained by the inhomogeneity of the spatial maps.

Next, we computed the PCFs for two surrogate data sets to test the validity of our statistical procedure. The surrogate data were selected to examine the null hypothesis of complete spatial randomness, both for a homogeneous Poisson process with constant intensity $\lambda(x,y) = \lambda_0$ (Fig. 1, central panel) and an inhomogeneous Poisson process with position-dependent intensity $\lambda(x,y)$ (Fig. 1, right panel). Analysis of our surrogate data reveal that both the homogeneous and inhomogeneous Poisson process give flat PCFs with $g_{inhom}(r) \approx 1$, which demonstrates the absence of clustering (with divergence at small scales being an artifact of numerical computations). We conclude that clustering in the experimental data is not a simple consequence of the inhomogeneous intensity.

4 Discussion and Further Work

Our analysis reveals that the pair correlation function (PCF) can be estimated for fixation locations from individual trials when normalized to the spatial density of all fixations. We observed strong clustering at small scales, i.e., at distances $r < 3^\circ$, reflecting a tendency of the eye to remain close to previously fixated regions. Finally, spatial statistics provide a number of useful tools to investigate the relation of fixation locations.

ACKNOWLEDGMENTS: This work was supported by Bundesministerium für Bildung und Forschung (BMBF, Bernstein Center for Computational Neuroscience Berlin, Project B3, Förderkennzeichen 01GQ1001F and 01GQ1001B to R.E. and F.A.W., respectively) and by Deutsche Forschungsgemeinschaft (DFG Research Unit 868, speaker: R.E.).

References

1. Baddeley A, Turner R (2005). Spatstat: An R package for analyzing spatial point patterns. Journal of Statistical Software 12:1-42.
2. Barthelmé S, Trukenbrod HA, Engbert R, Wichmann F (in press). Modeling fixation locations using spatial point processes. Journal of Vision.
3. Illian J, Penttinen A, Stoyan H, Stoyan D (2008). Statistical analysis and modelling of spatial point patterns. Chichester, UK: J. Wiley.
4. Law R, Illian J, Burslem DFRP, Gratzer G, Gunatilleke CVS, Gunatilleke IAUN (2009). Ecological information from spatial patterns of plants: Insights from point process theory. Journal of Ecology 97:616-628.
5. Mergenthaler K, Engbert R (2009). Microsaccades are different from saccades in scene perception. Experimental Brain Research 203:753-757.
6. R Development Core Team (2013). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. URL: http://www.R-project.org/

Part III
Session 03: Methods for Mobile Gaze Analysis

DOI: 10.2390/biecoll-saga2013_6

In Pfeiffer and Essig (2013): Proceedings of the First International Workshop on Solutions for Automatic Gaze-Data Analysis 2013 (SAGA 2013). DOI: 10.2390/biecoll-saga2013_0

Object recognition and person detection for mobile eye-tracking research. A case study with real-life customer journeys.

Stijn De Beugher, Geert Brône, and Toon Goedemé

1 Introduction

The recent development of user-friendly plug-and-play mobile eye-tracking technology has paved the way for research into visual behavior and real-life user experience in natural environments, such as public spaces, commercial environments or interpersonal communicative settings. The challenge for this new type of pervasive eye-tracking is the processing of data generated by the systems used in real-world environments [7]. Recently, several solutions to the analysis challenge have been proposed (see [2] for an overview). The best-known technique is the use of markers (infrared or natural) to predefine potential areas of interest (AOI), generating a two-dimensional plane within which eye gaze data can be collected for longer stretches of time and generalized across subjects. This paper presents an alternative to the AOI-based methods, building on recent studies combining object recognition algorithms with eye-tracking data [1] and [7].

2 Approach

By combining state-of-the-art object recognition [6] and person- [4] and face- [8] detection techniques for image processing (see figure 1), our system allows for a robust largely automatic analysis of relevant objects without the need for predefined areas of analysis or prior training. To process an eye-tracker experiment, one needs

Stijn De Beugher
EAVISE, ESAT - KU Leuven , Belgium e-mail: stijn.debeugher@thomasmore.be

Geert Brône
MIDI Research Group - KU Leuven, Belgium e-mail: geert.brone@arts.kuleuven.be

Toon Goedemé
EAVISE, ESAT - KU Leuven, Belgium e-mail: tgoedeme@esat.kuleuven.be

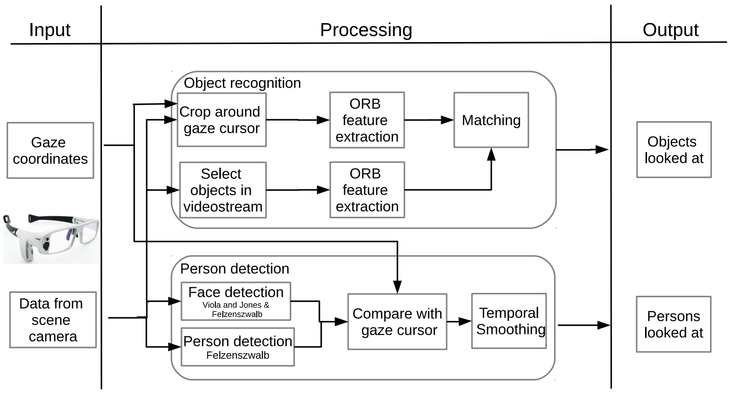

Fig. 1 High-level overview approach
Output is generated in the form of both a visual summary of the eye-tracking data and statistical information.

to select the objects of interest, by simply clicking on the objects while replaying the video of the eye-tracking experiment. The major advantage of this approach is that once the objects of interest are marked, they can be reused to process several eye-track recordings.

The next step of our automatic approach is the calculation of all fixations and fixation durations for the selected objects in the video. This is realized by searching for ORB feature [6] correspondences between the objects of interest and a region around the gaze cursor in each video frame, as shown by the green border in the left part of figure 2. At the same time, the system calculates how often and for how long one looked at another person, during this stage we make a distinction between specifically looking at a face, e.g. during talking, and looking at someone from a larger distance. In order to improve the detection-rate of faces and bodies, we applied two novel approaches. Firstly, we obtained an occlusion robust human torso detector by training a deformable part model [5] with only the top 60% of images from the VOC2009 [3] database, our model is illustrated at the right part of figure 2. The second novelty is a temporal smoothing system in which we use the gaze cursor as a tracker. This system assumes that a valid detection should stand for at least a certain time, thus preventing false detections and it allows us to solve gaps between detection sequences and therefore overcome missing detecitons.

The final step is the visualisation of our detection results. We chose to display the results on a timeline in terms of detected objects to give a chronological overview of the complete eye-tracking experiment, as shown in figure 3. The output of the detection results can be tuned through a set of parameters such as detection threshold, minimum fixation duration or the maximum gap between visual fixations.

3 Experiment

For this study, we conducted a real-life experiment to test the overall performance of our detection scheme for processing mobile eye-tracking data. In order to collect

Fig. 2 Illustration of our matching techniques. Left part of the image illustrates the object correspondences. Right part illustrates a person detection and our model for torso and upperbody detection.

representative data in a natural user environment, we selected the typical customer journey of a visitor to a museum, starting from the ticket counter all the way to the gift shop. Fourteen participants (7 male-7 female) were recorded while they visited a special exhibition at Museum M in Leuven (Belgium). Recordings were made with Tobii Glasses and Arrington Gig-E60 mobile systems and lasted between 35 and 65 minutes.

4 Results

Figure 3 shows the visual output of the detection system. Each fixation on a pre-trained object or person with a duration of at least a certain time (tunable via threshold, for example 150 ms) is displayed as a separate thumbnail of the relevant object of interest, human face or torso. In this figure we have selected 12 objects of interest, including specific works of arts, an elevator, the ticket counter, an Ipod, a route map, human torso or face, etc. In our GUI we visualize the objects of interest at the upper horizontal bar. The lower bar of images represents the actual visual fixations in the entire eye-tracker experiment. For each fixation the start and stop time is shown.

Such a visualisation may be a valuable tool to gain insights into user experience. Since we provide a summary of a complete eye-tracker experiment with respect to the viewing behaviour towards specific objects in a set of thumbnails, analyzing a customer journey is simplified. Our tool makes it possible to answer customer journey related questions such as: ”Did the participants use the elevator to enter the exhibition?”, ”How long did it take before they entered the exhibition?”, ”Did the participants notice the walking guides at the start of the exhibition?”, ”Did they notice there was an Ipod in the exhibition?”, ”How did they navigate through the different works of art (order, time spent looking at the different works, etc.)?”. Since it is possible to reuse the marked objects of interest it is possible to compare recordings of several visitors and produce more generalized statistics, as shown in table 1. These results correspond to the questions of the post-questionnaire.

At the workshop, we will discuss in more detail the results of the experiment, both in terms of precision recall curves for the techniques that were used, as well as computation time and general usability of the system.

Table 1 Questions to be answered in the context of the museum visit.

Question	Visitor 1	Visitor 2	Visitor 3	Visitor 4
Time at cash deck?	1m22s	42s	49s	20s
Looked at work of art in entrance hall?	NO	NO	NO	YES
Make use of elevator or stairs?	Elevator	Stairs	Elevator	Stairs
Time to get to exhibition?	1m43s	5m3s	1m21s	3m17s
Looked at walking guides at start of exhibition?	NO	YES	YES	NO
Looked at Ipod?	NO	NO	YES	YES
Total time at the exhibition?	28m58s	51m13s	35m3s	37m27s
Make use of elevator or stairs to get back from the exhibition?	Elevator	Stairs	Elevator	Stairs

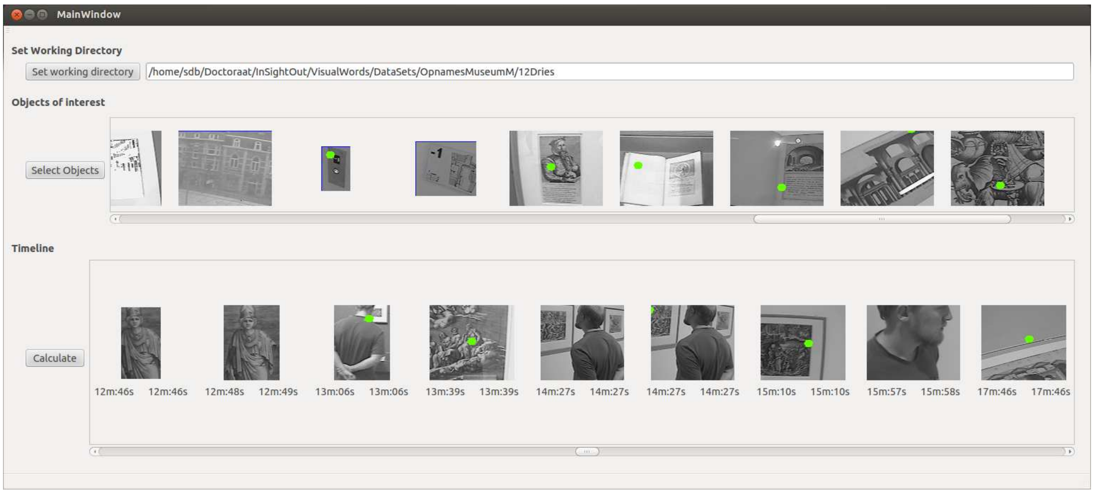

Fig. 3 Visualisation of the detection results. Top: manually selected objects. Bottom: timeline with objects and persons looked at.

References

1. S. De Beugher, G. Brône,T. Goedemé (2012). Automatic analysis of eye-tracking data using object detection algorithms. In Proc. of ACM Conference on Ubiquitous Computing pp. 677-680, Pittsburgh, USA.
2. K. M. Evans, R. A. Jacobs, J. A. Tarduno, J. B. Pelz (2012). Collecting and Analyzing Eye-tracking Data in Outdoor Environments. Journal of Eye Movement Research 5(2)6 pp. 1-19.
3. M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, A. Zisserman. The PASCAL Visual Object Classes Challenge 2009 (VOC2009) Results.
4. P. F. Felzenszwalb, R. B. Girshick, D. McAllester (2010). Cascade object detection with deformable part models. In Proc. of CVPR pp. 2241-2248, San Fransisco, USA.
5. P.F. Felzenszwalb, R. B. Girshick, D. McAllester. Discriminatively Trained Deformable Part Models, Release 4.
6. E. Rublee, V. Rabaud, K. Konolige, G. Bradski (2011). ORB: An efficient alternative to SIFT or SURF. In Proc. of ICCV 2564-2571, Barcelona, Spain.
7. T. Toyama, T. Kieninger, F. Shafait, A. Dengel (2012). Gaze guided object recognition using a head-mounted eye tracker. In Proc. of ETRA, pp. 91-98, California, USA.
8. P. Viola, M. Jones (2001). Rapid object detection using a boosted cascade of simple features. In Proc CVPR pp. 511-518, Kauai, USA.

DOI: 10.2390/biecoll-saga2013_7

In Pfeiffer and Essig (2013): Proceedings of the First International Workshop on Solutions for Automatic Gaze-Data Analysis 2013 (SAGA 2013). DOI: 10.2390/biecoll-saga2013_0

Studying joint attention and hand-eye coordination in human-human interaction: A model-based approach to an automatic mapping of fixations to target objects

Patrick Renner, Thies Pfeiffer

1 Introduction

If robots are to successfully interact in a space shared with humans, they should learn the communicative signals humans use in face-to-face interactions. For example, a robot can consider human presence for grasping decisions using a representation of peripersonal space (Holthaus & Wachsmuth, 2012). During interaction, the eye gaze of the interlocutor plays an important role. Using mechanisms of joint attention, gaze can be used to ground objects during interaction and knowledge about the current goals of the interlocutor are revealed (Imai et al., 2003). Eye movements are also known to precede hand pointing or grasping (Prablanc et al., 1979), which could help robots to predict areas with human activities, e.g. for security reasons.

We aim to study patterns of gaze and pointing in interaction space. The human participants' task is to jointly plan routes on a floor plan. For analysis, it is necessary to find fixations on specific rooms and floors as well as on the interlocutor's face or hands. Therefore, a model-based approach for automating this mapping was developed. This approach was evaluated using a highly accurate outside-in tracking system as baseline and a newly developed low-cost inside-out marker-based tracking system making use of the eye tracker's scene camera.

2 Experiment

The study is based on an adapted receptionist scenario: A map of a building with three floors is located between two participants. For each participant, there is one floor close-by, one in an intermediate distance within reaching space, and one in the distance which cannot be reached directly (Figure 1). Besides patterns of joint

Patrick Renner e-mail: prenner@techfak.uni-bielefeld.de
SFB 673, Faculty of Technology, Bielefeld University, 33615 Bielefeld

Thies Pfeiffer e-mail: tpfeiffe@techfak.uni-bielefeld.de
Cognitive Interaction Technology Center of Excellence (CITEC), Bielefeld University, 33615 Bielefeld

attention and patterns of hand-eye coordination that may predict pointing gestures, we are interested in differences in gaze behaviour and the use of modalities for the three distances. Head and hand movements of both participants are tracked using the high end system. Eye movements of one participant are tracked and both participants are recorded on two video cameras (audio and video).

The procedure of the main task is as follows: Firstly, one of the participants draws a miniature floor plan on which a starting point and a target room are marked. She is asked to show both of them on the real plan and describe the route. Afterwards, the partner draws a card with blockings drawn in. She is supposed to mark these with small tokens. Then, both participants jointly plan the fastest remaining route. The roles of the task are alternated each two of twelve trials altogether.

As we suppose that friends and acquaintances will act differently than strangers, in this stage of the experiment, each pair of participants is required to know each other. The participants firstly fill out a questionnaire which is meant to retrieve information about gender, age, debility of sight and prophecy in order to be able to account for possible behavioural differences.

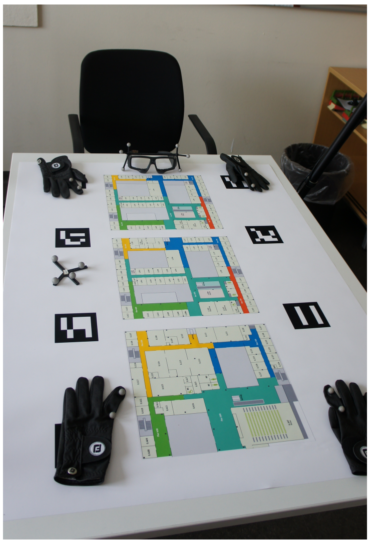

Fig. 1 Setup of the experiment: Floor plans, tracking markers, gloves and glasses.

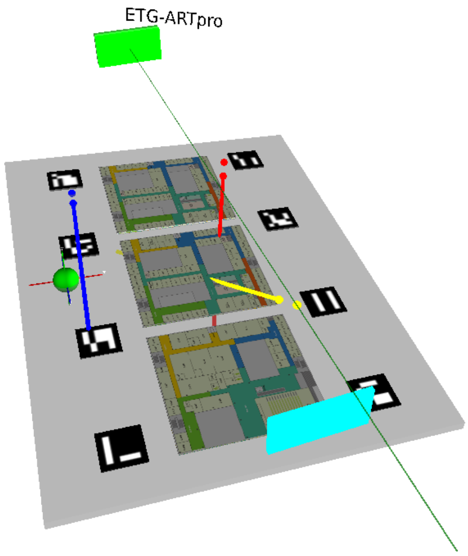

Fig. 2 Virtual reality simulation of the experiment including eye- (green), head- (green and blue) and hand (red and yellow) tracking.

3 Method: Model-based annotation of fixations

In experiments like ours, users move their head freely and gaze analysis requires a manual mapping of all fixations from the scene camera video to real world objects. To enable an automatic fixation mapping, the flexible head positions have to be tracked. Pfeiffer (2012) proposed a method for measuring attention in 3D space by using an optical tracking system and a tracking target attached to the eye tracker.

Here, we use two approaches of tracking the position of the eye tracker in space. An ART optical tracking system is a highly accurate and fast outside-in solution. Additionally, the built-in scene camera of the eye tracker is used as a low-cost inside-out tracking approach: By calibrating the intrinsic camera parameters, it is possible to transform recorded images of geometries of known size to their corresponding 3D pose in the real world. For that purpose, we chose fiducial augmented reality markers (see Figure 2 for examples) that can easily be detected in the camera images.

We then have to map gaze directions and fixations to our target stimuli: Room positions and the counterpart's head and hands. For this we modelled the floor plan in virtual reality and fed eye tracking and marker tracking data into the system. By reconstructing the line of sight in 3D, we can then automatically detect fixations on the virtual floor plan. For detecting fixations on the interlocutor, we use a face detection algorithm. Pointing directions are detected using the ART tracking system: The participants wear light gloves with markers attached.

4 Results and Conclusion

Using our approach (see Figure 3), the interaction of the participants can be simultaneously simulated in the virtual scene (Figure 2): Gaze directions are cast as rays into the scene. By testing for collision of those rays with the target stimuli, higher-level events can be output, e.g. *Fixation of 300ms on room 101*, which can be analysed easily without the need for manual annotation.

The whole interaction can be analysed using either tracking solution, except for the finger tracking. Both systems are capable of real-time analysis: The outside-in approach has a frame rate of 60 Hz, but is restricted to the 30 Hz input of gaze di-

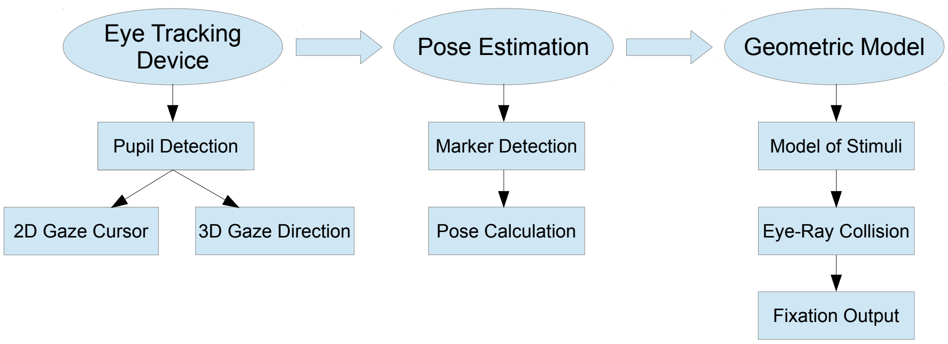

Fig. 3 The different steps of our inside-out tracking based approach: All gaze-related computations are done by the eye tracking device. The scene-camera image is used to estimate the pose of the device using marker tracking. Gaze and pose information are combined to identify fixations on target stimuli by means of ray-casting in a 3D scene model.

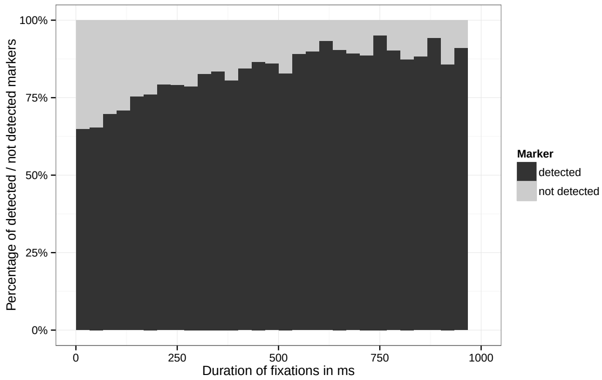

Fig. 4 The figure shows the percentage of detected markers depending on the duration of fixations.

rections from the eye tracker. The inside-out approach runs at the 24 Hz of the eye tracker's scene camera. It has, however, a mean delay of 379 ms (sd: 90 ms) compared to the outside-in approach, which is not sufficient for real-time interaction, but gaze and pose data are synchronized and thus the accuracy of the fixation classification is not affected by latency. In a comparison study, the inside-out solution could estimate the pose in 75.96% of the analysed frames, the outside-in solution covers all. In the remaining frames no markers were present or the image of the scene camera was smeared, caused by quick head movements. During fixations, however, the head remains relatively stable and losses of markers are less likely. Indeed, Figure 4 shows that the percentage of detected markers increases during longer fixation. The accuracy of the inside-out tracking is good, compared to the highly accurate outside-in tracking we have an average deviation of 1.11 cm (sd: 0.69 cm) in the 3D position and 1.39 degrees (sd: 0.68 degrees) in the orientation.

To conclude: The presented model-based approach combined with the inside-out tracking, which does not require additional expensive equipment except for the eye tracker itself, allows for an automatic analysis of gaze data in our scenario.

ACKNOWLEDGMENTS: This work has partly been funded by the SFB 673 ”Alignment in Communication”, in the project ”Interaction Space”.

References

1. Holthaus, P., Wachsmuth, S. (2012). Active Peripersonal Space for More Intuitive HRI. International Conference on Humanoid Robots: 508513.
2. Imai, M., Ono, T., & Ishiguro, H. (2003). Physical relation and expression: joint attention for human-robot interaction. IEEE Transactions on Industrial Electronics, 50(4):636643.
3. Prablanc, C., Echallier, J., Komilis, E., Jeannerod, M. (1979). Optimal response of eye and hand motor systems in pointing at a visual target. Biological cybernetics 124:113124.
4. Pfeiffer, T. (2012). Measuring and visualizing attention in space with 3D attention volumes. Proceedings of the Symposium on Eye Tracking Research and Applications:18.

DOI: 10.2390/biecoll-saga2013_8

In Pfeiffer and Essig (2013): Proceedings of the First International Workshop on Solutions for Automatic Gaze-Data Analysis 2013 (SAGA 2013). DOI: 10.2390/biecoll-saga2013_0

3D Gaze Recovery in Large Environments Using Visual SLAM

Lucas Paletta, Katrin Santner, Albert Hofmann, Georg Thallinger

DIGITAL – Institute for Information and Communication Technology
JOANNEUM RESEARCH Forschungsgesellschaft mbH, 8010 Graz, Austria

1 Introduction

This work describes a multi-component vision system that enables pervasive mapping of human attention. The key contribution is that our methodology enables full 3D recovery of the gaze pointer, human view frustum and associated human centered measurements directly into an automatically computed 3D model. We apply RGB-D SLAM and descriptor matching methodologies for the 3D modeling, localization and fully automated annotation of ROIs (regions of interest) within the acquired 3D model. This methodology enables fully automated processing of human attention, without artificial landmarks, in indoor natural environments.

2. Implementation

This work presents a computer vision system methodology that, *firstly*, enables to precisely estimate the 3D position and orientation of human view frustum and gaze [1] and from this enables to precisely analyze human attention in the context of the semantics of the local environment (objects [10], signs, scenes, etc.). *Secondly*, the work describes how ROIs (regions of interest) are automatically mapped from a reference video into the model and from this prevents from state-of-the-art laborious manual labeling of tens / hundreds of hours of eye tracking video data. This provides a scaling up of nowadays still small sketched attention studies. With the presented methodology, extended natural environments, such as shop floor departments, analysis of navigation guidance, and human-robot interaction, can be studied first time in large scale, statistically significant usability studies.

For a spatio-temporal analysis of human attention in the 3D environment, we firstly build a spatial reference in terms of a three-dimensional model of the environment using RGB-D SLAM methodology [2]. Secondly, the user's view is gathered with eye tracking glasses (ETG) within the environment and localized from extracted local descriptors [3]. Then ROIs are marked on imagery and automatically detected in video and further mapped into the 3D model. Finally, the distribution of saliency onto the 3D environment is computed for further human attention analysis, such as, evaluation of the attention mapping with respect to object and scene awareness. Saliency information can be aggregated and further evaluated in the frame of user behaviors of interest. The performance evaluation of the presented methodology firstly refers to results from a dedicated test environment [4] demonstrating very low projection errors, enabling to capture attention on daily objects and activities (package logos, cups, books, pencils).

3 Gaze recovery in 3D without artificial markers

Human Attention Analysis in 3D. 3D information recovery of human gaze has in principle been targeted by Munn et al. [2] who introduced monocular eye-tracking and triangulation of 2D gaze positions of subsequent key video frames, obtaining observer position and gaze pointer in 3D with angular errors of ≈3.8°. Pirri et al. [3] achieved accuracy indoors about ≈3.6 cm at 2 m distance to the target compared to our ≈0.9 cm [1]. Previous attempts focused on single 3D point recovery. Our approach maps fixation within a 3D environment model with the possibility of real-time tracking of attention with mass marketed eye-tracking hardware.

Visual Map Building and Camera Pose Estimation. For realistic environment modeling we make use of an RGB-D sensor providing per pixel color and depth information at high frame rates. Our environment consists of a sparse point-cloud, where each landmark [4] is attached for data association during pose tracking. Estimated camera poses are stored in a 6DOF manner. Incremental camera pose tracking assuming an already existing map is done by keypoint matching followed by a least-square optimization routine minimizing the reprojection.

Densely Textured Surface Generation. For realistic environment visualization, user interaction and subsequent human attention analysis, a dense, textured model of the environment is constructed. Depth images are integrated into a 3D occupancy grid [5] using the previously corrected camera pose estimates.

3D Gaze Recovery from Monocular Localization. To estimate the proband's pose, SIFT keypoints are extracted from ETG video frames and a *full 6DOF pose* is estimated using the perspective n-Point algorithm [6].

Automated 3D Annotation of Regions of Interest. Annotation of ROIs in 2D or even 3D information usually causes a process of massive manual interaction. In order to map objects of interests, such as, logos, package covers, etc. into the 3D model, we first use logo detection in the high resolution scanning video to search for occurrences of predefined reference appearances, using vocabulary trees [4].

Semantic Mapping of Attention. The automatic detection of ROIs in 3D enables statistical evaluations, such as on ROIs called AOI hit, which states for a raw sample or a fixation that its coordinate value is inside the ROI [7]. From this, the dwell time distribution for ROIs can be plotted over all participants, and some of the captured fixations are related to human object recognition which is known to trigger from 100 ms of observation / fixation [8].

4 Experimental results

Eye Tracking Device. The mass marketed SMI™ eye-tracking measure the gaze pointer for both eyes with 30 Hz. The gaze pointer accuracy of 0.5°–1.0° and a tracking range of 80°/60° horizontal/vertical assure a precise localization of the human’s gaze in the HD 1280x960 scene video with 24fps. We recorded data on a shop floor covering an area of about 8x20m² (Figure 1). We captured 2366 RGB-D images, reconstructed the model from 41700 natural visual landmarks.

5 Conclusions and future work

We present a complete system for (i) wearable data capturing, (ii) automated 3D modeling, (iii) automated recovery of human pose and gaze, and (iv) automated ROI based semantic interpretation of human attention. The presented system is a significant step towards a mobile mapping framework [9] for quantitative analysis of human attention measures [7,10] in natural environments (Figure 2). Future work will focus on improved tracking of the human pose across image blur and uncharted areas as well as study human factors in the frame of stress and emotion in the context of the 3D space.

Figure 1. Hardware (a) for the 3D model building process (Kinect and HD camera), (b) study with packages, (c) 3D environment model, a large shop floor.

ACKNOWLEDGMENTS: The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 288587 (MASELTOV) as well as from the Austrian FFG via contracts No. 832045 (FACTS) and No. 836270 (EVES), and by the Provinc. Gov. of Styria (NeoAttrakt). We kindly thank INTERSPAR Graz and CITYPARK GmbH for the permission to capture the data.

References

1. Paletta, L., Santner, K., Fritz, G., Mayer, H., & Schrammel, J.: 3D Attention: Measurement of Visual Saliency Using Eye Tracking Glasses, *Proc. CHI 2013*.
2. Munn, S. M., & Pelz J. B.: 3D point-of-regard, position and head orientation from a portable monocular video-based eye tracker. *Proc. ETRA 2008*, pp. 181-188.
3. Pirri, F., Pizzoli, M., & Rudi, A.: A general method for the point of regard estimation in 3D space. *Proc. CVPR 2011*, pp. 921-928.
4. Nistér, D. & Stewénius, H.: Scalable Recogn. with a Vocabulary Tree, Proc. CVPR 2006.
5. Marks, T. K., Howard, A., Bajracharya, M., Cottrell, G. W. & Matthies, L.: Gamma-SLAM: Using stereo vision and variance grid maps for SLAM, *Proc. ICRA 2008*.
6. Lepetit V., Moreno-Noguer F. and Fua P.: EPnP: An Accurate O(n) Solution to the PnP Problem, *International Journal of Computer Vision*, pp. 155-166, 2009.
7. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and van de Weijler, J.: *Eye Tracking*, Oxford University Press, 2011, pp. 187.
8. Grill-Spector, K. and Sayres, R. Object Recognition: Insights From Advances in fMRI Methods, Current Direct. *Psychol. Science*, Vol. 17, No. 2. 2008, pp. 73-79.
9. Paletta, L., Santner, K., Fritz, G., Hofmann, A., Lodron, G., Thallinger, G., and Mayer, H. FACTS - A Computer Vision System for 3D Recovery and Semantic Mapping of Human Factors, *Proc. ICVS 2013*, Springer-Verlag, LNCS 7963, pp. 62-72.
10. Fritz, G., Seifert, C., Paletta, L., and Bischof, H. (2005). Attentive object detection using an inform. theoret. saliency measure, *Proc. WAPCV 2004*, Springer-Verlag, LNCS 3368.

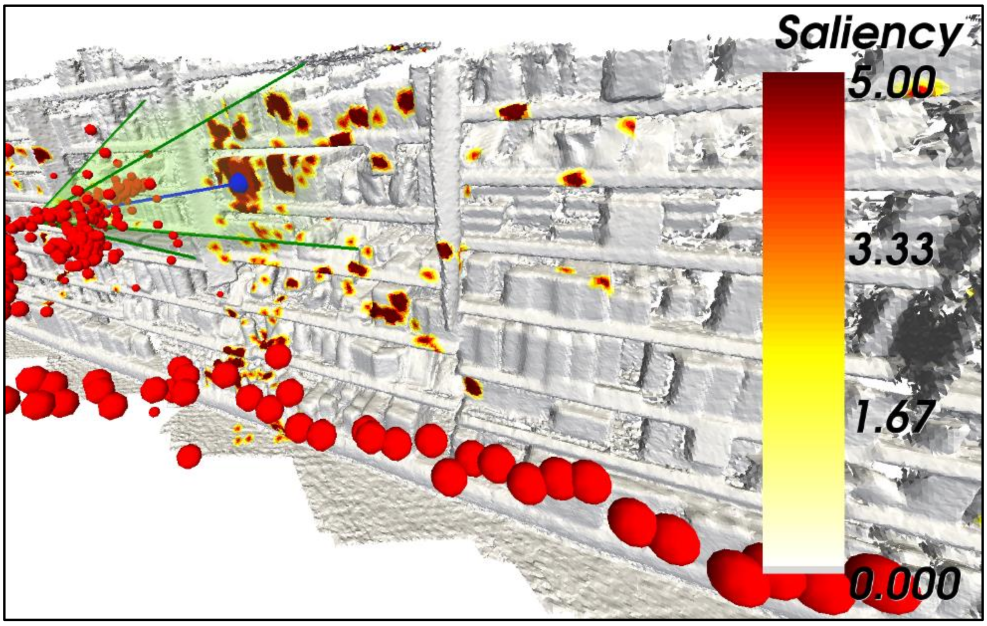

Figure 2. Mapping of user saliency onto the acquired 3D model and automated recovery of the trajectory of ETG camera positions (spheres), as well as recovery of frustum (green lines) and gaze pointer (blue).

Part IV
Poster Session

DOI: 10.2390/biecoll-saga2013_9

In Pfeiffer and Essig (2013): Proceedings of the First International Workshop on Solutions for Automatic Gaze-Data Analysis 2013 (SAGA 2013). DOI: 10.2390/biecoll-saga2013_0

Visual Context and Language Comprehension: The Resilience of the Recent – Event Preference

Dato Abashidze, Pia Knoeferle, & Maria Nella Carminati
Cognitive Interaction Technology Excellence Cluster, Bielefeld University, Bielefeld, Germany
dabashidze@cit-ec.uni.bielefeld.de

Eye tracking results suggest that people prefer to look at recently depicted over possible future events during spoken sentence comprehension. In a study by Knoeferle et al., 2011 (Exp 2) participants saw an actor performing an action (e.g. sugaring strawberries) and then listened to a sentence (NP1-VERB-ADV-NP2). The sentence referred either to the just depicted action (sugaring strawberries) or a possible future action (sugaring pancakes) (sentence: Lit transl. 'The experimenter sugared just now the strawberries/sugars in a moment the pancakes'). The NP1-VERB fragment was referentially ambiguous between the "recent", and another, equally plausible "future" event (performed post-sentence). Recent and future events were shown equally often. At the verb, participants inspected the recent over the future event target. Looks to the future target only rose as the future tense ADV cued the future event. Importantly, throughout the sentence an overall preference for the recent vs. future target prevailed, irrespective of tense.

Two visual-world eye-tracking experiments (each N=32) examined the robustness of this preference by introducing a frequency bias in favor of the future event. In Experiment 1, 80% of all sentences referred to a future event, and filler trials showed a future event only. In Experiment 2, 75% of all sentences referred to a future event and the fillers showed both a recent and future event. All experimental trials showed a past and future event, with half referring to a future event.

If our biases are effective, inspection of the future event target should happen earlier, and the overall recent target preference should reverse or drastically decrease. Results of both studies show that during the verb participants preferentially inspected the object that had recently been acted upon (the recent target). At the same time, both frequency manipulations were effective and resulted in earlier inspection of the future event target (during the late VERB). However the overall preference for the recent target was still present until sentence end.

We suggest a philosophical explanation for this bias relating to the epistemic interpretation of past and future tense statements. Past tense statements are based on stronger evidence about the truth of events (they can be verified) than future tense assertions (McFarlane 2003, see Staub & Clifton, 2011). For future events, one can predict at most their likelihood, based on previous experience; however even a highly likely future event appears to be insufficient to overcome the evidence of a past event.

References:

Knoeferle, P., Carminati, M.N., Abashidze, D., & Essig, K. (2011). Preferential inspection of recent real-world events over future events: evidence from eye tracking during spoken sentence comprehension.
McFarlane, J. (2003). Future contingents and relative truth.
Staub, A., & Clifton, C.,Jr. (2011). Processing effects of an indeterminate future.

DOI: 10.2390/biecoll-saga2013_10

In Pfeiffer and Essig (2013): Proceedings of the First International Workshop on Solutions for Automatic Gaze-Data Analysis 2013 (SAGA 2013). DOI: 10.2390/biecoll-saga2013_0

Location-based Online Identification of Objects in the Centre of Visual Attention using Eye Tracking

Kai Harmening, Thies Pfeiffer

1 Introduction

Modern mobile eye trackers calculate the point-of-regard relatively to the current image obtained by a scene-camera. They show where the wearer of the eye tracker is looking at in this 2D picture, but they fail to provide a link to the object of interest in the environment. To understand the context of the wearer's current actions, human annotators therefore have to label the recorded fixations manually. This is very time consuming and also prevents an online interactive use in HCI.

A popular scenario for mobile eye tracking are supermarkets. Gidlöf et al. (2013) used this scenario to study the visual behaviour in a decision-process. De Beugher et al. (2012) developed an offline approach to automate the analysis of object identification. For usage of mobile eye tracking in an online recommender system (Pfeiffer et al., 2013), that supports the user in a supermarket, it is essential to identify the object of interest immediately. Our work addresses this issue by using location information to speed-up the identification of the fixated object and at the same time making detection results more robust.

2 Approach

The purpose of our framework is to identify the object in the focus of visual attention based on the scene-camera image and the point-of-regard in a mobile interaction. In our application example, we want to identify the fixated cereal package in a decision situation in a supermarket.

Kai Harmening, kharmening@techfak.uni-bielefeld.de
SFB 673, Faculty of Technology, Bielefeld University, Germany

Thies Pfeiffer, tpfeiffe@techfak.uni-bielefeld.de
CITEC – Cognitive Interaction Technology Center of Excellence, Bielefeld University, Germany

Figure 1: Example of an annotated panorama image of a shelf. The objects relevant for the study are manually identified once for the setup using shapes (here rectangles). The annotation is done using a general purpose vector graphics format (SVG).

We use a hierarchical database to store our objects of interest while preserving topological information as far as possible. For the supermarket, the top-level elements are panorama-pictures taken from the relevant shelves. The second level holds all relevant objects, here cereal packages, which are pre-labelled by human annotators in the panorama picture (see Figure 1). Using this process, we get image examples of relevant objects together with topological information. For each extracted image, multi-dimensional fingerprints are computed (computer vision, feature extraction) using a set of pre-configured methods. Currently we support SIFT, SURF and ORB.

To find the focused object, an area in the scene-image around the point-of-regard is extracted. Based on the assumption that the user has not moved very much since the last frame (i.e. within 40 ms), the search process starts within the neighbourhood of the previously fixated and successfully identified object. This step uses the topological information encoded in the hierarchical database, here the panorama images, to significantly speed up the search process. Only if this approach is not successful, the search domain is expanded by climbing up the hierarchy. Finally, the found object and useful information like the name of the product is visualized (see Figure 2).

3 Preliminary Results

In a first evaluation of the system, the area in the panorama-picture fixated by the participant could be successfully identified in 416 of 470 consecutive images (88.51%). An object could be classified in 336 of these images (71.49%) and in 329 images (70%) the classified object was the one being focussed.

4 Discussions and Further Work

The presented approach uses a database of objects to identify fixated objects of interest. Using information about the topology of the objects, the system is able to classify sequences of fixations in the same area, e.g. within one shelf, in real-time. Only when the environment completely changes, a re-localization based on computer vision may take some time. To solve this problem, the database is prepared to support other means of localizations, such as GPS.

The presented approach can be used with online and offline data. While our current focus is on the online capabilities for human-computer interaction the offline modus works even better because there is more time for the computer vision processes and when the position in the environment is lost, the full database can be searched in the worst case. This is not feasible in the online approach.

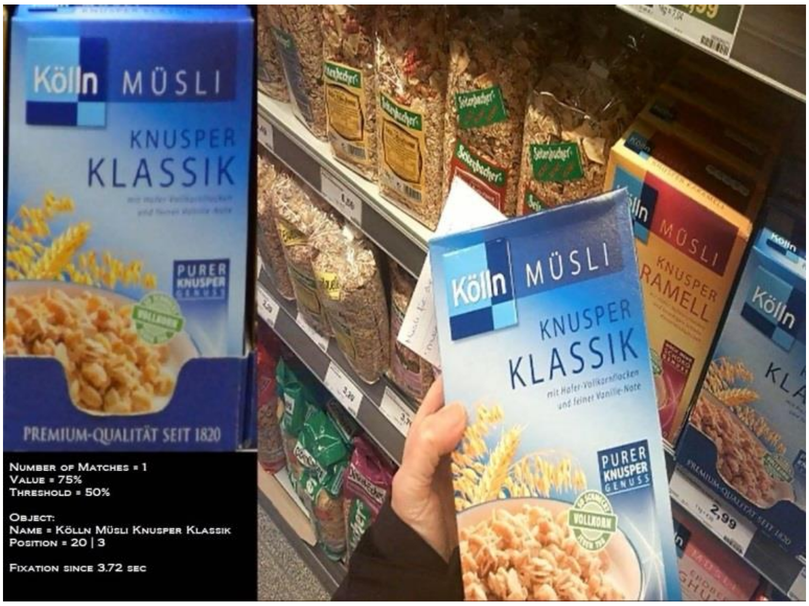

Figure 2: Application of the present work in a supermarket, Right: Consumer with an object of interest, Left: Found object in the database with information

References

1. Gidlöf K, Wallin A, Dewhurst R, Holmqvist K (2013). Using Eye Tracking to Trace a Cognitive Process: Gaze Behaviour During Decision Making in a Natural Environment. Journal of Eye Movement Research 6(1):3, 1-14
2. De Beugher S, Younes I, Brône G, Goedemé T (2012). Automatic analysis of eye-tracking data using object detection algorithms. UbiComp ’12 677-680
3. Pfeiffer T, Pfeiffer J, Meißner M (2013). Mobile Recommendation Agents Making Online Use of Visual Attention Information at the Point of Sale. Gmunden Retreat on NeuroIS

DOI: 10.2390/biecoll-saga2013_11

In Pfeiffer and Essig (2013): Proceedings of the First International Workshop on Solutions for Automatic Gaze-Data Analysis 2013 (SAGA 2013). DOI: 10.2390/biecoll-saga2013_0

Searching for salient locations in topographic maps

Vassilios Krassanakis, Alexandra Lelli, Ismini-Eleni Lokka, Vassiliki Filippakopoulou and Byron Nakos

1 Introduction

Maps have proven to be of importance in everyday life tasks as well as in scientific studies in different research fields. A basic classification of maps includes two categories, static and animated maps. In animated maps the element of motion is dominant. Psychological literature suggests that motion is detected in a primary stage of vision during a visual scene perception (e.g. Wolfe and Horowitz 2004). The attributes that are processed pre-attentively (or in a bottom-up process) can be used to control the selective attention (Wolfe 2005). Bottom-up processes are occurred in any map reading task (Lloyd 2005). Hence, pre-attentive features have to be considered in map design (Sluter 2001). Example that indicates the importance of an immediate response from map users can be an animated map for navigation. In this case the symbols that are used must be processed rapidly for decision making. The evaluation of the effectiveness of graphic elements that are used to represent geographic phenomena might lead to crucial suggestions for map design. Map readers' reaction is crucial for such evaluation.

In a previous study, an eye tracking experiment was conducted in order to investigate the reaction of map users in two different variables of map design; duration and rate of change (Krassanakis et al. 2013). The experiment was based on a bottom-up process (free viewing conditions). In this study the duration of changes during the observation of a moving point symbol on two different backgrounds was tested, while the magnitude of change was held constant. Both a blank background and a topographic map were used in order to examine how the detection of the moving symbol is affected due to the variety of spatial information. Despite the fact that moving targets pop out among stationary distractors (Wolfe 2000), the results indicated that the detection of the moving symbol is affected from the huge variety of graphic elements.

Extending the results of the experiment cited above, the principal aim of the present study is to investigate the subjects' post-reaction after the detection of the moving point symbol. The subjects' post-reaction is meaningful in the case of the topographic map. The results of eye movement analysis are compared with the outcomes of the application of three saliency models (Itti et al. 1998, Harel et al. 2006, Hou et al. 2012). The comparison is essential in order to examine the possibility of applying saliency models, which have been developed in order to predict fixations in complex physical images, on topographic maps.

2 Experimental design

Subjects are asked to observe a computer monitor during some stimuli projection without requiring the completion of any task. That means that the experiment is performed under free viewing condition (bottom-up process). The stimuli consist of a moving point symbol on two different backgrounds; a blank background and a topographic map. Different durations of the moving symbols are examined. Totally 32 subjects participated in the experiment for the observation of 98 different stimuli. A full description about the design of the stimuli can be found in the study described above (Krassanakis et al. 2013). Additionally, more information about the eye tracker equipment and the laboratory setup is presented by Krassanakis et al. (2011) and Krassanakis (2013).

The performance analysis includes fixations' computation for each visual scene. In order to examine fixations that correspond to locations in the backgrounds, which are not linked to the moving point symbol, a spatial buffer is applied to separate fixations referring to these locations from fixations dedicated to the moving point symbol. Heatmaps based on fixations' durations and not on fixations' distribution are used to indicate a qualitative comparison of the relative dwell times on observed locations for both backgrounds. Heatmaps can be considered more appropriate than an approach that examines different regions (ROIs analysis) as the study aims to test the fixations in overall rather than in particular areas. Heatmaps have already used in the analysis of eye tracking experiments performed in cartographic research (e.g. Coltiken et al. 2009). Moreover it is difficult to define a way to relate the density of different information with the definition of discriminant regions on the topographic map.

3 Results

The qualitative results of the analysis are presented in Figure 1. Specifically, on the blank background the heat-map reveals a homogeneous distribution of fixations, as it was expected. On the contrary, fixations are clustered at specific locations in the case of topographic map. Additionally, the total number of fixations on the topographic map is significantly larger than the number on blank background. This phenomenon indicates that there is a great influence on the perception of the topographic map' s graphic elements. Moreover, the application of saliency models demonstrates a sole common salient area of the topographic map. This area differs from the salient locations suggested by the heatmaps. Especially, the red regions correspond to the common areas that are indicated by the application of the saliency models, while the green regions reveal the most salient locations as they are indicated by the heatmap visualization. This means that the tested salient models cannot be applied directly in order to indicate salient locations.

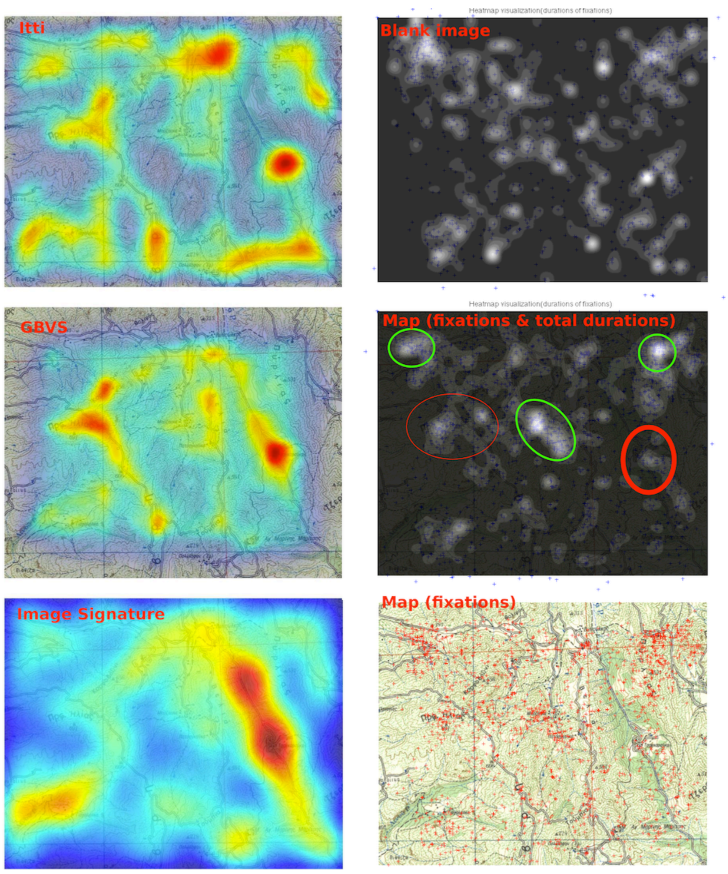

Fig 1: The results of saliency models' performance (left) in comparison with the most fixated areas after the moving point detection (right). Fixations are presented as crosses. Green areas correspond to the most salient locations according to fixations' durations while red areas correspond to the common salient areas indicated by the application of the salient models.

4 Conclusion

Eye tracking methodology has already been used in cartographic experiments where the perception of topographic maps is examined (e.g. Brodersen et al. 2002). Moreover, in the study of Garlandini and Fabrikant (2009) eye tracking methodology is combined with the application of saliency models in order to evaluate the effectiveness and the efficiency of graphic tools for map design. In the present study, the salient locations after the detection of a moving point on a topographic map (which is an artificial representation of the real world and not a physical image) are examined. The findings of researches related to visual saliency and models are very important for the validation of map design (Fabrikant et al. 2010). According to the results of the experiment the salient locations of the topographic map, as they are indicated from the analysis, are different from the common locations revealed by the application of the saliency models.

5 Future work

The present study is a first attempt to examine the post-reaction of map readers after the detection of a moving point symbol. Additionally, the possibility of applying three saliency models in topographic maps is examined. The analysis is mainly based on a qualitative approach. The performance of a quantitative analysis using specific metrics (e.g. number of fixations, mean fixations, number of saccades etc.) would be much more promising in order to evaluate the influence in map reading process. Additionally the same methodology can be used in order to examine corresponding reactions of different user groups on different types of cartographic backgrounds (e.g. maps for navigation, weather maps etc.).

References

1. Brodersen L., Andersen H.H.K., Weber S. (2002). Applying Eye-Movement Tracking fo the study of Map Perception and Map Design. Natinal Survey and Cadastre, Denmark 9(2).
2. Coltiken A., Heil B., Garlandini S., Fabrikant S.I. (2009). Evaluating the Effectiveness of Interactive Map Interface Designs: A Case Study Integrating Usability Metrics with Eye-Movement Analysis. In: Cartography and Geographic Information Science, 36(1): 5-17.
3. Fabrikant S.I., Hespanha S.R., Hegarty M. (2010). Cognitively Inspired and Perceptually Salient Graphic Displays for Efficient Spatial Inference Making. In: Annals of the Association of American Geographers, 100(1): 13-29.
4. Garlandini S., Fabrikant S. (2009). Evaluating the Effectiveness and Efficiency of Visual Variables for Geographic Information Visualization. In: Hornsby K.S. et al. (eds). 9th International Conference COSIT 2009 (pp. 195-211), Springer-Verlag Berlin.
5. Harel J., Koch C., Perona P. (2006). Graph-Based Visual Saliency. In: Advances in neural information processing systems, 545-552.
6. Hou X., Harel J., Koch C. (2012), Image Signature: Highlighting Sparse Salient Regions, IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(1): 194-201.
7. Itti L. Koch C., Niebur E. (1998). A model of saliency-based visual attention for rapid scene analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(11): 1254-1259.
8. Krassanakis V., Filippakopoulou V., Nakos B. (2011). An Application of Eye Tracking Methodology in Cartographic Research, Proc. of the EyeTrackBehavior 2011(Tobii), Frankfurt.
9. Krassanakis V., Lelli A., Lokka I.E., Filippakopoulou V., Nakos B. (2013). Investigating dynamic variables with eye movement analysis. Proc. of the 26th International Cartographic Conference, Dresden.
10. Krassanakis V. (2013) Exploring the map reading process with eye movement analysis. Proc. of the 1st International Workshop on Eye Tracking for Spatial Research (In conjunction with COSIT 2013), Scarborough.
11. Lloyd R. (2005). Attention on Maps. In: cartographic perspectives, 52: 28-57.
12. Sluter R.S. (2001). New theoretical research trends in cartography. Revista Brasileira de Cartografia, 53: 29-37.
13. Wolfe J.M. (2000). Visual Attention. In: De Valois K.K. (ed.). Seeing (pp. 335-386) Academic Press, San Diego.
14. Wolfe J.M. (2005). Guidance of visual search by preattentive information, Neurobiology of attention, 101-104.
15. Wolfe J.M., Horowitz T.S. (2004). What attributes guide the deployment of visual attention and how do they do it?. Nature Reviews Neuroscience, 5(6): 495-501.

DOI: 10.2390/biecoll-saga2013_12

In Pfeiffer and Essig (2013): Proceedings of the First International Workshop on Solutions for Automatic Gaze-Data Analysis 2013 (SAGA 2013). DOI: 10.2390/biecoll-saga2013_0

Embodied attention for gaze analysis in daily life activities

William Welby Abbott, Andreas A. C. Thomik, and Aldo Ahmed Faisal

Significant progress has been made towards understanding the role of vision in the control of action, particularly in the last 25 years with the emergence of mobile eye-tracking solutions which measure gaze in relation to the subject's scene view (2D video feed) (Eileen, 2011). In addition, increasing emphasis has been placed on the ecological validity of gaze studies, taking them out of the lab and into the "wild" (Hayhoe & Ballard, 2005; Kingstone et al., 2003; Land & Tatler, 2009). These steps have challenged many of the fundamental results of static eye-tracking experiments and are changing the way we look at eye-movement ethology. However unlike static experiments they require painstaking annotation of gaze position in relation to the scene, objects and the body. Recent developments have been made to automate elements of this process by estimating gaze with respect to the 3D world (Paletta, Santner, & Fritz, 2013) and physical objects in the world (Essig et al., 2012). However, we are also interested in the relationship between eye-movements and full body kinematics in unconstrained natural tasks.

In this work we present a new paradigm for probing this relationship: annotating eye-movements with 3D full-body kinematics in every-day natural tasks. We use a portable eye-tracker (SMI Eye Tracking Glasses, Sensomotoric Instruments, Teltow Germany) in combination with a portable full body motion capture suit, measuring 51 degrees of freedom (DOF) from the body with 16 inertial measurement units (IGS-180 Animazoo, Brighton UK) and 22 DOF from the right hand (Cyberglove 1) and 18 from the left (Cyberglove 3, Cyberglove Systems, San Diego California USA). All tracking equipment is marker less and thus allows extensive behavioral monitoring "in the wild". The eye-tracker data is processed post-hoc to give 3D gaze position relative to the subjects head using our own software (Abbott & Faisal, 2011; Abbott & Faisal, 2012) and the motion capture data. The experiment had 3 natural scenarios: breakfast time, evening chores and navigation. During each scenario, task level instructions were given and subjects conducted activities such as laying the table for breakfast, sweeping the floor and walking to a specified location in the building. During the experiment we had a manual online annotator to separate tasks and sub tasks, and an usher to guide the subject from task to task.

William Welby Abbott, Andreas A. C. Thomik
Imperial College London, Dept. Bioengineering, UK

Aldo Ahmed Faisal
Imperial College London, Dept. of Computing and Dept. Bioengineering, UK

This data intensive approach, in ecologically valid environments gives detailed information of the simultaneous output from the motor plant and thus allows us to extract meaningful statistics from the data without the need of hand coding eye-movement data frame by frame. This level of data richness (98 DOF), coupled with the extensive recordings of natural behavior (total>30 hrs) will allow us to answer general questions about the natural statistics of movement, highlighting variation and consistencies, both inter and intra subject, to be understood and verified quantitatively in natural tasks.

References

1. Abbott, W. W., & Faisal, A. A. (2011). Ultra-low cost eyetracking as an high-information throughput alternative to BMIs. BMC Neuroscience, 12(Suppl 1), P103.
2. Abbott, W. W., & Faisal, A. A. (2012). Ultra-low-cost 3D gaze estimation: an intuitive high information throughput compliment to direct brain–machine interfaces. Journal of Neural Engineering, 9(4), 046016.
3. Eileen, K. (2011). Eye movements: The past 25 years. Vision Research, 51(13), 1457–1483.
4. Essig, K., Dornbusch, D., Prinzhorn, D., Ritter, H., Maycock, J., & Schack, T. (2012). Automatic analysis of 3D gaze coordinates on scene objects using data from eye-tracking and motion-capture systems. In Proceedings of the Symposium on Eye Tracking Research and Applications (pp. 37–44). New York, NY, USA: ACM.
5. Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. Trends in Cognitive Sciences, 9(4), 188–194.
6. Kingstone, A., Smilek, D., Ristic, J., Friesen, C. K., & Eastwood, J. D. (2003). Attention, Researchers! It Is Time to Take a Look at the Real World. Current Directions in Psychological Science, 12(5), 176–180.
7. Land, M. F., & Tatler, B. W. (2009). Looking and acting: vision and eye movements in natural behaviour. Oxford University Press, USA.
8. Paletta, L., Santner, K., & Fritz, G. (2013). An Integrated System for 3D Gaze Recovery and Semantic Analysis of Human Attention. arXiv preprint arXiv:1307.7848.

DOI: 10.2390/biecoll-saga2013_13

In Pfeiffer and Essig (2013): Proceedings of the First International Workshop on Solutions for Automatic Gaze-Data Analysis 2013 (SAGA 2013). DOI: 10.2390/biecoll-saga2013_0

Large-field study of gaze based ultra low-cost, non-invasive task level BMI

William Welby Abbott and Aldo Ahmed Faisal

The pursuit of an effective brain machine interface (BMI) holds the hope to enable patients with severe motor disorders to interact with their surroundings. Different approaches can be categorised as non-invasive cortical interfaces (e.g. EEG), invasive cortical interfaces, e.g. implanted multi-electrode arrays (MEA), or non-invasive and non-cortical interfaces (e.g. EMG). The clinical aim, however, remains the same: to extract an intention signal from a patient, for which conventional approaches such as joystick, mouse movement or sip and puff control are not possible. Current approaches however come at considerable clinical and post-clinical cost (Shih et al., 2012), while posing limitations for use in daily applications due to low information transmission bandwidths. Powered continuous wheelchair control requires, about 15.3 bit/s and full-finger hand prosthetics would require 54.2 bit/s, well beyond the reported performance of current BMI approaches (EEG 1.63 bit/s, cortical multielectrode arrays (MEA) - 3.3 bit/s, non-cortical non-invasive BMIs, e.g. EMG - 2.66 bit/s (Tonet et al., 2008).

We propose a non-invasive and ultra-low cost alternative - action intention decoding from 3D gaze signals (Abbott & Faisal, 2012). This enables real-time closed-loop control that outperforms invasive (and non-invasive) BMIs in terms of cost and read-out data rates (Abbott & Faisal, 2012) - and hence has enabled robotic arm control in conjunction with other low data-rate signal sources (EMG (Corbett, Kording, & Perreault, 2013), EEG (Onose et al., 2012) or tongue-flick-switches (Buckley, Vaidyanathan, & Mayol-Cuevas, 2011). As we have previously estimated, our GT3D approach could yield bit rates up to 43 bit/s at a system cost of <30 USD and we have framed this performance within a comparison of BMI approaches in terms of cost and information throughput (Abbott & Faisal, 2012). Although eye-tracking and even low-cost eye-tracking is nothing new (Li, Babcock, & Parkhurst, 2006; San Agustin, Skovsgaard, Hansen, & Hansen, 2009; Schneider, Bex, Barth, & Dorr, 2011), we showed that gaze location, particularly in 3D and at high-data rates (matching those of eye movements) provides a real-time decodable and graded control signal that should be utilised in the BMI field.

William Welby Abbott
Imperial College London, Dept. Bioengineering, UK
Aldo Ahmed Faisal
Imperial College London, Dept. of Computing and Dept. Bioengineering, UK

Our system demonstrated a clear improvement on these low-level measures of BMI performance, but such technical measures mask the complexities of learning to use and operate BMIs in the clinic and daily-life. Thus building on our previous work, we present here a large field study (N=2224 subjects) that aimed to understand how efficient our approach is at allowing subjects, from first use, to operate our BMI on the Pong BMI benchmark task. To achieve this we built an eye controlled arcade booth that allowed members of the public to walk up and then simply "sit", "scan" (calibrate) and "play". Within the first 30 seconds of first time use, the majority of subjects were able to successfully play the arcade game pong against a computer. Subjects made on average 6.6 +- 6.2 ball returns compared to the chance level of 2.6+-2.5 obtained without input (mean +- SD). Almost 20% of players even managed to beat the computer, despite having never used their eye-movements as a control input. This performance was achieved with members of the public at a scientific engagement event, not in stringent lab conditions and with minimal system calibration (30s) and negligible user control learning (5s countdown before ball released). This demonstrates the intuitive nature of gaze control and thus the clinical applicability of our approach.

References

1. Abbott, W. W., & Faisal, A. A. (2012). Ultra-low-cost 3D gaze estimation: an intuitive high information throughput compliment to direct brain–machine interfaces. Journal of Neural Engineering, 9(4), 046016. doi:10.1088/1741-
2. Buckley, M., Vaidyanathan, R., & Mayol-Cuevas, W. (2011). Sensor suites for assistive arm prosthetics. In Computer-Based Medical Systems (CBMS), 2011 24th International Symposium on (pp. 1 –6).
3. Corbett, E. A., Kording, K. P., & Perreault, E. J. (2013). Real-Time Evaluation of a Noninvasive Neuroprosthetic Interface for Control of Reach. IEEE Transactions on Neural Systems and Rehabilitation Engineering, Early
4. Land, M. F., & Lee, D. N. (1994). Where we look when we steer. Nature, 369(6483), 742–744. doi:10.1038/369742a0
5. Li, D., Babcock, J., & Parkhurst, D. J. (2006). openEyes: a low-cost head-mounted eye-tracking solution. In Proceedings of the 2006 symposium on Eye tracking research & applications (pp. 95–100). New York, NY, USA:
6. Mace, M et al. (2012). On the feasibility of using motor imagery EEG-based brain–computer interface in chronic tetraplegics for assistive robotic arm control: a clinical test and long-term post-trial follow-up. Spinal Cord, 50(8),
7. San Agustin, J., Skovsgaard, H., Hansen, J. P., & Hansen, D. W. (2009). Low-cost gaze interaction: ready to deliver the promises. In Proceedings of the 27th international conference extended abstracts on Human factors in computing systems (pp. 4453–4458). New York, NY, USA: ACM.
8. Schneider, N., Bex, P., Barth, E., & Dorr, M. (2011). An open-source low-cost eye-tracking system for portable real-time and offline tracking. In Proceedings of the 1st Conference on Novel Gaze-Controlled Applications (p. 8).
9. Shih, J. J., Krusienski, D. J., & Wolpaw, J. R. (2012). Brain-Computer Interfaces in Medicine. Mayo Clinic Proceedings, 87(3), 268–279.
10. Tonet, O., Marinelli, M., Citi, L., Rossini, P. M., Rossini, L., Megali, G., & Dario, P. (2008). Defining brain-machine interface applications by matching interface performance with device requirements. Journal of Neuroscience Methods, 167(1), 91–104.

DOI: 10.2390/biecoll-saga2013_14

In Pfeiffer and Essig (2013): Proceedings of the First International Workshop on Solutions for Automatic Gaze-Data Analysis 2013 (SAGA 2013). DOI: 10.2390/biecoll-saga2013_0

Ultra low-cost 3D gaze estimation: an intuitive high information throughput complement to direct Brain-Machine-Interfaces

William Welby Abbott and Aldo Ahmed Faisal

The advancement of Brain Machine Interface (BMI) technology for controlling neuromotor prosthetic devices holds the hope to restore vital degrees of independence to patients with motor disorders, improving quality of life. Unfortunately, emerging rehabilitative methods come at considerable clinical and post-clinical operational cost, beyond the means of the majority of patients (Shih et al. 2012). We present an ultra-low cost alternative: using eye-tracking. Monitoring eye movement provides a feasible alternative to traditional BMIs because the ability to control eye-movements can be retained in cases of severe traumas or pathologies in which all other motor functions are lost (Kaminski et al. 2002; Kaminski et al. 1992). Based on disease statistics, we find that within the EU alone, there were over 16 million people in 2005 (3.2% of the population) with disabilities that would benefit from such gaze based communication and control systems (Jordansen et al. 2005). Despite this, eye tracking is not widely used as control interface for movement impaired patients due to high cost, poor signal interpretation and lack of control flexibility.

We propose that tracking the gaze position in 3D rather than 2D (as is used in most gaze interaction methods (Majaranta et al., 2011)) provides a considerably richer signal for neuroprosthetic control by allowing direct interaction with the environment rather than via computer displays. We demonstrate that by using mass-produced video-game hardware that an ultra-low cost binocular eye-tracker with comparable performance to commercial systems more than 800 times as expensive is possible (Abbott & Faisal 2012). Our head-mounted system has 30 USD material costs and operates at over 120 Hz sampling rate with a 0.5-1 degree of visual angle resolution. We perform 2D and 3D gaze estimation, controlling a real-time volumetric cursor, essential for driving complex user interfaces. Our approach yields an information throughput of 43 bits/s, more than ten times that of invasive and semi-invasive BMIs that are vastly more expensive.

William Welby Abbott
Imperial College London, Dept. Bioengineering, UK

Aldo Ahmed Faisal
Imperial College London, Dept. of Computing and Dept. Bioengineering, UK

In contrast, BMI information rates from direct recording of neuronal activity are ultimately constrained by noise in the recording systems and the nervous system itself (Faisal et al. 2008). In particular physical noise sources inside central neurons (Faisal et al. 2002; Faisal 2009) and peripheral axons (Faisal et al. 2005) will limit decoding performance from limited numbers of independent neuronal sources. Thus, to compensate for noise, BMI signal decoders have to observe signals for longer periods of time, thereby increasing response latencies for direct BMIs. While these issues will be ameliorated by the steady progress of sensor quality and density, eye movements already offer a highly accurate, low-latency (and low cost) read out. This is because the brain has already evolved to minimise the role of noise and delays in eye movements, which form an aggregated output of the nervous system. The leap in readout performance (in terms of readout performance and latency) enables closed-loop real-time control of rehabilitative and domotic devices beyond what is achievable by current BMIs: e.g. it was estimated that powered wheelchair control requires, on average, 15.3 bits/second and full-finger hand prosthetics require 54.2 bits/second (Tonet et al. 2008). We have demonstrated how an ultra-low cost, non-invasive eye-tracking approach can form the basis of a real-time control interface for rehabilitative devices – making it a low-cost complement or alternative to existing BMI technologies.

References

1. Abbott, W.W. & Faisal, A.A., 2012. Ultra-low-cost 3D gaze estimation: an intuitive high information throughput compliment to direct brain–machine interfaces. Journal of Neural Engineering, 9(4), p.046016.
2. Faisal, A.A., 2009. Stochastic Simulation of Neurons, Axons, and Action Potentials. In C. Laing & G. J. Lord, eds. Stochastic Methods in Neuroscience. Oxford University Press, pp. 297–343.
3. Faisal, A.A., Laughlin, S.B. & White, J.A., 2002. How reliable is the connectivity in cortical neural networks? In Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on. pp. 1661–1666.
4. Faisal, A.A., Selen, L.P.J. & Wolpert, D.M., 2008. Noise in the nervous system. Nature Reviews Neuroscience, 9(4), pp.292–303.
5. Faisal, A.A., White, J.A. & Laughlin, S.B., 2005. Ion-Channel Noise Places Limits on the Miniaturization of the Brain's Wiring. Current Biology, 15(12), pp.1143–1149.
6. Jordansen, I. et al., 2005. Report on a market study and demographics of user population (cogain) ist-2003-511598: Deliverable 7.2),
7. Kaminski, H.J. et al., 2002. Differential Susceptibility of the Ocular Motor System to Disease. Annals of the New York Academy of Sciences, 956(1), pp.42–54.
8. Kaminski, H.J. et al., 1992. Extraocular muscles are spared in advanced duchenne dystrophy. Annals of Neurology, 32(4), pp.586–588.
9. Majaranta, P., Aoki, H., Donegan, M., Hansen, D. W., Hansen, J. P., Hyrskykari, A., & Räihä, K.-J. (2011). Gaze Interaction and Applications of Eye Tracking. IGI Global.
10. Shih, J.J., Krusienski, D.J. & Wolpaw, J.R., 2012. Brain-Computer Interfaces in Medicine. Mayo Clinic Proceedings, 87(3), pp.268–279.
11. Tonet, O. et al., 2008. Defining brain-machine interface applications by matching interface performance with device requirements. Journal of Neuroscience Methods, 167(1), pp.91–104.

Part V

Keynotes

1 Vision in Action - Ben Tatler

Successful completion of real world activities requires precise control over where and when we move our eyes. Eye movements target behaviourally relevant information in our surroundings. Behaviourally informative locations change with progress through a task, so gaze allocation must be to the right places at the right times to serve behaviour. Current computational models of fixation selection offer high explanatory power for some aspects of static scene viewing, and models of dynamic scene viewing are emerging. However, few engage with the need to consider visual selection as being fundamentally and intricately linked to action. Across a wide variety of natural tasks common fixation selection principles can be identified. These principles change the emphasis of what should be modeled and identify a need for new classes of models for explaining visual behaviour in natural task settings. A framework incorporating behavioral rewards provides a powerful potential framework for explaining eye movement behaviour, and for the development of formal models of eye guidance.

2 Eye Tracking as an Assistive Technology in Professional Training Processes for Disabled People - Ellen Schack

The Bodelschwingh Foundation (Bethel) as Europes largest institution of Christian social welfare aims to offer a high amount of assistance and support for disabled people. One crucial intention is hereby to enable people with special needs to live a self-determinant life to the greatest extent possible. Assistive technologies can help to realize this aim substancially. To develop appropriate accesses to assistive technologies for people in need different research projects were set up in a cooperation between Bethel and CITEC (Cluster of Excellence, Bielefeld University). In the first part of the talk particular tasks and institutions of Bethel are presented, whereas in a second part a short overview about Bethel-CITEC-cooperation is provided with focus on a project concerning the development of user and context specific attentive systems. Studies on attentive systems and task dependent attentional processes in disabled persons involve the measurement of attentional processes such as mobile eye-based interaction and eye-based context-awareness. These projects go far beyond usual classical assistive technologies using gaze measurements to facilitate communication with the environment by looking at control keys or cells on a computer display. Research on this topic and derived practical solutions could play an interesting role for Bethel workshops (proWerk) and may help to support self regulation and attention within a zone of proximal development for disabled people. In the final part of the talk, perspectives concerning the link between research on attentional processes and the practical needs within education and work in a caregiving system will be discussed. In this context, e.g. eye tracking is used as a technology to support a training process which enables handicapped people to enter the professional world and therefore to build up an independent existence.

3 Guidance of Visual Attention by Low- and High-Level Features in Real-World Scenes - Marc Pomplun

Several studies have shown that low-level visual features in naturalistic scenes, such as color or contrast, guide attention during inspection and search. I will present the results of our computational modeling studies showing that this guidance depends on the composition of a scene across feature dimensions. Furthermore, high-level features such as object and scene semantics are known to influence the allocation of attention as well. I will discuss our recent studies showing that even the semantic relations among scene objects are significant predictors of shifts of attention. Integrating these guidance mechanisms operating at various levels into one model may lead to a more comprehensive understanding of attentional control in real-world scenes.

Part VI

Appendix

Version: 23.10.2013

SAGA Solutions for Automatic Gaze-Data Analysis

SAGA 2013 Workshop,
October 24-25th, 2013
Bielefeld University, Germany

Programm

The workshop is hosted at the **NEW** CITEC building (Inspiration 1/33615 Bielefeld).

Thursday, October 24th		
09:30	Registration desk opens / Coffee	Foyer
10:15	Welcome	Lecture Hall
10:30	Keynote by Ben Tatler - **Vision in Action** Chair: Pia Knoeferle	Lecture Hall
12:00	Lunch	Foyer
13:00	Session 01: Gaze Analysis in Basic Research Chair: Jella Pfeiffer - *A Pilot Study on Eye-tracking in 3D Search Tasks* Ekaterina Potapova, Valsamis Ntouskos, Astrid Weiss, Michael Zillich, Markus Vincze and Fiora Pirri - *Mental representation and mental practice: The influence of imagery rehearsal on representation structure, gaze behavior and performance* Cornelia Frank, Kai Essig, and Thomas Schack - *Mobile Eyetracking for Decision Analysis at the Point-of-Sale* Martin Meißner, Jella Pfeiffer, and Thies Pfeiffer	Lecture Hall
14:15	Poster Fast Forward Chair: Pia Knoeferle	Lecture Hall
14:25	Poster Session and Demos / Coffee - *Searching for salient locations in topographic maps* Vassilios Krassanakis, Alexandra Lelli, Ismini-Eleni Lokka, Vassiliki Filippakopoulou, and Byron Nakos - *Embodied attention for gaze analysis in daily life activities* William Welby Abbott, Andreas A. C. Thomik, and Aldo Ahmed Faisal - *Location-based Online Identification of Objects in the Centre of Visual Attention using Eye Tracking* Kai Harmening and Thies Pfeiffer - *Visual Context and Language Comprehension: The Resilience of the Recent – Event Preference* Dato Abashidze, Pia Knoeferle, and Maria Nella Carminati	VR Lab 0.113
15:15	Keynote by Ellen Schack **Eye Tracking as an Assistive Technology in Professional Training Processes for Disabled People** Chair: Kai Essig	Lecture Hall
16:15	Coffee & Cake Break	Foyer
16:30	Session 02: Methods for Gaze Analysis Chair: Marc Pomplun - *Experiences from eye tracking chat-communication using the INKA-SUITE* – Andrea Kienle, Christian Schlösser and Philipp Schlieker-Steens - *Using spatial statistics to investigate within-trial correlations of human gaze positions* – Hans Trukenbrod, Simon Barthelmé, Felix Wichmann, and Ralf Engbert	Lecture Hall
17:20 – 18:00	Presentation of Commercial Solutions for Gaze Analysis Chair: Marc Pomplun - SMI Semantic Gaze Mapping – Dynamic video analysis taken to the next level – Ingmar Gutberlet (SMI) - *EyeTracking and CAPTIV physiological wireless sensors* – Stephane Folley (TEA)	Lecture Hall
19:30 - 23:00	Social Event at Bernstein, Niederwall 2, 33602 Bielefeld	City Center, close to “Jahnplatz”

Version: 23.10.2013

SAGA Solutions for Automatic Gaze-Data Analysis

SAGA 2013 Workshop,
October 24-25th, 2013
Bielefeld University, Germany

The workshop is hosted at the **NEW** CITEC building (Inspiration 1/33615 Bielefeld).

Friday, October 25th		
09:00	Keynote by Marc Pomplun **Guidance of Visual Attention by Low- and High-Level Features in Real-World Scenes** Chair: Thies Pfeiffer	Lecture Hall
10:30	Coffee & Cake Break	Foyer
10:45	Session 03: Methods for Mobile Gaze Analysis Chair: Thies Pfeiffer - *Object recognition and person detection for mobile eye-tracking research. A case study with real-life customer journeys* Stijn De Beugher, Geert Brône, and Toon Goedemé - Studying joint attention and hand-eye coordination in human-human interaction: A model-based approach to an automatic mapping of fixations to target objects Patrick Renner, Thies Pfeiffer - *3D Gaze Recovery in Large Environments Using Visual SLAM* Lucas Paletta, Katrin Santner, Albert Hofmann, and Georg Thallinger	Lecture Hall
12:00	Coffee Break	Foyer
12:15	Session 04: Applications of Mobile Gaze Analysis Chair: Peter Reuter - *What you think is what you get - Toward mind and gaze controlled assistive systems* Andrea Finke	Lecture Hall
12:40	Demo Fast Forward Chair: Kai Essig	Lecture Hall
13:00	Demo Session & Buffet see http://saga.eyemovementresearch.com/technical-demos/	Foyer / VR Lab 0.113
14:30	Discussion Session / Position Paper - Trends in Basic Research / Visions / Goals / Challenges Chair: Rebecca Förster - Benchmark / Technical Aspects / Competition Chair: Ingmar Gutberlet [2 Groups]	Seminar Rooms 1.015 1.016
15:30	Coffee	Foyer
15:45 - 16:30	Summary of the Discussion Session / Panel Official Closing	Lecture Hall
20:00	*Optional Social Event for guests leaving on Saturday* *If you are interested, please contact the reception*	*To be announced*

2 Location / Map of the CITEC building

The SAGA Workshop 2013 is hosted by CITEC in the new building with address Inspiration 1, 33615 Bielefeld. The main events of the workshop are distributed over a small number of rooms on the ground floor of the building (see Figure 1).

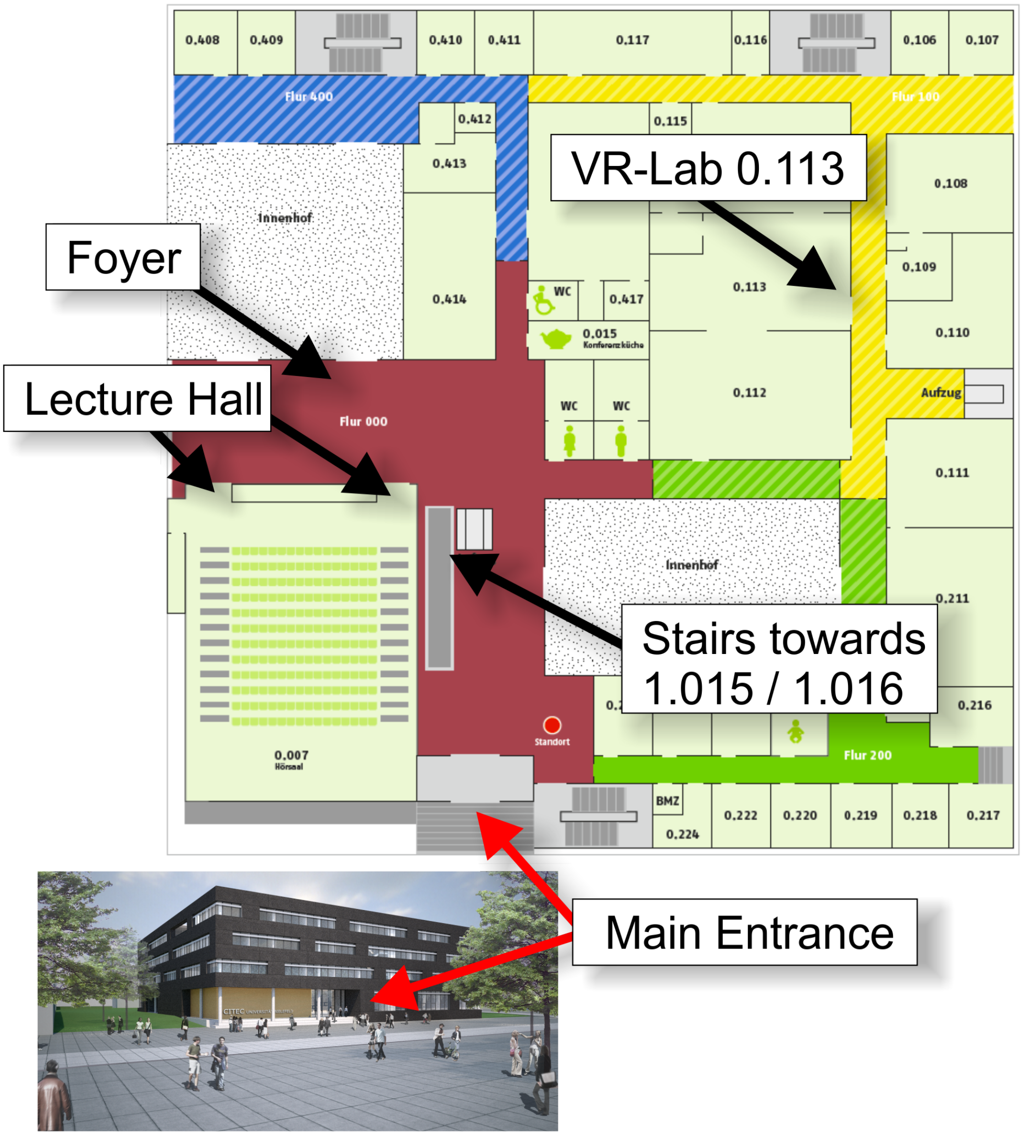

Figure 1: Map of the ground floor of the new CITEC building. The main events of the SAGA Workshop take place in the Lecture Hall, the Foyer and the VR-Laboratory in 0.113.