Universität Bielefeld

CITEC
Cognitive Interaction Technology
Center of Excellence
Bielefeld University

Proceedings of the SAGA

Solutions for Automatic Gaze-Data Analysis

Workshop 2015

@ CITEC, Bielefeld University
Bielefeld, Germany

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

Proceedings of the Second International Workshop on Solutions for Automatic Gaze-Data Analysis 2015 (SAGA 2015)

Printed on September 23rd, 2015

This is a paper-preprint of the online proceedings of the SAGA Workshop 2015. The online proceedings will be available as OpenAccess under the following URL: `http://biecoll.ub.uni-bielefeld.de/` and the DOI: 10.2390/biecoll-saga2015_0.

Editors

- Dr. Thies Pfeiffer
- Dr. Kai Essig

Address

Cluster 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

Team of Student Organizers Caroline Schwarzer, Claas Braun, Clara Matheus, Henri Neumann, Kai Harmening, Matthias Sterz, Manjunath Prasad

Reviewers Dr. Thies Pfeiffer, Dr. Kai Essig, Jun.-Prof. Dr. Pia Knoeferle, DI Dr. Lucas Paletta, Prof. Dr. Marc Pomplun, Dr. Maria Staudte, Dr. Peter Kiefer, Prof. Dr. Martin Meiner, Prof. Dr. Jacob Lund Orquin

Acknowledgment

We are thanking the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG), for funding and hosting the SAGA 2015 Workshop.

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

Contents

Foreword

When we held our first SAGA Workshop on Solutions for Automatic Gaze-Analysis at CITEC in Bielefeld in 2013, a new era of tremendous technological development was kicked off and we were introduced to the first wearable mobile products. We are talking about mobile, intuitive and intelligent applications to support humans intuitively and unobstrusively in their every-day life.

Since then we have seen extraordinary developments in a number of areas over the last two years. These developments have already started to, or in the near future will, permeate many aspects of our life. To mention just a few of them: Smart Glasses, Wearables, Internet of Things, Smart Home, natural HCI- Interfaces, Industry 4.0, Personal Coaches, mobile Assistance and Diagnostic Systems. For example, we are voluntarily carrying Wearables around and allow them to invade our personal space.

People are becoming more and more mobile. Because of the demographic change, people want to live a self-sufficient life in an age appropriate way according to their mental and physical capabilities. These developments go along with a simultaneous decrease in leisure time and number of available nursing personal, as well as an increase in life complexity. For example, the working world changes: we need optimal working environments and conditions which flexible adapt to these emerging challenges while supporting people with regard to a lifelong education. Thus, there is a need to get from mainly stationary status diagnostic systems to mobile and dynamical-adaptive action support and monitoring systems which are able to react to failures and to provide individualized support.

Additionally, the times where humans have to adapt to technical devices are over. People don't want to deal with stubborn technical devices. Most of you remember the wasted hours of time and the frustrations while we were dealing with a video recorder. The new era of technical systems shall be intuitive, intelligent, context aware and be able to individually adapt to the user: We want the machines to adapt to us and not vice versa.

Here, mobile eye tracking will play an important role. By knowing where people focus their attention and therefore by providing a "window to the mind" future systems will be able to identify problems in actual action and behaviour processes and to react when mistakes are made. This will result in cognitive systems that will be able to adapt individually, naturally and in a context sensitive way to the user's need. Eye tracking has also become more prominent in combination with virtual reality technologies. Here, virtual reality allows researchers to conduct studies in highly controlled life-size scenarios, while at the same time providing easy to analyze data, without the requirements for manual labelling of areas of interest. These developments are amplified by the growth in consumer virtual reality technology, such as Google Cardboard or Oculus Rift.

Time has brought many more mobile eye-tracking solutions, some of them even on open source basis. The eye-tracking hardware is improving continuously and the devices are becoming smaller and smaller. Recent developments will also allow for wireless transmission of the recorded data and therefore will significantly improve wearing comfort. Whereas we already have established analysis methods for screen or lab-based studies, the transfer to natural environments will provide new challenges and will require new solutions for the analysis of data gathered.

In order to address these changes and to show recent developments the SAGA Workshop 2015 once again aims to bring together researchers and practitioners to exchange thoughts about how to design, conduct, and analyze eye-tracking studies beyond 2D-desktop environments.

Join us as we embark on a quest to discover new and better ways for the analysis of eye-movement data gathered in natural environments. We are looking forward to an exciting event, with three keynote speakers from different research areas, about twenty research presentations and several technical demonstrations.

Thies Pfeiffer & Kai Essig

Part 1
Keynotes

1.1 What eye tracking researchers (dis)agree about reporting

Jacob Lund Orquin, Department of Management, Aarhus University, Aarhus, Denmark

It has previously been pointed out that eye tracking research as a field lacks a common terminology and standardization (Holmqvist et al., 2011). We put this claim to a test and ask two simple questions. First, what do eye tracking researchers report in their scientific articles and second, when directly asked do eye tracking researchers agree about what is important to report? In this talk I will provide answers to these questions as well as third and more prominent one: What should we report in scientific eye tracking papers in order to maximize transparency and reproducibility? The conclusions are based on coding of nearly 100 eye tracking studies as well as an expert survey among eye tracking researchers. The project is joint work with Susann Fiedler, Michael Schulte-Mecklenbeck, and Frank Renkewitz.

1.2 Studying gaze in spoken interaction

Maria Staudte, Computational Linguistics & Phonetics, Saarland University, Saarbrcken, Germany

Beyond the observation that both speakers and listeners rapidly fixate the visual targets of referring expressions, it has been argued that such gaze may constitute part of the communicative signal. However, listener gaze has been examined mostly in laboratory settings, typically for examining language comprehension processes in response to predefined spoken stimuli and without considering the influence of such listener behaviour onto the speaker again. I will present experimental work that explored the utility of listener gaze in natural(istic) and dynamic instruction-giving and -following scenarios. The data comprises scene view videos from the listeners perspective, their gaze data, and verbal instructions from a speaker when a) the speaker could see and use only the listeners scene view, or b) when the scene view along with the projected gaze cursor of the listener was available to the speaker. The analysis of such data is challenging and needs to deal with two problems: the dynamics of the context which require careful (manual) data annotation; and the reciprocal nature of such interactions, i.e. speech influences listener gaze and vice versa. We propose a combination of tools and analyses to (partly) overcome these issues but invite suggestions for and discussions of alternative approaches.

1.3 Visual search behaviour and expertise in high-performance environments

Mark Williams, Centre for Cognitive Neuroscience, Department of Life Sciences, Brunel University London

In recent decades there has been increasing interest in identifying the visual search behaviours underpinning expert performance both in time constrained domains involving anticipation and in more self-paced, tasks that require aiming at a target object (i.e., the so called quiet eye phenomenon). This body of research has helped increase our understanding of the processes and mechanisms underlying expert performance in sports and many other domains such as driving, law enforcement, and medicine. In this presentation, a critical overview is provided of contemporary research that explores the links between visual gaze behaviours and expert performance in both types of tasks. First, an attempt is made to highlight how gaze behaviours are shaped during performance in time-constrained situations by various task (e.g., conditions, rules, tactics) and individual (e.g., anxiety, fatigue) constraints. These constraints influence how information is picked up via the fovea and peripheral vision to guide perception and action. Second, recent research on the quiet eye phenomenon in aiming tasks such as archery is reviewed with a particular focus on identifying some of the methodological issues that impact upon its measurement as well as the theoretical mechanisms that provide an explanation for the effect. Finally, the implications of research on gaze behaviour and expertise is considered for performance enhancement in different high-performance domains with reference both to training anticipation and the quiet eye.

Part 2

Session 01: Automatic Analysis

GazeVideoAnalyser: A Modular Software Approach Towards Automatic Annotation of Gaze Videos

Kai Essig, Dato Abashidze, Manjunath Prasad and Thomas Schack

1 Introduction/Related Work

With the development of mobile eye-tracking systems over the last years, eye movements can now be recorded when humans are in sensomotoric contact with their environment. The analysis of eye movements in natural scenes yields valuable insights into the cognitive processes underlying scene perception and to explore the strategies our visual system uses in the initiation and guidance of actions [Land and Tatler 2009; Evans et al. 2012]. The application of eye-tracking techniques in real-world conditions led to a bunch of new problems amongst others: participants perceive the world from different perspectives, the lighting conditions change, relevant objects move over time and may be partially or fully occluded. Furthermore, in order to track participants' gaze positions in dynamic environments, eye-tracking systems have to meet completely diverse demands, such as careful eye-tracker calibration, tracking of eye features and gaze data analysis [Evans et al. 2012]. Whereas the hardware of the mobile systems is quite developed, automated methods for calibration, pupil- and fixation detection, and gaze analysis in field studies need further research [Evans et al. 2012].

There exist already some approaches to overcome the time-consuming and error prone manual annotation process of gaze videos. [Land and Mc Leod 2000] determine the gaze angle by combining head- and eye-in-head orientation, calculated from the movement of fixated objects and the gaze cursor in the scene video. For dynamic applications the target must be manually coded in the video. Paletta et al. [2013] generate first a 3D model of the scene by using a Kinect and the RGB-DSLAM methodology. Their multi-component vision system then outputs the 3D point of regard, the gaze positions, frustrum and saliency map overlaid onto the acquired 3D model. This approach needs a tripod to hold the Kinect and HD camera and was applied to a supermarket scenario. Beugher et al. [2013] describe a system for the analysis of recorded gaze videos by combining trained object recognition, person- and facetracking algorithms. After the desired target is recognized, the fixated object is labeled. Two extensions to the system are suggested to improve the detection performance of faces and bodies: 1.) a human torso detector is trained on images from the VOC2009 dataset, 2.) the gaze cursor is used as a tracker to prevent false detections and to overcome missing detections.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_2

The overview reveals that each solution has its own advantages and limitations. Either it 1.) is tailored to a particular setting (e.g. supermarket); 2.) is not applicable in unrestricted natural scenarios; 3.) needs data pre- or postprocessing; or 4.) only works with additional hardware. Furthermore, time-saving factors (e.g., ease of use, the ability to add AOIs (areas of interest) after the recording is finished, compatibility with statistic software) have to be considered [Evans et al. 2012].

2 Our Contribution

In order to provide a generally applicable and time-saving approach to the analysis process, we developed a modular software called GazeVideoAnalyser that allows for fully- and semiautomatic annotation of gaze videos recorded in unrestricted natural scenarios. The user can select target objects of interest by manually "roping" a rectangular lasso with the mouse around them at particular scene positions in the scene video provided by the mobile eye-tracking system. The selected part is then cut out of the video frame and responding feature vectors are calculated. The software overcomes the limits of existing, largely application specific solutions by combining different object recognition and tracking methods, without the need of any scene or data preparation (e.g., no markers or models are necessary). GazeVideoAnalyser provides a semi-automatic interpolation function, as well as Speeded-Up Robust Features (SURF) [Bay et al, 2006; Evens, 2009] and HSV Color Object Tracking [Smith and Chang, 1996]. Gaze information is used to tune tracking parameters (i.e., to weight fixated areas or to exclude false positives).

3 Discussion

In a preliminary evaluation study we compared the results of the two tracking algorithms against each other and to those of a manual annotation based on scene videos of a typical day-by-day task. Each recorded scene video (MPEG-4) has a resolution of 800 x 600 pixels at 25 fps and a duration of around 1 minute. The scene videos were manually labeled with the ELAN annotation software [ELAN Annotation Tool]. Additionally, all videos were analyzed fully-automatically using the GazeVideoAnalyser on a Quadro Core i7 -4810MQ CPU computer with 2.8 GHz and 8 GByte RAM. The manual annotation of each video took around 20-25 minutes, depending on its length. The automatic annotation lasts around 2-3 minutes (Color Tracker) and 6-7 minutes (SURF) (while feedback results were displayed online), resulting in significant time savings (factor up to 8).

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_2

All in all, the results indicate that the GazeVideoAnalyser provides a reliable automatic video analysis even under challenging recording conditions and can thus significantly speed up the annotation process. By providing no online feedback this factor can even be increased. Furthermore, the tracking algorithms have shown different performance advantages under diverse experimental conditions, making our modular approach with various tracking algorithms a suitable tool for the annotation of videos from natural scenes.

ACKNOWLEDGMENTS: This research/work was supported by the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG).

References

1. Bay H, Tuytelaars T, Shah M (2006). Surf: Speeded up robust features. European Conference on Computer Vision, 1, 404-417.
2. De Beugher S, Youne I, Brône G, Goedemé T (2012). Automatic analysis of eye-tracking data using object detection algorithms. *In Proc. of ACM Conf. on Ubiquitous Computing*, 677-680.
3. Evans C. (2009). Notes on the OpenSURF Library. CSTR-09-001, University of Bristol, January 2009.
4. Evans KM, Jacobs RA, Tarduno JA, Pelz J. (2012). Collecting and Analyzing Eyetracking Data in Outdoor Environments. Journal of Eye Movement Research, 5(2):6, 1-19.
5. Land MF, Mc Leod P. (2000). From eye movements to actions: how batmen hit the ball. *Nature Neuroscience, 3*, 1340-1345.
6. Land MF, Tatler BW (2009). Looking and acting –vision and eye movements in natural behaviour. Oxford: University Press.
7. Paletta L, Santner K, Fritz G, Mayer H, Schrammel J (2013). 3D Attention: Measurement of Visual Saliency Using Eye Tracking Glasses, *Proc. CHI 2013*.
8. Smith JR, Chang, SF (1996). Tools and techniques for colour image retrieval. In I.K.Sethi & R.C. Jain (Eds.), Proceedings SPIE Storage and Retrieval for Still Image and Video Databases IV, 2670, 426–437.
9. ELAN – Annotation Tool: http://www.lat-mpi.eu/tool/elan/.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_2

Semi-automatic annotation of eye-tracking recordings in terms of human torso, face and hands

Stijn De Beugher, Geert Brône, and Toon Goedemé

1 Introduction/Related Work

Research on interactive communication increasingly focuses on the role of eye gaze as an important signal in interaction management, reference and grounding (see [1] for an overview). Interlocutors may use eye gaze as a means to take, hold or give the floor in conversation (turn management), to refer to objects or persons in the conversational space (gaze cueing) or to give and elicit feedback (grounding). The use of unobtrusive eye-tracking technology (like eye-tracking glasses or table-top systems) has proven to be an invaluable resource for obtaining detailed information on the distribution of visual attention of multiple participants simultaneously ([3], [4], [2]). One of the key challenges in the use of mobile eye-tracking technology, however, resides in the processing and annotation of the obtained data stream. An example of an automatic annotation algorithm can be found in our previous work, see [5] for more details. In this paper we further extend our semi-automatic hand annotation algorithm [7], to apply it in the processing of mobile eye-tracking recordings. We integrate the detections of human face, body and hands with gaze data and produce ELAN compatible annotation files.

2 Our Approach

As mentioned above we present the integration of a semi-automatic annotation tool and the gaze data of mobile eye-tracker recordings in order to reduce the manual annotation cost. The annotation classes we tackle are: human torsos, faces and hands. The detection of the human torso and faces is built on our previous work [6]. In [7] we reduced the computational cost of those algorithms by utilizing the Kalman-tracker predictions to reduce the search area of both face and torso. We also implemented an extension to the face detection to left, right and frontal view. The detection of human hands on the other hand is a novel implementation of an accurate

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

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

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

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_1

segmentation combined with advanced tracking mechanisms as well as a validation of human poses. The foundation of our hand detection algorithm is a highly accurate skin segmentation [8] in combination with Kalman-trackers for both hands and arms. Finally we validate hand candidates against a probability map of possible hand/arm positions with respect to the human pose. We embedded those algorithms in a semi-automatic tool, which calculates the confidence of the hand detections. This confidence is a combination of the distance with respect to the detections in the previous frames and the result of the validation against the probability maps. If the confidence drops below a certain threshold, our automatic analysis is halted and the user is asked for manual correction. This threshold is determined empirically using several eye-tracking experiments, but can be adapted by the user making the system more or less strict. After this intervention, our system automatically continues processing the remaining frames. Using such an approach results in a highly accurate system at a minimal cost of manual interventions. For technical details see [7].

As explained above, our semi-automatic system processes each frame captured by a mobile eye-tracker and searches for faces, human torso and hands. Next we map the gaze coordinates of each frame on top of our detections. Our system automatically calculates whether the gaze coordinates overlap with one of our detection classes. The integration of our semi-automatic detections and the gaze data is shown in figure 1. The red dot illustrated the gaze cursor while the rectangles illustrate the detection class that overlaps with the gaze cursor. The green rectangle corresponds to a face detection, while the yellow and blue rectangles corresponds to respectively right and left hand. We applied our algorithm on two different sequences. Each sequence contains images of one person in front of the participant wearing the eye-tracker. The current implementation of our algorithm supports only a single person in the video, however the software is written to support multiple persons in the future. In a next step, our system clusters consecutive frames in which the gaze cursor overlaps with the same detection class. When the length of the cluster is larger than a user defined threshold (standard value of a visual fixation is 150 ms), the cluster is stored as a valid annotation. We assign the class label as annotation value. Finally we export these data to a file that is compatible with annotation tools such as ELAN, making our tool integratable with existing annotations.

3 Results

Although the main contribution of this paper is the integration of semi-automatic hand detection with gaze data, we first present the accuracy results of the hand annotations. We validate our approach on a dataset containing 4000 hand labels and achieve an average accuracy of around 90% at a cost of only 1.7% of manual annotations. The average processing time per frame (1280×720) is around 150ms and includes face, torso and hand detection. Our approach is substantially faster compared to our previous research [6] in which more than 30 seconds were needed to perform the same detections.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_1

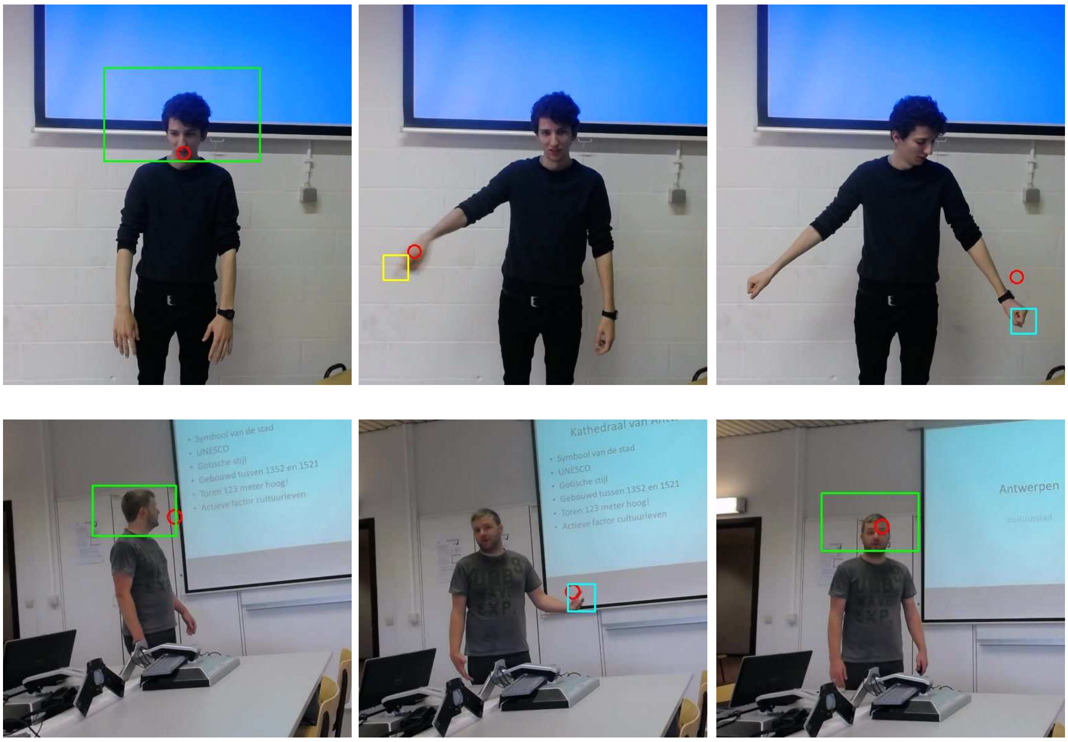


Fig. 1 Results on two sequences of our dataset. Red dot indicates the gaze cursor, coloured rectangles illustrate the detection class that overlaps with the gaze cursor: e.g. face (green), left hand (blue) and right hand (yellow).

Next to the accuracy of the hand detections, we also present the accuracy of our total system: face-, person- and hand detections and the integration of the gaze data. We used our tool to analyze a video sequence of approximately 1m30s and mapped the gaze data on top of the detections. Next we exported those results into an ELAN compatible format. We removed all the labels for validation and asked a participant to manually assign a label to each annotation. To validate our system we applied a statistical analysis, of which the results can be found in table 1. This table reveals the high accuracy of our system making it applicable in real-life situations. A final note should be made to the processing time: the semi-automatic analysis of the video took 5 minutes, while the manual labeling took 20 minutes. The software tool developed in this work will be made publicly available.

Table 1 Statistical analysis of semi-automatic analysis vs manual analysis.

	Percent Agreement	Scott's Pi	Cohen's Kappa	Krippendorff's Alpha
automatic vs manual analysis	95.6%	0.904	0.904	0.904

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_1

4 Conclusion

In this paper we proposed the integration of a novel semi-automatic hand annotation tool [7] and mobile eye-tracking recordings. Our semi automatic annotation tools detects human faces, bodies and hands in images with a high accuracy (around 90%) at the cost of only 1.7% of manual interventions at a speed of 6.5 frames per second. Next we integrate those detections with the gaze data to automatically decide to which part of the body a person is looking at. A comparison of our semi-automatic approach and a fully manual annotation reveals that our system is highly accurate (we scored 0.904 on a Cohen's Kappa test) and that our system is at least 4 times faster than manual analysis.

References

1. Rossano, F (2012). Gaze in Conversation. The Handbook of Conversation Analysis 308-329.
2. Holler, J Kendrick K. H. (2015). Unaddressed participants gaze in multi-person interaction: optimizing recipiency. Frontiers in Psychology 6-98.
3. Jokinen, K (2010). Non-verbal signals for turn-taking & feedback. In Proc. of 7th International Conference on Language Resources & Evaluation.
4. Oertel C, Salvi G (2013). A Gaze-based Method for Relating Group Involvement to Individual Engagement in Multimodal Multiparty Dialogue. Accepted for publication In Proc. of the 15th ACM International Conference on Multimodal Interaction ICMI. Syndney, Australia.
5. De Beugher S, Brône G, Goedemé T (2013). Object recognition and person detection for mobile eye-tracking research. A case study with real-life customer journeys. In Proc. of the First International Workshop on Solutions for Automatic Gaze Data Analysis 2013 (SAGA 2013)
6. De Beugher S, Brône G, Goedemé T (2015). Semi-automatic hand detection - a case study on real life mobile eye-tracker data. In Proc.of VISAPP, pages 121-129.
7. De Beugher S, Brône G, Goedemé T (2015) Fast and accurate semi-automatic hand annotation for human-human interaction analysis. To be published.
8. N. A. Abdul Rahim, C. W. Kit,J. See (2006). RGB-H-CbCr skin colour model for human face detection. In Proc. of M2USIC.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_1

Capturing and Visualizing Eye Movements in 3D Environments

Thies Pfeiffer, Cem Memili, Patrick Renner

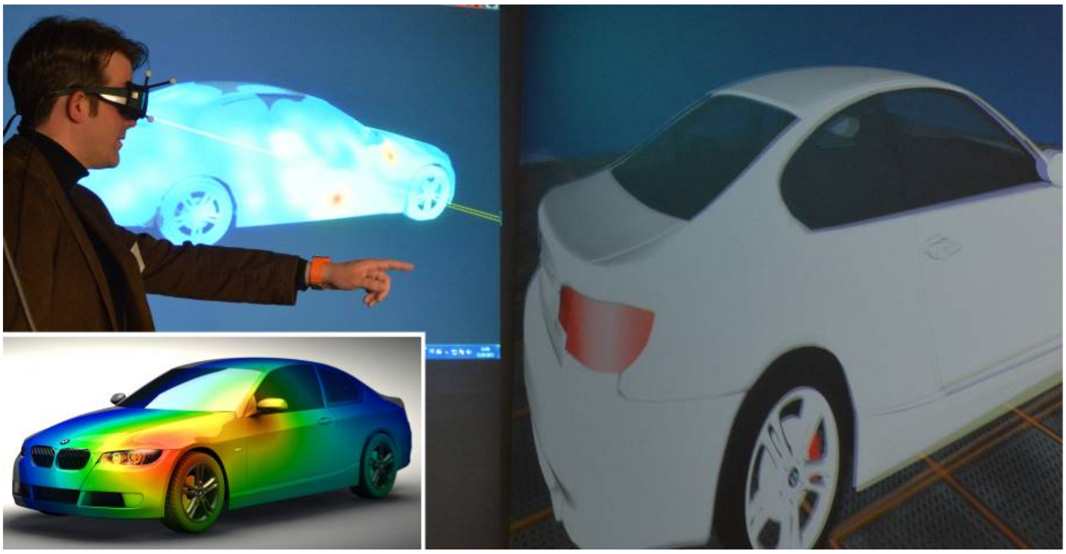


Figure 1: The presented approach is able to map fixations to 3D models of the environment and create 3D heatmaps in real-time. The example shows a 3D gaze recording in an immersive virtual environment (person to the left inspecting 3D object on the right). Real-time heatmap generation is shown in the left background (wrong colors due to photography of a projection screen). Lower left corner shows a rendering of the created 3D heatmaps.

1 Introduction

Visual attention can be a viable source of information to assess human behaviors in many different contexts, from human-computer interaction, over sports or social interactions, to complex working environments, such as to be found in the context of Industry 4.0. In such scenarios in which the user is able to walk around freely, mobile eye-tracking systems are used to record eye movements, which are then mapped onto an ego-perspective video. The analysis of such recordings then requires large efforts for manually annotating the recorded videos on a frame-by-frame basis to label the fixations based on their locations to the target objects present in the video. There are several problems scientists are faced:

- Manual annotations are cumbersome and error prone. The time required for the annotation of the data renders many studies unfeasible.
- The annotated material consists of 2D videos of a 3D environment, which is an abstraction and the results are difficult to integrate. In particular, it is difficult to create appropriate visualizations for the collected data.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_3

2 Our Contribution

First, we present a method to record eye movements in 3D scenarios and annotate fixations with corresponding labels for the objects of interest in real-time [2]. For this purpose, we rely on computer-vision methods for the detection of the camera position and orientation in the world. Based on a coarse 3D model of the environment, representing the 3D areas of interest, fixations are mapped to areas of interest. As a result, we can identify the position of the fixation in terms of local object coordinates for each relevant object of interest.

Second, we present a method for real-time creation and visualization of heatmaps for 3D objects [1]. Based on a live-streaming of the recorded and analyzed eye movements, our solution renders heatmaps on top of the object surfaces. The resulting visualizations are more realistic than standard 2D heatmaps, in that we consider occlusions, depth of focus and dynamic moving objects.

Third, we present a new method which allows us to aggregate fixations on a per object basis, e.g. similar to regions/areas of interest. This allows us to transfer existing methods of analysis to 3D environments.

We present examples from a virtual supermarket, a study on social interactions between two humans, examples from real-time gaze mapping on body parts of a moving humans and from studying 3D prototypes in a virtual reality environment.

3 Discussion

The presented work covers the full workflow from recording to aggregating, analyzing and visualizing data on visual attention over multiple participants. The toolchain is designed in such a way, that the results can be gathered in real-time, during the recording of the study. This not only significantly reduces the time required for the conduction of the studies. It also increases the quality of the recordings, as problems can be detected already during the recording session. The annotation is also objectified and thus random errors by human annotators are eliminated.

ACKNOWLEDGMENTS: This research was supported by the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG).

References

1. Pfeiffer, T., & Memili, C. (2015). GPU-accelerated Attention Map Generation for Dynamic 3D Scenes. In T. Höllerer, V. Interrante, A. Lécuyer, & J. E. S. II (Eds.),Proceedings of the IEEE VR 2015 (pp. 257–258). IEEE.
2. Pfeiffer, T., & Renner, P. (2014). EyeSee3D: a low-cost approach for analysing mobile 3D eye tracking data using augmented reality technology. Proceedings of the Symposium on Eye Tracking Research and Applications, 195–202.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_3

Part 3

Session 02: Human-Computer-Interaction

Online Visual Attention Monitoring for Mobile Assistive Systems

Patrick Renner, Thies Pfeiffer

1 Introduction

Every now and then there are situations in which we are not sure how to proceed and thus are seeking for help. For example, choosing the best product out of dozens of different brands in a supermarket can be difficult, especially when following a specific diet. There are, however, also people who have problems with decision making or sequencing actions in everyday life, e.g. because they suffer from dementia. In such situations, it may be welcomed when there is someone around noticing our problem and offering help. In more private situations, e.g. in the bathroom, help in shape of a human being cannot be expected or even is not welcomed.

Our research focuses on the design of mobile assistive systems which could assist in everyday life activities. Such a system needs to detect situations of helplessness, identify the interaction context, conclude what would be an appropriate assistance, before finally engaging in interaction with the user.

2 Our Contribution

To approach the problem of detecting the cognitive state of helplessness or demand for assistance, our idea is to observe the eye movements of the user. This way, the system is at the same time attentive but quiet in the background until it detects user behaviours which elicit assistance. Only then, the system awakes and enters in direct interaction with the user.

We report about our experiences with such mobile assistance systems which we have gathered in two scenarios: shopping and chess playing. For the shopping situation, the assistance system has been created as a functional prototype in a 3D immersive virtual supermarket scenario (Figure 1). The second system has already been tested in real-world scenarios: Here, computer vision-based detection of the configuration on a chess board provides the basis for context identification.

In the virtual supermarket scenario, we currently investigate which basic features of eye movements could be used to decide in which phase of a decision process the users are currently in [1,2]. In a pilot user study, we tested different ways of presenting augmented information to support the decision process, inter alia an approach in which additional product information is presented on simulated augmented reality glasses. They appear attached to the product currently focused by the

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_6

Figure 1: Prototyping of the mobile assistance system in a 3D immersive virtual supermarket.

user. The augmentation thus adapts in real-time to the current target object via analysis of gaze information. We report about first results of this pilot study.

For the chess scenario, we show how we map gaze positions to locations on the chessboard and present a way to provide feedback on the strength of the fixated chess figure based on sonification. We report on a pilot study in which we tested the user experience of that system.

3 Discussion

Eye movements could provide viable information for mobile assistance systems that automatically adapt to the context of use and the cognitive state of the user. We present work of several projects and report on first pilot studies on user experience.

ACKNOWLEDGMENTS: This work is partly funded by the BMBF in the frame of the project *Adaptive and Mobile Action Assistance in Daily Living Activities* (ADAMAAS) and by the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG).

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), 1–14.
2. Pfeiffer, J., Pfeiffer, T., & Meißner, M. (2015). Towards Attentive In-Store Recommender Systems. In Reshaping Society through Analytics, Collaboration, and Decision Support (S. 161–173). Springer. Fetched from http://link.springer.com/chapter/10.1007/978-3-319-11575-7_11

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0. Individual article archived under DOI: 10.2390/biecoll-saga2015_6

Gaze Tracking for Human Robot Interaction

Oskar Palinko, Francesco Rea, Giulio Sandini, Alessandra Sciutti

Robotics, Brain and Cognitive Sciences Dept., Istituto Italiano di Tecnologia

1 Introduction

Humans use a number of communication cues in their daily interaction with other humans: primarily speech but also gestures, pointing and gaze [1]. The main purpose of gaze is to provide visual information to the subject, but at the same time a person's gaze implicitly provides information to an outside observer about what the subjects are focusing their attention on. There is a number of ways how eye gaze is implicitly used during communication: gaze aversion, mutual gaze, gaze pointing, join attention, etc.

Humans are very good at reading other people's gaze, but robots are less so. This ability would be especially important for humanoid robots to be able to mimic human abilities. However, most human robot interaction experiments today use head pose as a proxy for real eye gaze often times because it's easier to extract than eye gaze [2] [3] [4] [5]. But "head gaze" does not provide all the information that eye gaze does [6], thus enabling robots to perform eye tracking could significantly improve its abilities and also its acceptance by humans. A proof of concept gaze tracker was realized by Matsumoto and Zelinsky [7] and implemented on the HRP2 humanoid [8]. More recently Sciutti et al. [9] implemented a mutual gaze detection system on the iCub which facilitated a teacher/student scenario. Still, so far no extensive use of eye gaze tracking has been done in human-robot interaction.

Figure 1. Left: example performance of our eye tracking system on the iCub robot. Right: human robot interaction example where the subject needs to ask for taking toy building blocks from the robot.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_5

2 Our Contribution

We implemented a monocular feature-based passive gaze tracking algorithm on the iCub platform with the goal of facilitating human robot interaction. The first step in eye tracking is detecting faces and finding face features. For this purpose we used King's implementation [10] of Khazemi and Sullivan's approach for finding features like the corners of the eyes and mouth [11]. We also used Baltrusaitis implementation of the constrained local models approach for tracking head pose [12]. Once these measures were found we proceeded to apply an eye model to the detected center of the pupil similarly as in [13]. The model finally provided the estimate of the gaze angle of the subject, see Figure 1. We then performed a validation experiments in which we found the gaze estimates to be quite acceptable for our setup: the absolute error in the horizontal plane was 5 degrees on average. The accuracy of our system was limited by the cameras used in the iCub setup. We employed PointGrey Dragonfly2 cameras in 1024x768 resolution with fixed-focus 4mm lenses, which produce images of the iris with 20 pixels in diameter when the subject is at 60cm. Knowing that the average diameter of the iris [14] is similar in size to the average eye radius (12mm) [15], then one pixel difference in the middle of the iris corresponds to about 3 degrees difference in gaze. Thus our accuracy is greatly influenced by the hardware used. It is foreseeable that the progressive development of cheaper and small cameras will allow future robotic platforms to have higher resolution sensors, with a consequent improvement of the accuracy of our system. In the meanwhile, the current hardware already enables a gaze estimation from the iCub robot that it can exploit to manage human-robot collaboration tasks.

We also conducted a proof of concept human robot interaction experiment in which subjects were seated opposite of the robot and experimenter, who held toy building blocks in their hands, see Figure 1. The subject's role was to ask for the blocks in specific order, but we did not provide information on how to communicate with the robot. Participants used a combination of speech, pointing and gaze to achieve the task, but the robot really only reacted to gaze. More precisely, the robot handed over pieces of toy building blocks when it detected a succession of mutual gaze and gazing at the requested object. The subjects were not aware of the robot's gaze reading ability, but could still complete the task of building a pillar out of these blocks just by using natural eye behavior (paper submitted to Humanoids 2015). Hence, the robot succeeded in exploiting naturally occurring human gaze behavior to control its helping actions in a collaborative manner.

Future benefits of a built-in gaze tracker in a humanoid robot can be manifold: it could improve turn taking, joint attention and in general the processing of all the communicative gaze cues typical of human interaction. Furthermore, the robot could potentially be used for diagnosing early behavioral problems associated with gaze processing as Autism Spectrum Disorders, by monitoring subjects' gaze in real time.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_5

ACKNOWLEDGMENTS: The research presented here has been supported by the European CODEFROR project (PIRSES-2013-612555).

References

[1] N. George and L. Conty, "Facing the gaze of others.," *Neurophysiol. Clin.*, vol. 38, no. 3, pp. 197–207, Jun. 2008.

[2] M. W. Doniec, G. Sun, and B. Scassellati, "Active Learning of Joint Attention," in *Humanoid Robots, 2006 6th IEEE-RAS International Conference on*, 2006, pp. 34–39.

[3] H. Kim, H. Jasso, G. Deák, and J. Triesch, "A robotic model of the development of gaze following," in *2008 IEEE 7th International Conference on Development and Learning, ICDL*, 2008, pp. 238–243.

[4] S. Ivaldi, S. M. Anzalone, W. Rousseau, O. Sigaud, and M. Chetouani, "Robot initiative in a team learning task increases the rhythm of interaction but not the perceived engagement," *Front. Neurorobot.*, vol. 8, 2014.

[5] S. Sheikhi and J.-M. Odobez, "Combining dynamic head pose–gaze mapping with the robot conversational state for attention recognition in human–robot interactions," *Pattern Recognit. Lett.*

[6] A. Borji, D. Parks, and L. Itti, "Complementary effects of gaze direction and early saliency in guiding fixations during free viewing.," *J. Vis.*, vol. 14, no. 13, p. 3–, Jan. 2014.

[7] Y. Matsumoto and A. Zelinsky, "An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement," in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, 2000, pp. 499–504.

[8] J. Ido, Y. Matsumoto, T. Ogasawara, and R. Nisimura, "Humanoid with interaction ability using vision and speech information," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2006, no. November 2008, pp. 1316–1321.

[9] and G. S. A. Sciutti, L. Schillingmann, O. Palinko, Y. Nagai, "A Gaze-contingent Dictating Robot to Study Turn-taking," in *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction*, 2015.

[10] D. E. King, "Dlib-ml : A Machine Learning Toolkit," *J. Mach. Learn. Res.*, vol. 10, pp. 1755–1758, 2009.

[11] V. Kazemi and J. Sullivan, "One Millisecond Face Alignment with an Ensemble of Regression Trees," in *Computer Vision and Pattern Recognition (CVPR), 2014*, 2014.

[12] T. Baltrusaitis, P. Robinson, and L. P. Morency, "3D Constrained Local Model for rigid and non-rigid facial tracking," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2610–2617.

[13] T. Ishikawa, S. Baker, I. Matthews, and T. Kanade, "Passive Driver Gaze Tracking with Active Appearance Models," in *Proc. World Congress on Intelligent Transportation Systems*, 2004, pp. 1–12.

[14] S. Thainimit, L. A. Alexandre, and V. M. N. de Almeida, "Iris surface deformation and normalization," *Communications and Information Technologies (ISCIT), 2013 13th International Symposium on*. pp. 501–506, 2013.

[15] I. Bekerman, P. Gottlieb, and M. Vaiman, "Variations in Eyeball Diameters of the Healthy Adults," *J. Ophthalmol.*, vol. 2014, 2014.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_5

Part 4
Session 03: Visual Expertise & Motor Performance

Exploring the Relationship Between Motor Imagery, Action Observation, and Action Execution in Motor Skill Learning

A.D'Aquino, C. Frank, K. Essig, and T.Schack[1]

1 Introduction/Related Work

Whether or not the occurrence of eye movements during Motor Imagery (MI) informs and guides the motor system during the mental simulation of an action is still a topic of debate. Rodionov et al. (2004) were the first who examined eye gaze in movement imagery. Their data suggested that nystagmic activity in the horizontal plane could be elicited during movement imagery providing evidence that eye movements could be used as an objective measure of online cognitive processes. Using a cyclical aiming task, Heremans et al. (2008) reported that the number and amplitude of the eye movements during imagery resembled closely those of eye movements made during the physical execution of the task. The findings contrast, in part, those of McCormick et al. (2013) who reported that additional fixations were made during physical execution. The differences may be explained through the demands of the actions performed. Heremans et al. (2008) used a relatively low demand cyclic wrist extension action that was cued externally whereas McCormick et al. (2013) employed a task that involved the optimal movement of a stylus to a target in the sagittal plane. These data suggest that the neural coupling that exists between eye and hand movements during physically executed movements remains partially, but not totally, intact in imagery (i.e., fixation location is preserved) for relatively simple tasks that involve movement towards a motionless target. Therefore, it still remains unexplored whether for more complex tasks, such as the ones that require acting upon a moving target, the neural coupling between eye and hand-movement is maintained or further disrupted during the physical or mental simulation action.

[1]Alessio D'Aquino, mail: alessio.daquino@uni-bielefeld.de
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

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_7

2 Our Contribution

To address this issue, a first study will be aimed at exploring the congruency of gaze metrics between motor imagery and action execution for a complex aiming task. 20 university students will be asked to intercept or to imagine Intercepting a moving target displayed on a computer screen. Total fixation duration, number of fixations on the target and for the whole scene, and saccade amplitude will be then compared between performance conditions in which the presence /absence of the visual stimulus has been manipulated (unguided motor imagery, guided motor imagery target visible, guided motor imagery target disappearing, physical execution, and a control condition) and between different target speeds (fast vs slow) in a within-subject design. It should be noted that, by adding multiple performance conditions, we wanted to provide more detailed evidence on how the presence/absence of visual guidance affects participants' gaze behavior during the mental simulation of the task. Hence, 5 (performance conditions) X 2 (target speed) RM ANOVAs will be used to compare the aforementioned gaze metrics. We expect gaze metrics to be similarly affected by task complexity (target speed) and eye movement variables in the more overt imagery condition (guided motor imagery target visible) to be more congruent to the physical execution of the interceptive task rather than the more covert conditions. The data from this experiment will be then compared to previous findings and will serve the purpose of testing whether specific gaze metrics, such as number of fixations, are congruent or differ for a more dynamic task, such as those ones that require a target to be intercepted.

3 Discussion

In this talk we will present preliminary results of our computer study. Furthermore, we will provide an outline of our future research, where our aim will be to compare the findings of the laboratory study to a more naturalistic task with a mobile target (e.g. dart throwing, pistol shooting). Subsequently, we will compare eye movements between action execution and motor imagery and explore whether adaptations in gaze behavior are maintained in simulation conditions after observational learning, i.e., whether such adaptation will be transferred to the mental simulation of movement. These insights will not only provide additional evidence on how eye movements inform and guide the motor system during mental simulation of an action, but also investigate these processes in more natural and complex situations. In the applied settings, the results from these three studies could inform practitioners in different fields. In cognitive psychology the data could provide an insight about differences in information processing during motor imagery when compared to action execution. In performance psychology the results from these

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_7

study could provide precious evidence regarding how to optimizing motor imagery training (e.g. visual-cues vs no-visual cues). In more technical areas the capability of generating efficient and embodied mental simulations is still lacking. Therefore, studying the visual-perceptual differences between simulation conditions could provide innovative evidence which can then be implemented in humanoid robots to construct more realistic systems.

ACKNOWLEDGMENTS: This research is supported by the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG).

References

1. Heremans E, Helsen WF, Feys P (2008).The eyes as a mirror of our thoughts: Quantification of motor imagery of goal-directed movements through eye movement registration. Behav Brain Res 187:351-60.
2. McCormick SA, Causer J, Holmes PS (2013). Active vision during action execution, observation and imagery: Evidence for Shared Motor Representation. PLoS ONE 8(6): e67761. doi:10.1371/journal.pone.0067761.
3. Rodionov V, Zislin J, Elidan, J (2004). Imagination of body rotation can induce eye movements. Acta Otolaryngol 124: 684–689.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_7

Into the wild – Musical communication in ensemble playing. Discerning mutual and solitary gaze events in musical duos using mobile eye-tracking

Sarah Vandemoortele, Stijn De Beugher, Geert Brône, Kurt Feyaerts, Toon Goedemé, Thomas De Baets, and Stijn Vervliet

1 Introduction/Related Work

The communicative behaviour of musicians is multimodal: it involves gazing behaviour, gesture and musical sound. The current study focuses specifically on the function and timing of interactive gazing behaviour. Thus far, only little systematic research has been dedicated to this specific aspect of musical interaction, with the exception of Kawase (2014a, 2014b). Using external cameras, Kawase investigated two communicative aspects: the impact of leader- follower roles on gazing behaviour (Kawase 2014a) and the influence of gaze events on coordination (Kawase 2014b). Eye-tracking technology has found its way into the musical research area in studies on music reading (Drai- Zerbib et al. 2012, Penttinen & Huovinen 2011, Wurtz et al. 2009). By implementing mobile eye trackers into the study of musical ensemble interaction this study ventures into a relatively new and unknown territory.

2 Our Contribution

The present study focuses on solitary and mutual gaze events in musical duos. Solitary events are defined in this study as the occasions where one musician looks at the other. Mutual events refer to those occasions where both musicians look at each other. The research questions are:

(i) Do solitary and mutual gaze events tend to reoccur at the same places in the musical piece?
(ii) Do these events correlate with specific musical characteristics?
(iii) Do they correlate with specific problems in the rehearsal process?

Five duos (bachelor and master level students of the Lemmens Institute, Leuven) were recorded while playing and working on a piece of their choice. The members of each duo had previously played together and had already performed or were planning to perform the piece of their choice. The instrumentation of each duo was unique, ranging from relatively unchallenging (two flutes, two guitars) over moderately challenging (harp-violin, clarinet-piano) to challenging (two percussion-

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_8

ists) as regards the implementation of mobile eye-tracking. Each duo was recorded during two or three rehearsal sessions of around 30 minutes and was asked to work for each session on the same piece. At the beginning and end of each rehearsal the participants were asked to play through the entire piece, which resulted in a total amount of 26 run-throughs.

Two ‘Pupil Pro’ eye trackers recorded the eye movements of both players during the entire rehearsal session. The rehearsal was interrupted twice for the purpose of checking and adjusting the calibration: after the first and before the second run-through. External cameras captured the full body of each player and an audio recorder guaranteed a reasonable sound quality. After each session the musicians completed a questionnaire asking them to specify difficulties in the musical score. The synchronized data (from the eye trackers, external cameras and audio recorder) are currently being annotated for gaze distribution using the annotation software ELAN.

3 Discussion

We aim to present preliminary results drawn from segments of about 1 minute out of each run-through, during which the same musical fragment is heard. The annotated solitary and mutual gaze events will be compared across several run-throughs within the same duo as well as across different duos. One general finding so far has been the fact that *mutual* gaze events tend to be reserved for the beginnings and endings of pieces. The causes for solitary gaze events seem to be rather diverse and related to unexpected events in the performance (e.g. a mistake by the partner), important structural moments in the music, the reversal of leader-follower roles, etc. However, analysis of a more substantial amount of data is needed in order to confirm and further specify these results.

References

1. Drai-Zerbib V, Baccino T, Bigand E (2012). Sight-reading expertise: Cross-modality integration investigated using eye tracking. Psychology of Music 40:216-235.
2. Kawase S (2014a). Assignment of Leadership Role Changes Performers' Gaze During Piano Duo Performances. Ecological Psychology 26:198-215.
3. Kawase S (2014b). Gazing behavior and coordination during piano duo performance. Attention, Perception, & Psychophysics 76:527- 540.
4. Keller PE (2008). Joint Action in Music Performance. In: Morganti F, Carassa A, Riva G. (ed), Enacting intersubjectivity: A cognitive and social perspective to the study of interactions (pp. 205-221), Amsterdam: IOS Press.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_8

5. Penttinen M, Huovinen E (2011). The Early Development of Sight-Reading Skills in Adulthood: A Study of Eye Movements. Journal of Research in Music Education 59:196-220.
6. Williamon A, Davidson J (2002). Exploring co-performer communication. Musicae Scientiae 6:53-72.
7. Wurtz P, Mueri RM, Wiesendanger M (2009). Sight-reading of violinists: eye movements anticipate the musical flow. Experimental Brain Research 194:445-450.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_8

Part 5

Session 04: Methods for Measuring & Quantification

Analyzing Patterns of Eye Movements in Social Interactions

Nadine Pfeiffer-Leßmann, Patrick Renner, Thies Pfeiffer

1 Introduction

Eye gaze plays an important role in human communication. One foundational skill in human social interaction is joint attention which is receiving increased interest in particular in the area of human-agent or human-robot interaction. We are focusing here on patterns of gaze interaction that emerge in the process of establishing joint attention. The approach, however, should be applicable to many other aspects of social communication in which eye gaze plays an important role.
Attention has been characterized as an increased awareness [1] and intentionally directed perception [2] and is judged to be crucial for goal-directed behavior. Joint attention can be defined as simultaneously allocating attention to a target as a consequence of attending to each other's attentional states [3]. In other words: Interlocutors have to deliberatively focus on the same target while being mutually aware of sharing their focus of attention [2] [4].

2 Our Contribution

2.1. Study
In a study with 18 pairs of participants (i.e. 36 different participants in total), we recorded eye movements while participants were solving a visual search task based on a verbal description of the experimenter and a communication task where they both had to cooperate on agreeing on a specific target stimulus.

2.2. Automatic labeling of fixation targets
Each pair of participants had to solve 32 tasks, which took about 30 minutes, resulting in 18 hours of mobile eye movement videos. Annotation time was reduced to zero by mapping from fixations to semantic regions automatically based on the EyeSee3D approach [5].

2.3. Analysis of interaction patterns
Our approach for the analysis is to define several indices to describe basic gazing behaviors, such as

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_10

Figure 1: Interaction scenario of our study: Two participants, both eye-tracked, have to solve cooperative tasks. Eye movements were recorded to analyze patterns of eye gaze interaction.

- Searching $\frac{|fixations(other\ stimuli)|}{|fixations(all)|} \rightarrow 1$
- Identifying $\frac{|fixations(target\ stimulus)|}{|fixations(all)|} \rightarrow 1$
- Looking at partner $\frac{|fixations(partner)|}{|fixations(all)|} \rightarrow 1$
- Looking for help $\frac{|fixations(other\ stimuli)|+|fixations(partner)|}{|fixations(all)|} \rightarrow 1$
- Helping $Identifying \gg Looking\ at\ partner$
- Following $Looking\ for\ help \gg Looking\ at\ partner$

This provides an abstraction of the concrete gazing behavior, which will then be used for the analysis of dyadic patterns. We will present the results of this analysis and discuss our approach with existing alternatives.

3 Discussion

We present an approach to characterize and analyze eye movement data in a way that is compatible with an online analysis. The EyeSee3D approach [5] can be applied to studies in the domain of social interactions, to significantly reduce the time required for the analysis of the eye movement data. In combination, the index-based classification and EyeSee3D will allow us to interpret gaze patterns of interacting persons in real-time, which is highly relevant for realizing human-agent interactions.

ACKNOWLEDGMENTS: This research was partly supported by the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) and the SFB 673 Alignment in Communication; both are at Bielefeld University and funded by the German Research Foundation (DFG).

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0. Individual article archived under DOI: 10.2390/biecoll-saga2015_10

References

1. Brinck, I. (2003): The objects of attention. In: Proc. of ESPP2003, Torino.
2. Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005): Understanding and sharing intentions: The origins of cultural cognition. Behavioral and Brain Sciences 28
3. Deak, G.O., Fasel, I., and Movellan, J. (2001): The emergence of shared attention: Using robots to test developmental theories. In: Proc. of the First Intl. Workshop on Epigenetic Robotics, Lund University, Cognitive Studies, 85.
4. Hobson, R.P. (2005): What puts the jointness into joint attention? In Eilan, N., Hoerl, C., McCormack, T., Roessler, J., eds.: Joint attention: communication and other minds. Oxford University Press.
5. Pfeiffer, T., & Renner, P. (2014). EyeSee3D: a low-cost approach for analysing mobile 3D eye tracking data using augmented reality technology. Proceedings of the Symposium on Eye Tracking Research and Applications, 195–202.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_10

A Metric to Quantify Shared Visual Attention in Two-Person Teams

Patrick Gontar, Jeffrey B. Mulligan

1 Introduction

Critical tasks in high-risk environments are often performed by teams, the members of which must work together efficiently. In some situations, the team members may have to work together to solve a particular problem, while in others it may be better for them to divide the work into separate tasks that can be completed in parallel. We hypothesize that these two team strategies can be differentiated on the basis of shared visual attention, measured by gaze tracking.

2 Method

Gaze recordings were obtained for two-person flight crews flying a high-fidelity simulator (Gontar & Hoermann, 2014). Inherent for those teams is that they normally do not face each other while working in time-critical situations (in contrast to infant-parents' shared attention research, such as Redcay, Kleiner, and Saxe, 2012). Gaze was categorized with respect to 12 areas of interest (AOIs). We used these data to construct time series of 12 dimensional vectors, with each vector component representing one of the AOIs. At each time step, each vector component was set to 0, except for the one corresponding to the currently fixated AOI, which was set to 1. This time series could then be averaged in time, with the averaging window time (Δt) as a variable parameter. For example, when we average with a Δt of one minute, each vector component represents the proportion of time that the corresponding AOI was fixated within the corresponding one minute interval. We then computed the Pearson product-moment correlation coefficient between the gaze proportion vectors for each of the two crew members, at each point in time, resulting in a signal representing the time-varying correlation between gaze behaviors. We determined criteria for concluding correlated gaze behavior using two methods: first, a permutation test was applied to the subjects' data. When one crew member's gaze proportion vector is correlated with a random time sample from the other crew member's data, a distribution of correlation values is obtained that differs markedly from the distribution obtained from temporally aligned samples. In addition to validating that the gaze tracker was functioning reasonably well, this also allows us to compute probabilities of coordinated behavior for each value of the correlation. As an alternative, we also tabulated distribu-

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_20

tions of correlation coefficients for synthetic data sets, in which the behavior was modeled as a first-order Markov process, and compared correlation distributions for identical processes with those for disparate processes, allowing us to choose criteria and estimate error rates.

3 Discussion

Our method of gaze correlation is able to measure shared visual attention, and can distinguish between activities involving different instruments. We plan to analyze whether pilots' strategies of sharing visual attention can predict performance. Possible measurements of performance include expert ratings from instructors, fuel consumption, total task time, and failure rate. While developed for two-person crews, our approach can be applied to larger groups, using intra-class correlation coefficients instead of the Pearson product-moment correlation.

ACKNOWLEDGMENTS: This work was supported by the TUM Graduate School and a LuFo IV-2 grant (PG), and NASA's Airspace Operations and Safety Program (JBM). The authors are grateful to Immanuel Barshi for his help and support.

References

1. Gontar, P., & Hoermann, H.-J. (2014). Flight Crew Performance and CRM Ratings Based on Three Different Perceptions. In A. Droog (Ed.), *Aviation Psychology: facilitating change(s): Proceedings of the 31st EAAP Conference* (pp. 310–316). Malta.
2. Redcay, E., Kleiner, M., & Saxe, R. (2012). Look at this: the neural correlates of initiating and responding to bids for joint attention. *Frontiers in human neuroscience*, 6, 169.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_20

The impact of image size on eye movement parameters

Ricardo Ramos Gameiro, Kai Kaspar, Sontje Nordholt, Peter König

1 Introduction/Related Work

Eye tracking studies have intensely addressed what attracts attention while viewing natural images. Hereby, it was found that image features (e.g. luminance, contrast) [1] and also semantic content of the scenery [2] play a high role to guide visual exploration. However, image size as causal factor has been rarely considered. Indeed, in typical laboratory setups visual stimuli lay in a small central part of the visual field only. Furthermore, image and screen sizes differ across laboratories. Thus, it is unclear in how far observed eye movement parameters indicate properties of the human visual/occulomotor systems in the real world or of typical laboratory setups.

2 Our Contribution

Here, we investigated the effect of image sizes on eye movement parameters. Specifically we test the hypotheses that visual exploration scales with images size or, alternatively, remains constant in absolute measures.

Participants were exploring images of different categories (web pages, urban scenes, nature scenes) varying in image size (7, 10, 15, 21, 30") in a free viewing task, while their eye movements were recorded. The distance to the screen was set to 80cm. In the 30" full screen condition, this provided an almost full coverage of the image on the visual field to imitate scenarios for watching urban places and landscapes in the real world.

Results revealed a central spatial bias of fixations in urban and nature scenes, whereas web pages showed an upper-left bias. These bias effects increased linearly with stimulus size, as indicated by measures of fixation eccentricities in horizontal and vertical direction as well as entropy. Additionally, the mean saccadic amplitudes also increased linearly with size. To test whether this effect arose of a general shift of amplitudes towards higher values or of a few single very large saccades in large images, we looked at the histogram of amplitudes. We saw that amplitudes were log-normally distributed within each stimulus size. However, with increasing size the distributions became more broadly shifting towards higher amplitudes. The maximum of the log-normal curves decreased for larger images but remained at the same location over all sizes. Thus, we did not find a sharp general shift of saccadic amplitudes but rather more spread distribution. According to single fixation analyses we showed that larger images led to a higher amount of fixations combined with a decrease of the mean duration of single fixations. Hereby, both parameters followed a logarithmic rather than linear trend (positive and negative, respectively).

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_9

3 Discussion

Consequently, the size in which visual stimuli are presented significantly affects those parameters that are commonly investigated in many studies on overt attention. Still, we cannot clearly claim either if larger images are treated as such leading to a visual exploration in different absolute measures or if exploration is up scaled.

References

[1] 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, pp. 1254–1259.

[2] Einhäuser W, Spain M, Perona P (2008). Objects predict fixations better than early salience. Journal of Vision November 2008 Vol.8, 18.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_9

Part 6

Session 05: Language & Memory

Language and gaze cues: findings from the real world and the lab

Ross Macdonald, Maria Staudte, and Benjamin Tatler

1 Introduction

People follow the gaze cues of others. Lab-based studies have shown that we reflexively follow gaze cue stimuli [1, 2], however it is unclear whether this occurs in the real world, as lab-based paradigms often lack key features of natural gaze cueing, such as the social and communicative context. We explored how social context and language affect gaze cues using real world tasks and complemented these tasks with more controlled lab-based studies.

2 Our Contribution

2.1 Exploratory real world tasks. In Study 1 (Figure 1a) two participants worked together to make a cake, while wearing portable eye-trackers [3]. Half of the pairs were given roles ("Chef" or "Gatherer") and the other half were not. We found that our social manipulations within and between these pairs affected the way participants aligned their gaze with each other and how they sought out the gaze cues of their partners. Study 2 (Figure 1b) used a real world search-task to explore how gaze cue utilisation varied depending on the form of concurrent language used [4]. An instructor varied his referring expressions (featural or spatial) and the presence of gaze cues (absent, congruent, or incongruent). Participants used the inherently spatial gaze cues strategically; cues were sought out and followed more when they were more informative relative to the accompanying verbal referring expressions. Combined, the results of these studies show that far from reflexively following gaze cues, people strategically used gaze cues in real world interactions depending on their beliefs about the gazer and the language accompanying the gaze cues. Moving away from traditional lab-based gaze cueing paradigms and using real-world eye-tracking methods was crucial to identifying these effects.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_12

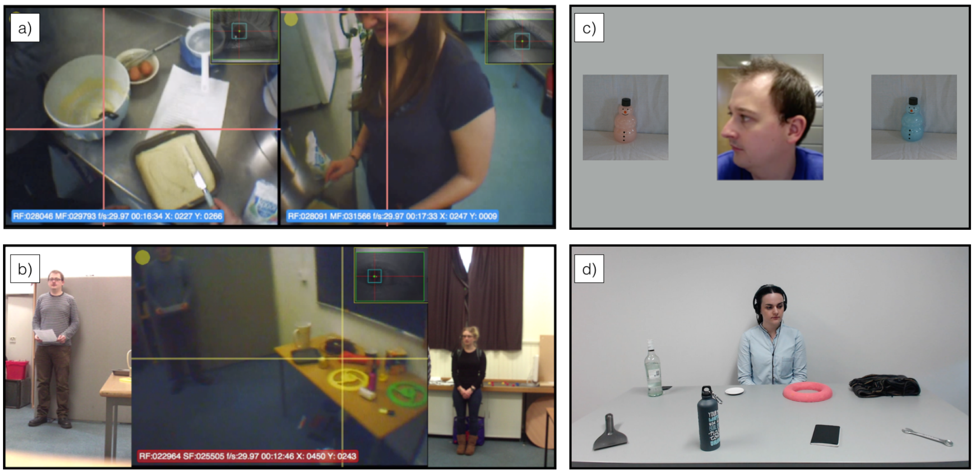


Figure 1. The left panels show stills from real world eye-tracking studies: a) a natural collaboration and b) a search task. The right panels show monitor-stills from lab-based eye-tracking studies, investigating c) the effects of language and gaze cue reliability and d) gazer intention.

2.2 Controlled lab-based tasks. Due to the complexity of real world tasks, many of our manipulations were necessarily course-grained. We have therefore used lab-based tasks to more thoroughly explore factors we found to be relevant in the real world. The effect of the informativeness of gaze found in Study 2 was investigated in Study 3 (Figure 1c). Gaze and language cues were reduced to equivalent artificial stimuli and the reliability of each was manipulated in a fine-grained manner [5]. Language was preferred when cues were equally informative. Reflexive gaze cueing effects were found, however these effects were modulated by gaze cue reliability. Study 4 (Figure 1d) further explored the effect of the beliefs about a gazer (Study 1) by manipulating the intention of the gazer in a controlled setting. Preliminary results show no evidence of perceived intention affecting gaze following, but some evidence that participants looked more at a gazer's face when the gazer and participant's intentions were aligned compared to when they were distinct.

3 Discussion

Our real world studies have provided insights into gaze following that could not be achieved with typical gaze cueing paradigms and these insights were used to inform the design of more controlled lab-based studies. Combining real world and lab-based paradigms is essential to fully understand the use of gaze cues in natural interactions.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_12

References

1. Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin and Review, 5,* 490–495.
2. Ricciardelli, P., Bricolo, E., Aglioti, S. M., & Chelazzi, L. (2002). My eyes want to look where your eyes are looking: Exploring the tendency to imitate another individual's gaze. *Neuroreport: For Rapid Communication of Neuroscience Research, 13,* 2259–2264.
3. Macdonald, R. G., & Tatler, B. W. (2013). *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 942-947). Austin, TX: Cognitive Science Society.
4. Macdonald, R. G., & Tatler, B. W. (2015). *Journal of Experimental Psychology: Human Perception and Performance*, 41 (2), 565-575.
5. Macdonald, R. G., & Tatler, B. W. (2014). *Proceedings of the 36th Annual Conference of the Cognitive Science* (pp. 910-915). Austin, TX: Cognitive Science Society.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_12

Collecting memories: the impact of active object handling on recall and search times

Dejan Draschkow[1] & Melissa L.-H. Võ[1]

Johann Wolfgang Goethe-Universität, Frankfurt, Germany [1]

1 Introduction

In natural behavior we not only visually inspect our environment, but often actively interact with objects which are part of it. Our perception and cognition is modulated by actual as well as planned actions towards objects (for a review see Witt & Riley, 2014). Priorities for selection and representations of a purely visual task do not reflect those present in a natural task in which objects are actively manipulated (Tatler et al., 2013). There is a strong dependence of memory representations on behavioral goals (Droll, Hayhoe, Triesch, & Sullivan, 2005; Droll & Hayhoe, 2007; Triesch, Ballard, Hayhoe, & Sullivan, 2003), which becomes particularly evident in natural behavior (Tatler & Land, 2011).

2 Our Contribution

To investigate how physically engaging with objects influences both identity memory (free recall) and location memory (subsequent search for these objects), we constructed a real-world paradigm in which participants equipped with a mobile eye tracker (SMI Eye Tracking Glasses) first either searched for cued objects via verbal response without object interaction (Passive condition) or actively collected the objects they found (Active condition) (Figure 1). Participants conducted this task within an actual four-room apartment (kitchen, bedroom, living room, and study); with each room containing 10 designated Active, Passive, and Distractor objects. Additionally, a unique category of objects was designated as relevant for each room (e.g. “objects needed for making a sandwich” in the kitchen) and participants were instructed to decide if an object was relevant upon finding it. After the 80 trials were completed, a surprise recall test was conducted in which participants were asked to list every object they remembered from the rooms in order to access identity memory performance. After completion, participants had to search trial-by-trial for all previously searched (Passive), searched + collected (Active), never searched but present (Distractors), and not included (Absent) objects.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_11

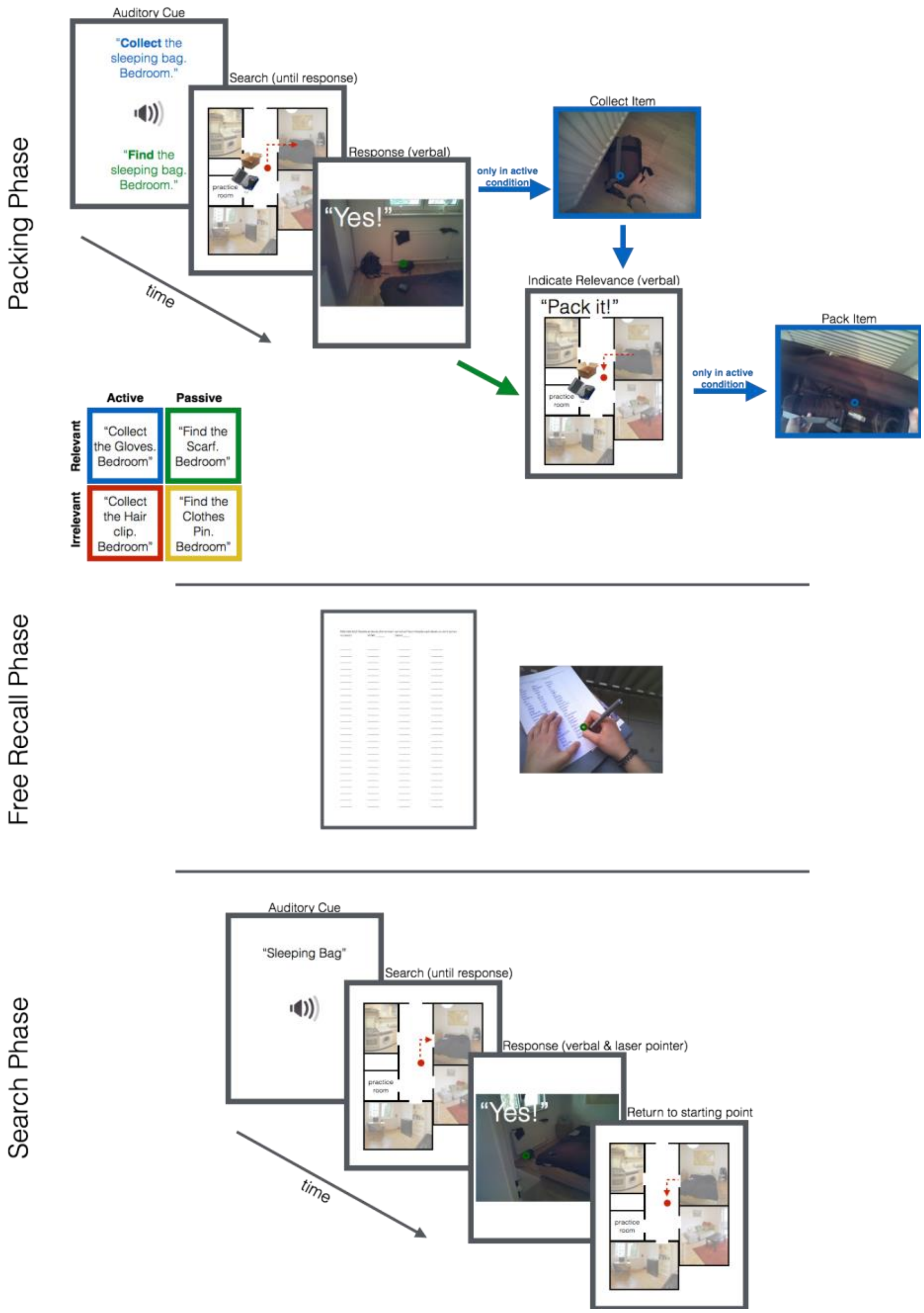


Figure 1: A depiction of the experimental procedure. The top part of the figure illustrates an example trial in the Packing Phase (80 trials): in the Passive as well as Active condition participants responded as soon as they found the target object, but participants were instructed to collect the object only in the Active condition. In both condition the participants had to indicate if the object was relevant. Experimental conditions and object searches were randomized in a trial-by-trial fashion. In the subsequent Free Recall Phase participants wrote down all objects they remembered. In the final Search Phase a randomized search of all the objects from the Packing Phase followed, supplemented with Distractor and Absent trials.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_11

3 Discussion

Identity memory was substantially modulated by task relevance, with a higher memory performance for relevant objects in both the Active and the Passive condition. Distractors were recalled less than targets, but were modulated by relevance as well, i.e. there was a higher recall performance for relevant distractors, even though they were never the target of a search.

Time to first fixation, as an indicator of location memory, was shorter for task-relevant objects following physical interaction (Active condition), but did not differ between relevant and irrelevant objects in the Passive condition. Time to first fixation for Distractors was slower than for targets. Task-relevant Distractors were fixated faster than irrelevant ones.

In the current study we demonstrate that active object manipulation interacts with task-relevance. Time to first fixation on relevant objects was faster compared to irrelevant ones, but only if these objects were previously manipulated.

ACKNOWLEDGMENTS: This work was supported by Deutsche Forschungsgemeinschaft (DFG) Grant VO 1683/2-1 to MLV. We wish to thank Sage Boettcher, Daniela Gresch, Maximilan Scheuplein and Leonie Polzer for valuable help with data collection and analysis.

References

Droll, J. A., & Hayhoe, M. M. (2007). Trade-offs between gaze and working memory use. *Journal of Experimental Psychology. Human Perception and Performance*, *33*(6), 1352–65. doi:10.1037/0096-1523.33.6.1352

Droll, J. A., Hayhoe, M. M., Triesch, J., & Sullivan, B. T. (2005). Task demands control acquisition and storage of visual information. *Journal of Experimental Psychology. Human Perception and Performance*, *31*(6), 1416–38. doi:10.1037/0096-1523.31.6.1416

Tatler, B. W., Hirose, Y., Finnegan, S. K., Pievilainen, R., Kirtley, C., & Kennedy, A. (2013). Priorities for selection and representation in natural tasks. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *368*(1628), 20130066. doi:10.1098/rstb.2013.0066

Tatler, B. W., & Land, M. F. (2011). Vision and the representation of the surroundings in spatial memory. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *366*(1564), 596–610. doi:10.1098/rstb.2010.0188

Triesch, J., Ballard, D. H., Hayhoe, M. M., & Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision*, *3*(1), 86–94. doi:10:1167/3.1.9

Witt, J. K., & Riley, M. A. (2014). Discovering your inner Gibson: Reconciling action-specific and ecological approaches to perception–action. *Psychonomic Bulletin & Review*, *21*(6), 1353–1370. doi:10.3758/s13423-014-0623-4

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_11

Part 7

Session 06: Real World Studies & Applications

Gaze Analysis in Mobile Pedestrian Navigation: Socio-Cultural Aspects and Wayfinding

Lucas Paletta[1], Stephanie Schwarz[2], Jan Bobeth[2], Michael Schwarz[1], and Manfred Tscheligi[2]

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

[2]AIT Austrian Institute of Technology GmbH, Vienna, Austria

1 Introduction

Using augmented reality is recognized as a suitable alternative to map-based interfaces for mobile pedestrian navigation, as it provides route instructions directly into the real visual context of the user [1].

The analysis of the concrete orientation behavior of using the phone, of paying attention to the immediate environment and the navigation landmarks has been of minor interest so far. However, researching the impact of socio-cultural aspects on wayfinding styles requires detailed information about the user's focus of attention, its preference for information types associated with particular affordances and environments, and continuous measurements about the user interaction, across extended periods of time.

An exploratory analysis of gaze behavior was conducted to identify point-of-regards (POR) on predefined areas of interest (AOI) within the smartphone display and towards the environment. To acquire video and eye movement data, SMI Eye Tracking Glasses were used, with 30 Hz sampling rate of gaze information and 1280 x 960 pixels scene camera (Figure 1). A screencast video of the navigation app was recorded and synchronized with the eye tracking data. For the post-processing of the gaze data, the smartphone eye tracking (SMET) system [2] was applied, which was demonstrated to be feasible for large scale studies [4].

2 Mobile eye tracking study

Navigation modes. In an outdoor study, gaze movements were recorded to investigate participants´ gaze during a wayfinding task on a predefined route on the campus at University Hospital Graz, Austria (Figure 1). A mobile navigation tool was used providing two alternative presentation modes on the smartphone (Figure

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_14

2a) to indicate the recommended route: (1) a two-dimensional map based view (MAP) and (2) an augmented reality-supported view (AR).

Data capture. The SMET system [2] enables fully automated analysis of attention in user studies and showed highly accurate POR mappings on smartphone displays. Figure 2b shows the automated smartphone localization on a sample video frame from the eye tracking scene camera. Synchronization and image analysis provide a correlated data stream of smartphone events, geometric transformations and heat-mapping for further attention analysis. For the analysis of the gaze patterns we investigated content-dependent AOI assignments (Figure 2**Fehler! Verweisquelle konnte nicht gefunden werden.**b) including the two interface elements on the smartphone (MAP, AR) as well as gaze into the physical surrounding (SUR).

Study. In total 20 women participated in the study, including 10 immigrants from Turkey and 10 local citizens from Graz, Austria. Among the Turkish participants the average duration of stay in Graz was 8.55 ±5.43[1] years. The age of the participants was 28.3 ±6.5 (Turkish) and 28.3 ±6.5 years (Austrian). All participants had experience with smartphones and were familiar with the overall area; they all had previously visited the hospital. Turkish participants reported to be less experienced with maps.

The objective of the study was to investigate whether there are principal differences between the viewing behaviors of the two participant groups, in the context of using a mobile interface for navigation, in particular, in the usage of map or augmented reality based services for wayfinding.

3 Results

Figure 3 shows the average count of PORs with respect to the selected AOIs, i.e., AOIs of the augmented reality mode (AR), the map mode (MAP) and the spatial surrounding (SUR).

Analysis revealed that the Austrian participants preferred to look on the MAP than on the AR display region whereas Turkish users looked on average more often on the AR view than on the MAP. Results indicate a tendency in the context of the users’ socio-cultural background on the amount of PORs regarding the functional use of the MAP and SUR views, respectively.

Consequently, the findings drawn from eye tracking data as well as from qualitative feedback were capable to reveal that there exist relevant differences in the interface preferences of the Turkish users in the study in contrast to behaviors of the native Austrian participants. Future work will investigate the attention processes in more depth and in the context of the socio-cultural variables, for example, whether the social (immigrants vs. permanent citizen) or the cultural back-

[1] *Note: M ±S with M (=mean) and S (=standard deviation)*

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_14

ground (Turkish vs. Austrian) of the users definitely impacts the wayfinding and the tool use in mobile navigation.

References

1. Walther-Franks, B (2007) Augmented reality on handhelds for pedestrian navigation. Master thesis, University of Bremen, Germany.
2. Paletta, L., Neuschmied, H., Schwarz, M., Lodron, H., Pszeida, M., Ladstätter, S., Luley, P. (2014) (2014), Smartphone Eye Tracking Toolbox: Accurate Gaze Recovery on Mobile Displays, *Proc. Symposium on Eye Tracking Research and Applications*, ETRA 2014, Safety Harbor, FA, pp. 367-68.
3. Goldberg, J. and Kotval, X. (1999) Computer interface evaluation using eye movements: methods and constructs, *International Journal of Industrial Ergonomics*, 24, 631-645.
4. Paletta, L., Neuschmied, H., Schwarz, M., Lodron, H., Pszeida, M., Luley, P., Ladstätter, S., Deutsch, S., Bobeth, J., and Tscheligi, M. (2014). Attention in Mobile Interactions: Gaze Recovery for Large Scale Studies, *Proc. Conference on Human Factors in Computing Systems*, CHI'14 extended abstracts, pp. 1717-1722.

(a) (b)

Figure 1. Eye tracking glasses used by (a) local and (b) immigrant users.

(a) (b)

Figure 2. Navigation app: (a) AR (left) and MAP (right) presentation modes, (b) view from eye tracking scene camera with automated gaze recovery (green dot) [2].

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_14

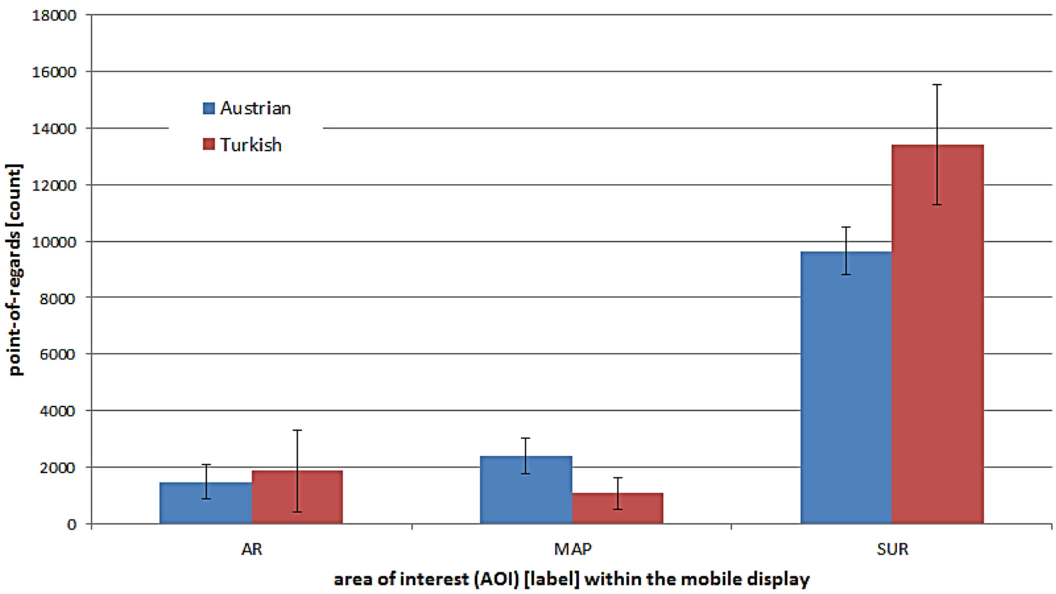


Figure 3. Average gaze count on areas-of-interest (AR-augmented reality, MAP-map, SUR-surrounding) with respect to the mobile display by Austrian (blue) and Turkish immigrant (red) participants. The error bars indicate the 95% confidence intervals. Austrian participants clearly preferred to look at the map based information than on the augmented reality interface component. Turkish participants focused more on the surrounding than Austrian participants and looked on average more on intuitive navigation information (AR) than on the map based information display.

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 Research Promotion Agency (FFG) under contract No. 832045 (FACTS).

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_14

Effect of Familiarity on Visual Attention and Choice at the Point of Purchase

Kerstin Gidlöf, Martin Lingonblad, and Annika Wallin

1 Introduction

Being familiar with an environment can have great effects on our eye movements and visual attention. With experience, we learn to attend to things that are important to us and to ignore less relevant information (Droll, Gigone & Hayhoe, 2007; Haider & Frensch, 1999; Jovancevic-Misic & Hayhoe, 2009; Meisner & Decker, 2010). This is one of the reasons why it is preferable to take studies out of the lab and into more natural environments. The present study investigated the effect of familiarity on visual attention and choice of consumers doing their grocery shopping. In a familiar environment where we make decisions repeatedly, our visual attention will change over time. For instance, familiarity with the task and task environment reduces time and effort and the relative influence of bottom up factors will decrease (Orquin, Bagger & Loose, 2013).

2 Experiment

The eye movements of fifty consumers were recorded in their familiar supermarket (group 1). All participants were instructed to buy a product from three different product categories during their regular shopping. These consumers were later recorded in another, unfamiliar supermarket of the same supermarket chain, again instructed to buy products from the same categories. A control group of twenty five consumers, familiar with the second supermarket, was also recorded (group 2), performing the same task, to determine if there was in fact familiarity and not differences in shelf organization or any other peculiarities of the second supermarket that made an impact on the results.

The SMI ETG glasses were used for this study, recording binocular eye movements at a rate of 30Hz. The collected eye tracking data was analysed manually frame-by- frame, using semantic gaze mapping in the SMI BeGazeTM Software by four independent coders. The fixations on AOI's where then divided into dwells according to the definition stated in Gidlöf et al. (2013) i.e. all fixations within an AOI for a duration of at least 120msec.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_13

The quality of the decision was calculated by having participants fill out a questionnaire of how important each attribute of a product category were for them when choosing a product. By summing up the values each product got an option quality (Gidlöf et al., 2013).

3 Preliminary Results and Discussion

All results are based on two of the three product categories since the third category is still being analysed. In general participants were more familiar with the yogurt category compared to the pasta category. Participants did spend more time inspecting the products in the unfamiliar supermarket compared to the familiar one, specifically in the pasta category. A paired t-test on group1 suggests an increase in total time between group1 in supermarket No.1 and in supermarket No.2. This difference was also reflected in the quality of their choices with participants making significantly better choices in pasta category in their familiar supermarket compared to the unfamiliar supermarket. The results showed no general differences between group 1 in supermarket no.1 and the control group in supermarket no.2.

The present findings are a first step in examining the effect of familiarity on visual attention and choice. These effects will be investigated further by for example studying the top-down and bottom-up influences on visual attention in these environments.

ACKNOWLEDGMENTS: The authors would like to thank Erik Jansson for his contributions in recording and analysis of the data.

References

1. Droll, J. A., Gigone, K., & Hayhoe, M. M. (2007). Learning where to direct gaze during change detection. *Journal of Vision*, *7*(14), 6.
2. Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 172.
3. Jovanevic-Misic, J., & Hayhoe, M. M. (2009) Adaptive Gaze Control in Natural Environ ments. *Journal of Neuroscience*, *29*, 6234 – 6238
4. Meissner, M., & Decker, R. (2010). Eye-tracking information processing in choice-based conjoint analysis. *International Journal of Market Research*, *52*(5).
5. Orquin, J. L., Bagger, M. & Loose, S. M. (2013),"Learning affects top down and bottom up modulation of eye movements in decision making", *Judgment and Decision Making*, 8(6),

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_13

700-716.

6. Gidlöf, K., Wallin, A., Møgelvang-Hansen, P., & Holmqvist, K. (2013). Material Distortion of Economic Behaviour and Everyday Decision Quality. *Journal of Consumer Policy*, *36*(4), 389-402.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_13

Part 8

Poster Session

Towards Using Eyetracking Data as Basis for Conversation Analysis on Real-World Museum Interactions

Raphaela Gehle, Antje Amrhein, Maximilian Krug, and Karola Pitsch

1 Introduction

The growing amount of disciplines using (mobile) eyetracking (glasses) leads to the need of reflecting the opportunities and restrictions that derive from using these devices in interdisciplinary contexts or when expanding traditional methods. In this paper, we evaluate aspects of compatibility of eyetracking with methods of qualitative video-analysis based on Conversation Analysis [4].

For studies of real-world human interactions, analysing the gaze conduct is as important as difficult. It's not just that a manual annotation of gaze is time-consuming but also there might be insecurity depending on the videoquality. Eyetracking appears to be promising for improvements of interaction analytic methods. Nevertheless, it requires to take a step backwards during the analysis to reflect the effects of using mobile eyetracking for the analysis.

In recent years, the interest in developing mobile eyetracking technology in interactional studies grows (e.g. [1]), because it presents the prospect of higher precision and lower costs. Recent work seeks for ways of automated annotation [3] depending on fixation estimation.

2 Developing Conversation Analysis on two types of data of the same situation

We will compare results from a step-by-step sequential analysis of the same sitiation with (1) external cams and (2) data of a mobile eyetracker.

Does eyetracking data alone provide for solid results about the structures and functionality of gaze in interaction?

As common in Conversation Analysis, we reconstruct the participants' process of interpretation during the ongoing interaction ("members perspective") [cf. 4]. In a pre-study we equipped a museum guide with a pair of mobile eyetracker glasses (SMI, version 1) and additionally deploied two external cameras [2].

(1) The video data of the *external cams* show, that the guide explains an exhibit while visitor A changes his orientation towards another visitor B. A's

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_15

reorientation towards B gets recognized and made interactionally relevant through the guide's reaction (looking at A and afterwards addressing him verbally) - *Result:* It appears that the A's reorientation towards another group member *elicits* the guide's gaze movement.

(2) The *eyetracking data* shows, that the guide looks at the exhibit, then moves his gaze over the group and ends with a fixation on another visitor. Afterwards the guide shifts his gaze back to A, so that the fixation is on A, who is already oriented to B. - *Result*: It appears that the guide *accidentally recognizes* A's orientation towards B while moving his gaze.

3 Discussion

Using data of mobile eyetracking glasses for interaction studies encompasses the risk of considering fixations as key-element of the participants' gaze conduct. Considering the members' perspective, this appears to be only *one* aspect. Other important information consists of the peripheral view and anticipation processes of the participants' conduct. Humans use multiple aspects of interactional coordination to make decisions over their next actions.

For our ways of using eyetracking data in interactional studies, these observations raise the question, how to deal with precise information about fixations and how to consider aspects of perception in peripheral vision when attempting to develop (semi-)automatic annotation tools in interdisciplinary collaboration.

ACKNOWLEDGMENTS: The authors acknowledge the financial support from the Volkswagen Foundation (Dilthey Fellowship 'Interaction & Space', K. Pitsch).

References

1. Oben, Bert; Brone, Geert (2011). Bidirektionalität in der multimodalen Interaktion: zum Potential von Eyetracking-Methoden zur Studie interaktiver Alignmentprozesse. 15. Arbeitstagung zur Gesprächsforschung. Kognition in Interaktion. Mannheim
2. Pitsch, K., Gehle, R., Wrede, S. (2013) A museum guide robot: Dealing with multiple participants in the real-world. Workshop ICSR 2013. Bristol, UK
3. Renner, P., Pfeiffer, T. (2014). Model-based acquisition and analysis of multimodal interactions for improving human-robot interaction. Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '14
4. Sidnell, J., Stivers, T. (2012) Handbook of Conversation Analysis. Wiley-Blackwell.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_15

Online Calibration of Gaze Tracking on the Computer Screen using Particle Filtering

Marc Halbrügge

1 Introduction

Visual salience is usually not evenly distributed on the computer screen. Most of the time, there is only a limited number of highly salient points instead. These points naturally attract fixations. If the points of visual salience are known to a gaze tracking system, its calibration could be inspected and also enhanced on their basis. But how should gaze be mapped to one of the salient points during this process? This mapping can not be established from a single observation. Particle filtering (Kantas et al., 2009) allows to maintain several hypotheses about where the user is currently looking and integrates this with where the user has looked before, finally converging to a most likely gaze position. We are presenting a particle filter that estimates the calibration drift of a head-based gaze tracking system.

2 The Problem of Calibration in Eye-Tracking

Eye-tracking, whether applied in the lab or in the field, needs calibration (Poole and Ball, 2005). While this is usually bearable for normal subjects, standard gaze calibration procedures often fail for children or some patient groups because they cannot keep their gaze on a fixation point for the required amount of time. In this case, the identification of "correct" gazes during calibration has to be done by the experimenter in a tedious and lengthy process (c.f., Franchak et al., 2011). Pursuit calibration (Pfeuffer et al., 2013) could be a solution to this, but is still in its infancy.

How can we alleviate this situation? We propose to use particle filtering to estimate the gaze position. Particle Filters have been applied to the machine vision side of eye-tracking, i.e., iris detection and tracking, before (Hansen and Pece, 2005).

Quality & Usability Lab, Telekom Innovation Laboratories, Technische Universität Berlin, Ernst-Reuter-Platz 7, 10587 Berlin e-mail: marc.halbruegge@tu-berlin.de

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_16

Here we extend this idea to the calibration and gaze estimation part of the problem. As a starting point, we try to estimate a time dependent drift of the estimated gaze position. Drift may be caused by, e.g., the headband being dragged down by the weight of the tracker, or subject actions like lifting the eyebrows.

3 Gaze Controlled HTML5 Shooting Game

As empirical basis for the drift estimation, data collected using a simple gaze controlled HTML5 shooting game based on Halbrügge (2015) was chosen. Task of the game was to "shoot" rectangles of a specific color by looking at them. To make gaze control easier, a small fixation cross was added to the center of the target objects (see Figure 1). The game took about three minutes to play with about 250 targets to shoot. The players' gaze was tracked using a head-based SR Research EyeLink II with 250 Hz sampling frequency; no chinrest was used.

The connection from the eye tracker to the game was established using SR Research's proprietary ethernet link and the web browser bridge described in Halbrügge (2015). Gaze interaction was provided by triggering virtual clicks on the browser's JavaScript layer after a dwell time of 300 ms. Neither eye tracker nor display latency was compensated as both were negligibly small relative to the dwell time.

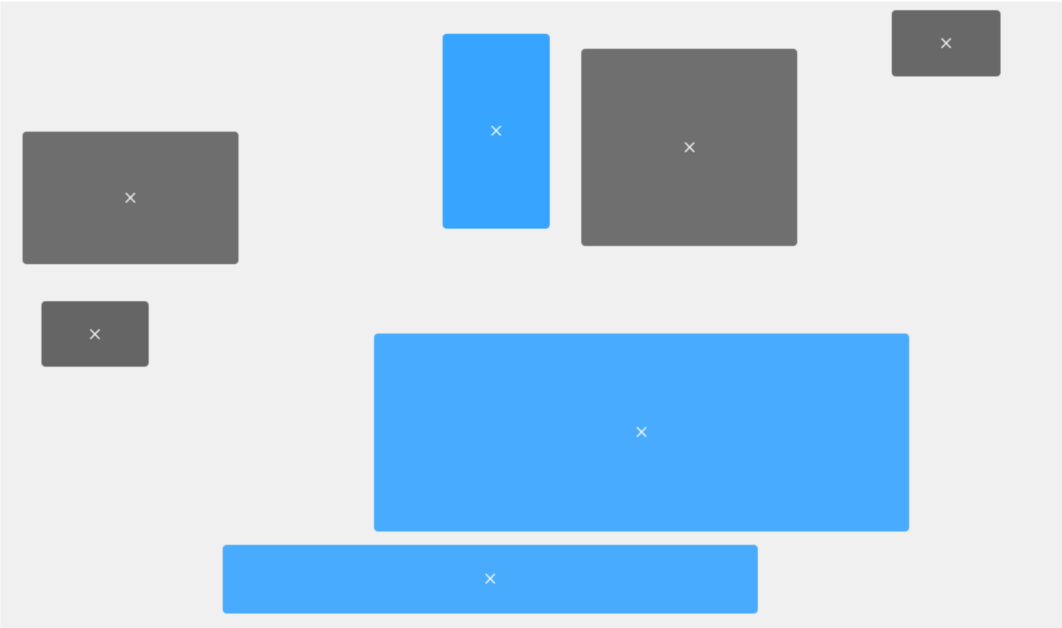


Fig. 1 Gaze controlled HTML5 game. The players have to "shoot" colored rectangles by looking at them. White crosses at the center of each target act as points of maximum salience.

Subsequent analysis showed that significant detoriation of the calibration happened during the gaming episodes. A fitted linear model showed substantial drift,

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_16

especially in the vertical dimension. This is consistent with the rather bulky and heavy headmount of the gaze tracker used for the game.

4 Particle Filter Results and Conclusions

Post-hoc analysis of the data using the particle filter approach showed that the detoriation of the calibration could easily be counteracted. As few as 1000 particles were enough to attain convergence and considerable error reduction. Each filter step needed less than 2 ms computation time on a consumer laptop computer, which would fast enough to run it in sync with the 250 Hz eye tracker used for the game.

As next step, we are planning to integrate the drift estimator into the tracking system. This should bring us closer to the goal of less intrusive calibration of gaze tracking systems. In the future, calibration could be performed by playing a simple game or watching a funny stickman video.

While this research has been guided by the goal of easier eye tracking for specific subject groups like children, and although it is currently restricted to HTML applications, it is not necessarily bound to these domains. Future applications could include calibration-free gaze interaction with large public displays, or continuous drift correction during mobile eye tracking.

ACKNOWLEDGMENTS: The authors gratefully acknowledges financial support from the German Research Foundation (DFG) for the project "Automatische Usability-Evaluierung modellbasierter Interaktionssysteme für Ambient Assisted Living" (MO 1038/18-1).

References

1. Franchak, J. M., Kretch, K. S., Soska, K. C., and Adolph, K. E. (2011). Head-mounted eye tracking: A new method to describe infant looking. *Child development*, 82(6):1738–1750.
2. Halbrügge, M. (2015). Automatic online analysis of eye-tracking data for dynamic HTML-based user interfaces. In Gramann, K. and Zander, T. O., editors, *Berliner Werkstatt Mensch-Maschine-Systeme*, Berlin. Technische Universität Berlin. accepted.
3. Hansen, D. W. and Pece, A. E. (2005). Eye tracking in the wild. *Computer Vision and Image Understanding*, 98(1):155–181.
4. Kantas, N., Doucet, A., Singh, S. S., and Maciejowski, J. M. (2009). An overview of sequential monte carlo methods for parameter estimation in general state-space models. In *15th IFAC Symposium on System Identification (SYSID), Saint-Malo, France.(invited paper)*, volume 102, page 117.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_16

5. Pfeuffer, K., Vidal, M., Turner, J., Bulling, A., and Gellersen, H. (2013). Pursuit calibration: Making gaze calibration less tedious and more flexible. In *Proc. of the 26th ACM Symposium on User Interface Software and Technology (UIST 2013)*, pages 261–270, New York, NY, USA. ACM.
6. Poole, A. and Ball, L. J. (2005). Eye tracking in human-computer interaction and usability research: Current status and future. In Ghaoui, C., editor, *Encyclopedia of Human-Computer Interaction*, pages 211–219, Hershey, PA. Idea Group.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_16

Automated Comparison of Scanpaths in Dynamic Scenes

Thomas C. Kübler and Enkelejda Kasneci

1 Introduction

According to the scanpath theory [6], the sequence of eye movements can give insights into the cognitive processes of visual perception. In this light, the comparison of scanpaths is of particular interest for many applications, e.g., driving [4], advertising, or activity recognition [1]. Although precise measurement devices and several approaches for the extraction of the visual scanpath in dynamic scenes exist, e.g. [7], the temporal and spatial comparison of scanpaths with their natural high variability remains challenging and laborious. Despite the variety of existing approaches, such as ScanMatch [2] or Multimatch [3], most scanpath studies are still limited to the comparison of time-integrated features, e.g. fixation duration or average saccade length. Due to the sequential nature of the scan pattern, such metrics are not sufficient to capture essential pattern characteristics of the viewing behavior. Recently, we proposed SubsMatch [5], a novel method for scanpath comparison in natural environments. In this work, we will demonstrate the performance of SubsMatch in comparison to the state-of-the-art when analyzing eye-tracking data derived from a driving scenario.

Thomas C. Kübler
University of Tübingen, Computer Engineering Department, Sand 13, 72076 Tübingen, Germany
e-mail: thomas.kuebler@uni-tuebingen.de

Enkelejda Kasneci
University of Tübingen, Computer Engineering Department, Sand 13, 72076 Tübingen, Germany
e-mail: enkelejda.kasneci@uni-tuebingen.de

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_17

2 SubsMatch Algorithm

Instead of addressing the issue whether two scanpaths focus on the same objects at the same time, SubsMatch addresses the question whether the scanpaths exhibit the same repetitive exploratory patterns. Figure 1 illustrates the main steps of SubsMatch as introduced in [5]. First, the scanpath is mapped to a string representation based on just one dimension of the scanpath, e.g. the horizontal position of fixations. Letter encoding is chosen to optimally map the scanpath data into equally frequent letters. This step also compensates for spatial offset and scaling, often caused by incorrect calibration, different seat positions while driving or different viewing distances towards the scene. The second step extracts all possible visual subpatterns from the encoded scanpath. The problem of scanpath comparison is thereby reduced to the comparison of the frequency of subpatterns.

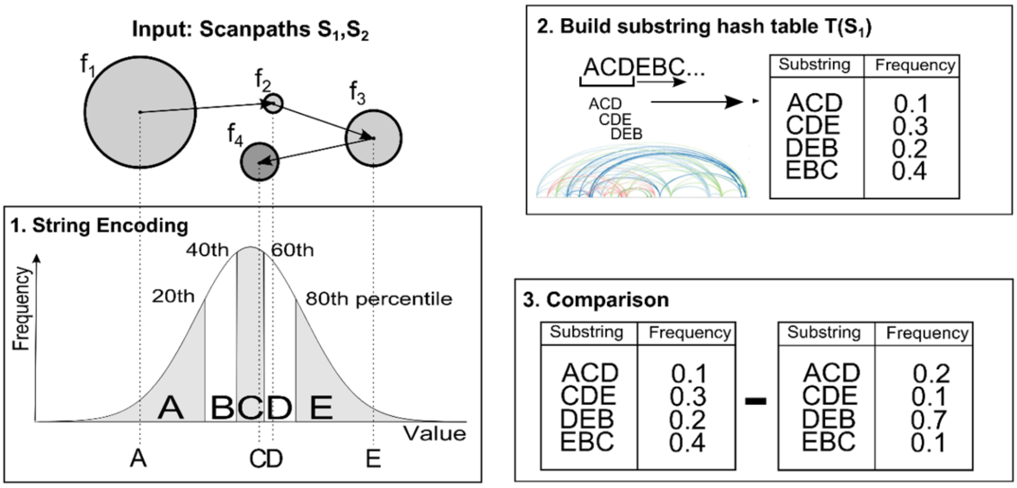


Fig. 1 Overview of the SubsMatch algorithm.

3 Exemplary Application

SubsMatch was applied to scanpaths derived from simulated driving sessions of 10 subjects with advanced glaucoma and 10 control subjects. Eye movements were recorded using a Dikablis eye tracker at 25Hz. 50% of the drivers with visual field defects were able to pass the driving test despite their visual impairment. This result indicates that the viewing behavior of patients who passed the test differs from the viewing behavior of patients who failed. Glaucoma subjects who passed (G_p) and failed (G_f) were compared to each other and to the control subjects (GC) to identify differences in the viewing patterns by means of SubsMatch, ScanMatch [2] and MutliMatch [3]. Table 1 shows the statistical evaluation of the pairwise scanpath

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_17

Table 1 Comparison of the pairwise distances within and between scanpath groups (G for glaucoma patients, p/f for passed or failed the driving test, C for control subjects).

	$\{G_f, G_p\}$	$\{GC, G_f\}$	$\{GC, G_p\}$
SubsMatch	**0.03**	**0.03**	0.67
ScanMatch	0.05	0.05	0.65
MultiMatch	0.89	0.89	0.89

distances between the groups. SubsMatch revealed that there are no additional compensatory eye movements performed by the G_p group, as often assumed. Instead, this group was able to maintain a normal exploratory behavior, similar to that in the GC group. The G_f group on the other hand exhibit-ed fewer horizontal, alternating saccades.

4 Conclusion

The presented algorithm is a significant step towards automated analysis and comparison of visual scanpaths derived from eye-tracking data in natural environments. Beyond further developments of SubsMatch, our future work will focus on the question, how the analysis of recurrent patterns can be used to provide added value for several applications.

References

1. Braunagel C, Kasneci E, Stolzmann W, Rosenstiel W. (2015). Driver-Activity Recognition in the Context of Conditionally Autonomous Driving. IEEE 18th International Conference on Intelligent Transportation Systems (ITSC), 2015
2. Cristino F, Matht S, Theeuwes J, Gilchrist ID. (2010). ScanMatch: a novel method for comparing fixation sequences. Behavior research methods, 42(3):692-700.
3. Dewhurst R, Nyström M, Jarodzka H, Foulsham T, Johansson R, Holmqvist K. (2012). It depends on how you look at it: scanpath comparison in multiple dimensions: MultiMatch, a vector-based approach. Behavior research methods 44(4):1079-100.
4. Kasneci E. (2013). Towards the Automated Recognition of Assistance Need for Drivers with Impaired Visual Field. PhD thesis,University of Tübingen, Wilhelmstr. 32, 72074 Tübingen.
5. Kübler TC, Kasneci E, Rosenstiel W. (2014). Subsmatch: Scanpath similarity in dynamic scenes based on subsequence frequencies. In Proceedings of the Symposium on Eye Tracking Research and Applications, ETRA 2014, 319–322, ACM.
6. Noton D, Stark L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. Vision research 11(9):929-42.
7. Tafaj E, Kübler TC, Kasneci G, Rosenstiel W, Bogdan M. (2013). Online classification of eye tracking data for automated analysis of traffic hazard perception. In Artificial Neural Networks and Machine Learning. ICANN 2013. 442–450. Springer Berlin Heidelberg.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_17

The influence of spatial occlusion on visual search behavior of karate athletes

Simon Salb, Markus Splitt, Nicole Bandow, Kerstin Witte

1 Introduction/Related Work

In martial arts, the ability to anticipate and response appropriately has to be seen as a performance determinant. Common methods in anticipation research are temporal and spatial occlusion, as well as eye tracking (Panchuk & Vickers, 2009; Savelsbergh, Williams, van der Kamp & Ward, 2002). For studying anticipatory skills in karate the presentation of attack scenes on a life size video screen (Williams & Elliott, 1999) using spatio-temporal occlusion proved to be a valid method (Bandow, Emmermacher, Stucke, Masik & Witte, 2014; Zerbe, E., Kirbach, M., Bandow, N., Emmermacher, P., Witte, K., 2013; Mori, Ohtani & Imanaka, 2002). Since there is evidence that spatial occlusion can influence the visual behavior of athletes (Hagemann, Schorer, Cañal-Bruland, Lotz & Strauss, 2010) it is important to consider this method critically. Prior studies did not record eye movements while applying spatial occlusion for studying anticipation in karate (Bandow et al., 2014). The aim of the present study is to explore the influence of spatial occlusion on the visual search behavior of karate athletes.

2 Our Contribution

Seven karate athletes with experience in national and international competition responded physically to life size videos of a karate reverse punch shown on a back projection screen (2 m x 2.5 m). Spatial occlusion was conducted by covering relevant body regions (hip, punching arm, front leg) of the opponent in the video sequences with background images. The participants' visual behavior was recorded with a binocular head mounted eye tracker (SensoMotoric Instruments; model: SMI Eye Tracking Glasses). Each video sequence was presented three times in randomized order to prevent a learning effect. Eye tracking data was analyzed using the software BeGaze (SensoMotoric Instruments, Teltow, Germany). Therefore nine areas of interest (AOI) were defined representing nine regions of the opponents' body relevant for karate kumite. As dependent measures relative fixation time per AOI, mean fixation duration per trial and the number fixations per second in each trial were chosen (Dicks, Button & Davids, 2010). A reliability analysis revealed no significant differences between three repeated measurements. By means of one-way ANOVA/ Kruskal-Wallis test significant differences between

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_18

the occlusion conditions were tested. The results matched with the findings of Williams & Elliott (1999) and showed primary fixation of the head and upper torso under all conditions (e.g. 58% head and 22% upper torso under non occluded condition). There was no difference in visual search behavior between occluded and non occluded stimuli. Neither mean fixation duration [$\chi^2(3)=4.543$, $p=.208$] nor number of fixations per second [$F(3,24)=1.747$, $p=.184$] showed significant differences between the occluded and non occluded conditions. Analyzing relative fixation duration per AOI confirmed those findings. For example, the relative fixation duration of the head did not significantly differ between conditions [$F(3,24)=.232$, $p=.873$].

3 Discussion

Due to the small sample size (n=7), variations between the subjects influence the results. Therefore it is required to gather more data in order to make general conclusions. It will be necessary to conduct further studies occluding regions of primary fixations (head, upper torso) to determine the influence of spatial occlusion on visual search of karate athletes. As long as spatial occlusion is conducted by making body regions invisible that are no main fixation locations, there seems to be no influence on visual search behavior.

References

1. Bandow, N., Emmermacher, P., Stucke, C., Masik, S. & Witte, K. (2014). Comparison of a video and a virtual based environment using the temporal and spatial occlusion technique for studying anticipation in karate. *International Journal of Computer Science in Sport, 13* (SPLISSUE.1), 44–56.
2. Dicks, M., Button, C. & Davids, K. (2010). Examination of gaze behaviors under in situ and video simulation task constraints reveals differences in information pickup for perception and action. *Attention, Perception, & Psychophysics, 72* (3), 706–720.
3. Hagemann, N., Schorer, J., Cañal-Bruland, R., Lotz, S. & Strauss, B. (2010). Visual perception in fencing: do the eye movements of fencers represent their information pickup? *Attention, perception & psychophysics, 72* (8), 2204–2214.
4. Mori, S., Ohtani, Y. & Imanaka, K. (2002). Reaction times and anticipatory skills of karate athletes. *Human Movement Science*, 21 (2), 213-230.
5. Panchuk, D. & Vickers, J.N. (2009). Using spatial occlusion to explore the control strategies used in rapid interceptive actions: Predictive or prospective control? *Journal of sports sciences*, 27 (12), 1249–1260
6. Savelsbergh, G. J., Williams, A. M., van der Kamp, J. & Ward, P. (2002). Visual search, anticipation and expertise in soccer goalkeepers. *Journal of sports sciences, 20* (3), 279–287.
7. Williams, A.M. & Elliott, D. (1999). Anxiety, expertise, and visual search strategy in karate. Journal of Sport & Exercise Psychology, 21 (4), 362–375.
8. Zerbe, E., Kirbach, M., Bandow, N., Emmermacher, P., Witte, K. (2013). Comparison of temporal and spatial occluded video karate attacks with regard to predictability. In Fac. of Elec-

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_18

trical Engineering (Hrsg.), International Scientific- Professional Symposium INFOTEH (Vol. 12, S. 944–950).

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_18

Capturing gaze behavior patterns of surfers during surfboard riding: A pre-study in testing a water housing system for mobile eye tracking technology

Martin Walz, Paul Günther, Guido Ellert

1 Introduction/Related Work

Due to the fact that the sport of surfboard riding occurs in a fluid and ultra dynamic environment it is to be expected that several visual focuses of attention are leading to either better or worse outcomes of performance [1]. Research in several sport disciplines shows that there are marked differences in visual search strategies between expert athletes and recreational athletes[3] and that gaze training improves athlete performance[4]. The quantification and visualization of gaze behavior patterns by fixation count and fixation duration on *AoI* (Areas of Interest) represents an extensively applied method in sport research [3; 5]. To record gaze behavior patterns from surfers, a water-housing prototype (see Fig. 1) for mobile eye tracking technology was built. Hence, the principle aim of this work is to test the operability of the water housing system. Through the theoretical baseline of the constraints-led perspective [2] and the expert performance approach [5], the secondary aim is a qualitative description of vision in action during surfboard riding for the first time and also to identify visual hot spots of expert surfers whilst performing carving turns.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_19

Fig. 1. An expert surfer performs a carve turn whilst wearing the *Tobii Glasses 2* in the water-housing system and the recording-unit on his back.

2 Our Contribution

To control the research constraint "environment" the study was held at the stationary river wave (Eisbach) in Munich. As subjects, 2 expert surfers (professional status) and 2 recreational surfers were analyzed. The constraint "task" was defined as doing 5 carve turns per surfing run. The subjects had to perform 2 runs. As research instrument the mobile eye tracker model *Tobii Glasses 2* was used and the software analysis was performed with the *Tobii Pro Glasses Analyzer*. To process and classify the recorded gaze data samples into fixations and other eye movement categories the *Tobii I-VT fixation filter* was used. Through qualitative gaze plots presented with heat maps and scan paths, the results are showing intersubjective differences in fixation duration and fixation count between experts and recreational surfers (see Fig. 2 and Fig. 3).

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_19

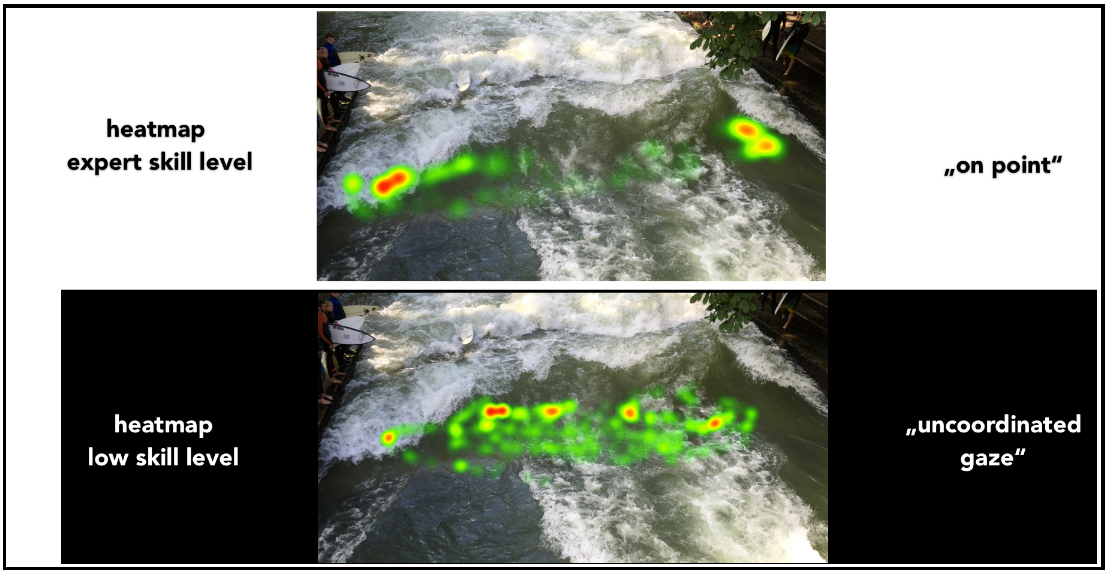


Fig. 2. Cumulative visual hot spots of expert surfers (n = 2) and recreational surfers (n = 2) showing subjective differences during 5 turns.

Fig. 3. Eye movement patterns showing qualitative relations to the surfboard direction.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_19

3 Discussion

Because of differences in gaze recognition rate by the hardware system, which varied from 32 percent to 87 percent between the subjects, no inference statistic analysis was performed. This research represents a pilot study for eye tracking in water sports. Therefore, the method, treatment, analysis and research prospects are open for discussion. How *quiet eye* studies can be performed with the presented materials in water environments, are to discuss on an interdisciplinary level.

ACKNOWLEDGMENTS: The authors would like to thank the engineering team of the water housing prototype: Heiko Pfisterer, Wolfrik Fischer, Martin Söllner and Manuel Angstl. The authors also want to thank the whole research team from the Macromedia University of applied sciences Munich: Fabio Rauscher Bascón, Johannes Michel and Simon Dallwig

References

1. Mendez-Villanueva, A., & Bishop, D. (2005). Physiological aspects of surfboard riding performance. *Sports Medicine (Auckland, N.Z.)*, *35*(1), 55–70.
2. Newell, K. M. (1985). Coordination, Control and Skill. In R. B. W. and I. M. F. David Goodman (Eds.), *Advances in Psychology*, *27*, 295–317, North-Holland.
3. Vickers, J. N. (2007). *Perception, Cognition, and Decision Training: The Quiet Eye in Action*. Human Kinetics.
4. 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.
5. Williams, A. M., & Ericsson, K. A. (2005). Perceptual-cognitive expertise in sport: Some considerations when applying the expert performance approach. *Human Movement Science*, *24*(3), 283–307.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_19

Gaze interaction for multi-display systems using natural light eye-tracker

Onur Ferhat, Arcadi Llanza, Fernando Vilariño

1 Introduction

Natural light-based gaze-tracking constitutes an interesting alternative to commercial eye-trackers, and it requires no additional hardware other than a camera (ubiquitously present on consumer devices). Methods developed to attack this problem include using raw eye appearance as the feature set [1], using the geometry of iris boundary ellipse for estimation [2, 3, 4] and using more complicated geometrical models [5, 6], among others. The problem with these methods is either they are using unrelated information (skin pixels), or they are summarising the necessary information too much (to model parameters).

We propose a system implementing a novel feature based approach for webcam gaze interaction. Our system relies on the automated detection of the iris bounding box, followed by iris segmentation by means of binarisation. We develop this approach using the OpenGazer eye-tracker as a baseline.

As a validation framework, we propose a two-display context-adaptive system which modifies in real time the data rendered in a secondary display depending on the particular object that is observed in the main display. We particularly tune this approach to Google-glass type devices. The application context of our research can be straightforwardly tested and extrapolated to different augmented reading expe-

Onur Ferhat
Universitat Autònoma de Barcelona and Computer Vision Center, Edificio O Campus UAB 08193 Bellaterra e-mail: oferhat@cvc.uab.es

Arcadi Llanza
Universitat Autònoma de Barcelona and Computer Vision Center, Edificio O Campus UAB 08193 Bellaterra e-mail: arcadi.llanza@e-campus.uab.cat

Fernando Vilariño
Universitat Autònoma de Barcelona and Computer Vision Center, Edificio O Campus UAB 08193 Bellaterra e-mail: fernando@cvc.uab.es

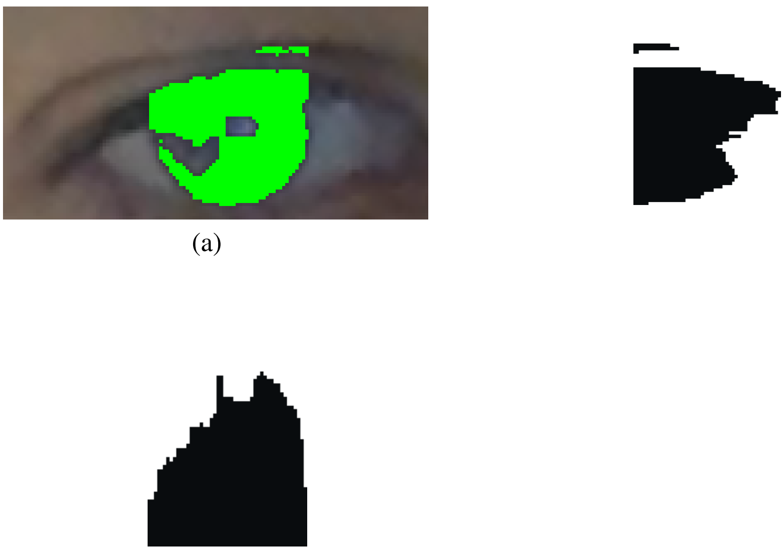
Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_4

riences such as maps analysis, artworks detailed observation and music reading, among others.

2 Our Approach

The system performs an **automatic selection** of 8 anchor points on the subject's face based on facial features such as eye corners, eyebrows, nose and mouth [7, 8]. Over the next frames, optical flow is used to update the points' positions in the new images.

The **feature extraction** stage takes as input the eye images extracted from the camera image. The iris bounding box is detected with template matching and iris segmentation is calculated using a binarisation inside this area. The features are extracted by projecting the segmented pixels into the vertical and horizontal axes and a feature vector is created by concatenation as seen in Fig. 1. For **gaze estimation**, we train a Gaussian Process with the standard squared exponential kernel tuned to work at a desired distance during the calibration process.



(a) (b)

(c)

Fig. 1 (a) Iris segmentation (b) Vertical features calculated by projecting iris pixels to vertical axis (c) Horizontal features

The system is complemented with a gaze-interactive interface as shown in Fig. 2. On the first display, visual selection is natively performed by guiding the gaze towards the desired object (left). Real-time output is then rendered in the secondary display (top-right) which adapts its information to the item visually selected (in the context of Google-glass type devices the secondary display plays the role for the glass display).

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_4

Fig. 2 System interface: (Left) Main visualization display where stimuli are shown. (Top-Right) Responsive display: The information is adapted to the object visualized. (Bottom-right) Subject's facial reference points and segmented images from the eyes shown for illustrative purposes.

3 Discussion

The proposed method decreases the average gaze estimation error by 34% in horizontal direction and by 12% in the vertical direction when compared to our previous work on standard databases [8]. The average errors are **2.35°and 1.82°** for the baseline system in both directions; whereas the proposed system achieves error rates of **1.54°and 1.61°**. Our experiments show that the errors are not uniform over the screen area, and horizontal errors can be up to 65% larger near the edges compared to the center of the screen.

The interaction obtained in this way is smooth and fast. The preliminary results are promising, and our current work is focused on **enriching the feature set** by introducing spatial feature correlation, and using other cues such as exact eye corner positions. Regarding interaction, the definition of suitable visual action triggers that handle with the information delivered in the secondary display, not defined in this work, is an open area of research.

ACKNOWLEDGMENTS: This work was supported by UAB grants and the Google Faculty Award.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_4

References

1. Holland C, Komogortsev OV (2012). Eye tracking on unmodified common tablets: challenges and solutions. Eye Tracking Research and Applications. 277280.
2. Wood E, Bulling A (2014). EyeTab: Model-based gaze estimation on unmodified tablet computers. Eye Tracking Research and Applications. 207-210.
3. Wang JG, Sung E (2000). Gaze determination via images of irises. British Machine Vision Conference.
4. Wu H, Chen Q, Wada T (2004). Conic-based algorithm for visual line estimation from one image. Automatic Face and Gesture Recognition. 260265.
5. Chen J, Ji Q (2000). 3D gaze estimation with a single camera without IR illumination. International Conference on Pattern Recognition. 14.
6. Ishikawa T, Baker S, Matthews I, Kanade T (2004). Passive driver gaze tracking with active appearance models. World Congress on Intelligent Transportation Systems.
7. Santana MC, Déniz-Suárez O, Hernández-Sosa D, Lorenzo J (2011). A comparison of face and facial feature detectors based on the Viola-Jones general object detection framework. Mach. Vis. Appl. 22(3), 481494.
8. Ferhat O, Vilarino F, Sanchez FJ (2014). A cheap portable eye-tracker solution for common setups. Journal of Eye Movement Research, 7(3):2, 1-10.

Published at SAGA Workshop 2015, proceedings archived under DOI: 10.2390/biecoll-saga2015_0.
Individual article archived under DOI: 10.2390/biecoll-saga2015_4

Part 9

Appendix

SAGA Solutions for Automatic Gaze-Data Analysis

Programm

The workshop is hosted at the CITEC building (Inspiration 1/33619 Bielefeld).

Tuesday, September 29th, 2015		
08:30	Registration desk opens / Coffee	Foyer
09:00	Welcome	CITEC Auditorium
09:15	**Keynote by Jacob Lund Orquin: What eye tracking researchers (dis)agree about reporting**	CITEC Auditorium
10:15	Coffee & Cake Break	Foyer
10:45	**Session 01: Automatic Analysis** > **GazeVideoAnalyser: A Modular Software Approach Towards Automatic Annotation of Gaze Videos** *by Kai Essig, Dato Abashidze, Manjunath Prasad and Thomas Schack* > **Semi-automatic annotation of eye-tracking recordings in terms of human torso, face and hands** *by Stijn De Beugher, Geert Brône and Toon Goedemé* > **Capturing and Visualizing Eye Movements in 3D Environments** *by Thies Pfeiffer, Cem Memili and Patrick Renner*	CITEC Auditorium
12:00	Lunch Break	Mensa
13:30	**Session 02: Human Computer Interaction** > **Toward implicit human-machine interaction: Single-trial classification of fixation-related potentials** *by Andrea Finke, Kai Essig, Giuseppe Marchioro and Helge Ritter* > **Online Visual Attention Monitoring for Mobile Assistive Systems** *by Patrick Renner and Thies Pfeiffer* > **Gaze Tracking for Human Robot Interaction** *by Oskar Palinko, Francesco Rea, Giulio Sandini and Alessandra Sciutti*	CITEC Auditorium
14:30	**Poster & Demo Session** / Coffee Break > **Towards Using Eyetracking Data as Basis for Conversation Analysis on Real-World Museum Interactions** *by Raphaela Gehle, Antje Amrheim, Maximilian Krug, and Karola Pitsch* > **Online Calibration of Gaze Tracking on the Computer Screen using Particle Filtering** *by Marc Halbrügge* > **Automated Comparison of Scanpaths in Dynamic Scenes** *by Thomas C. Kübler and Enkelejda Kasneci* > **The influence of spatial occlusion on visual search behavior of karate athletes** *by Simon Salb, Markus Splitt, Nicole Bandow, and Kerstin Witte* > **Capturing gaze behavior patterns of surfers during surfboard riding: A pre-study in testing a water housing system for mobile eye tracking technology** *by Martin Walz, Paul Günther, and Guido Ellert* > ***Gaze interaction for multi-display systems using natural light eye-tracking*** *by Onur Ferhat, Arcadi Llaza and Fernando Vilariño* > **Demos see http://saga.eyemovementresearch.com/program/**	Foyer & Labs
16:00	**Keynote by Mark Williams: Visual search behaviour and expertise in high-performance environments**	CITEC Auditorium
17:00	Coffee Break	Foyer
17:15	**Session 03: Visual Expertise & Motor Performance** > **Maintenance of perceptual-cognitive expertise in volleyball under different time constraints** *by Lennart Fischer, Joseph Baker, Judith Tirp, Rebecca Rienhoff, Bernd Strauss and Jörg Schorer* > **Exploring the Relationship Between Motor Imagery, Action Observation, and Action Execution in Motor Skill Learning** *by Alessio D'Aquino, Cornelia Frank, Kai Essig and Thomas Schack* > **Into the wild – Musical communication in ensemble playing. Discerning mutual and solitary gaze events in musical duos using mobile eye tracking** *by Sarah Vandemoortele, Stijn De Beugher, Geert Brône, Kurt Feyaerts, Toon Goedemé, Thomas De Baets and Stijn Vervliet*	CITEC Auditorium
19:30	Social Event at Bernstein, Niederwall 2, 33602 Bielefeld	City Center, close to "Jahnplatz"

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

September 30th, 2015		
09:15	**Keynote by Maria Staudte: Studying gaze in spoken interaction**	CITEC Auditorium
10:15	Coffee & Cake Break	Foyer
10:45	**Session 04: Methods for Measuring & Quantification** **> Analyzing Patterns of Eye Movements in Social Interactions** *by Nadine Pfeiffer-Leßmann, Patrick Renner and Thies Pfeiffer* **> A Metric to Quantify Shared Visual Attention in Two-Person Teams** *by Patrick Gontar and Jeffrey B. Mulligan* **> The impact of image size on eye movement parameters** *by Ricardo Ramos Gameiro, Kai Kaspar, Sontje Nordholt and Peter König*	CITEC Auditorium
12:00	Lunch Break	Mensa
13:30	**Session 05: Language & Memory** **> Language and gaze cues: findings from the real world and the lab** *by Ross Macdonald, Maria Staudte and Benjamin Tatler* **> Collecting memories: the impact of active object handling on recall and search times** *by Dejan Draschkow and Melissa L.-H. Võ*	CITEC Auditorium
14:20	Coffee Break	Foyer
14:30	**Session 06: Real World Studies & Applications** **> Gaze Analysis in Mobile Pedestrian Navigation: Culture Affects Wayfinding Styles** *by Lucas Paletta, Stephanie Schwarz, Jan Bobeth, Michael Schwarz and Manfred Tscheligi* **> Effect of Familiarity on Visual Attention and Choice at the Point of Purchase** *by Kerstin Gidlöf, Martin Lingonblad and Annika Wallin*	CITEC Auditorium
15:20	Closing Session	CITEC Auditorium
20:00	Optional social event for guests leaving the next day	To be announced

9.2 Location / Map of the CITEC building

The SAGA Workshop 2015 is hosted by CITEC with the 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).

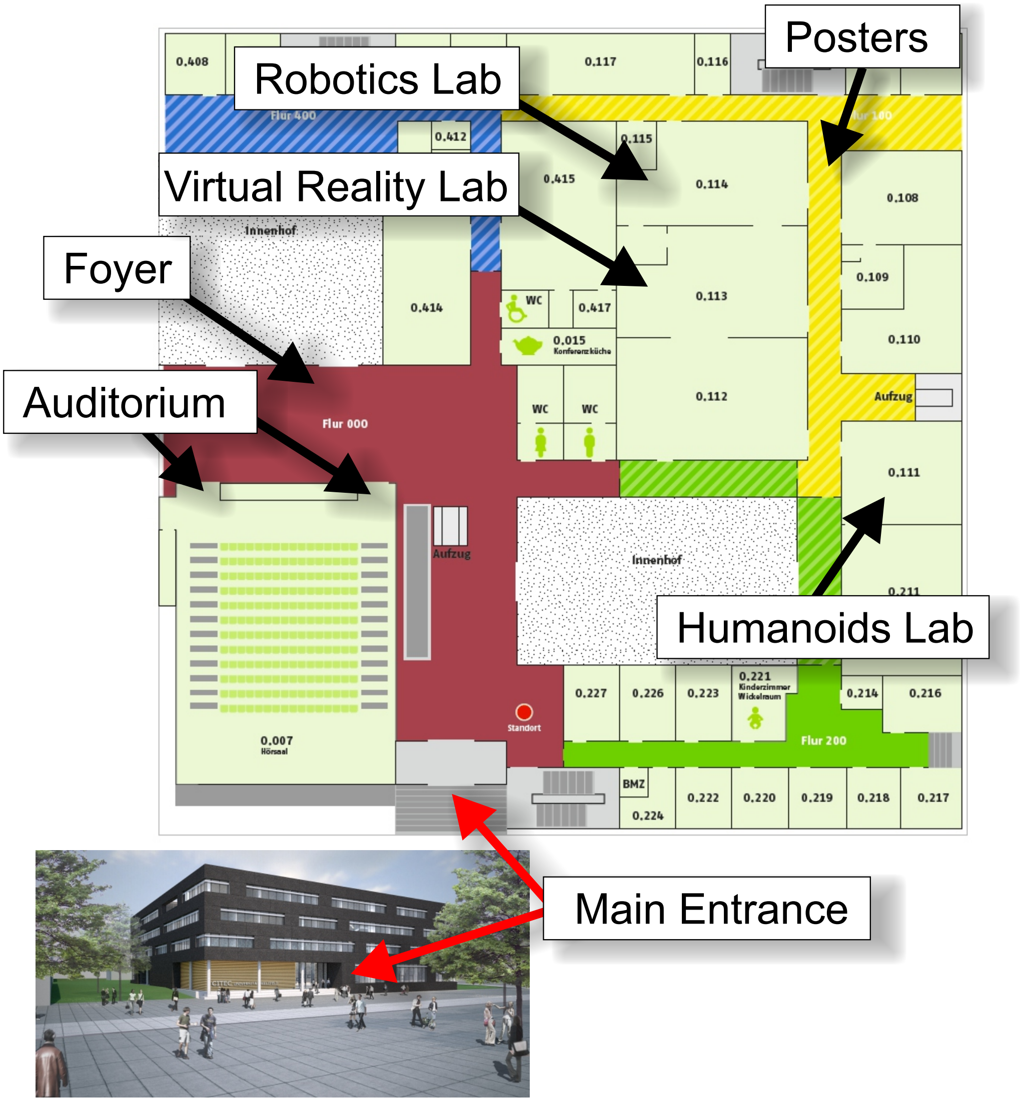


Figure 1: Map of the ground floor of the new CITEC building. The main events of the SAGA Workshop take place in the Lecture Hall, the Foyer and the VR-Laboratory in 0.113.

9.3 Social Event on Tuesday 29th

On the first evening, we will meet at the **Bernstein** for Dinner at 7:30 p.m. (19:30 Uhr). **Bernstein** is a Restaurant located at the City Center of Bielefeld, close to the "Jahnplatz" (accessible e.g. by Stadtbahn/Tram/Subway with all lines using station "Jahnplatz"). The address is:
Bernstein
Niederwall 4, entry in "Renteistrae"
33602 Bielefeld

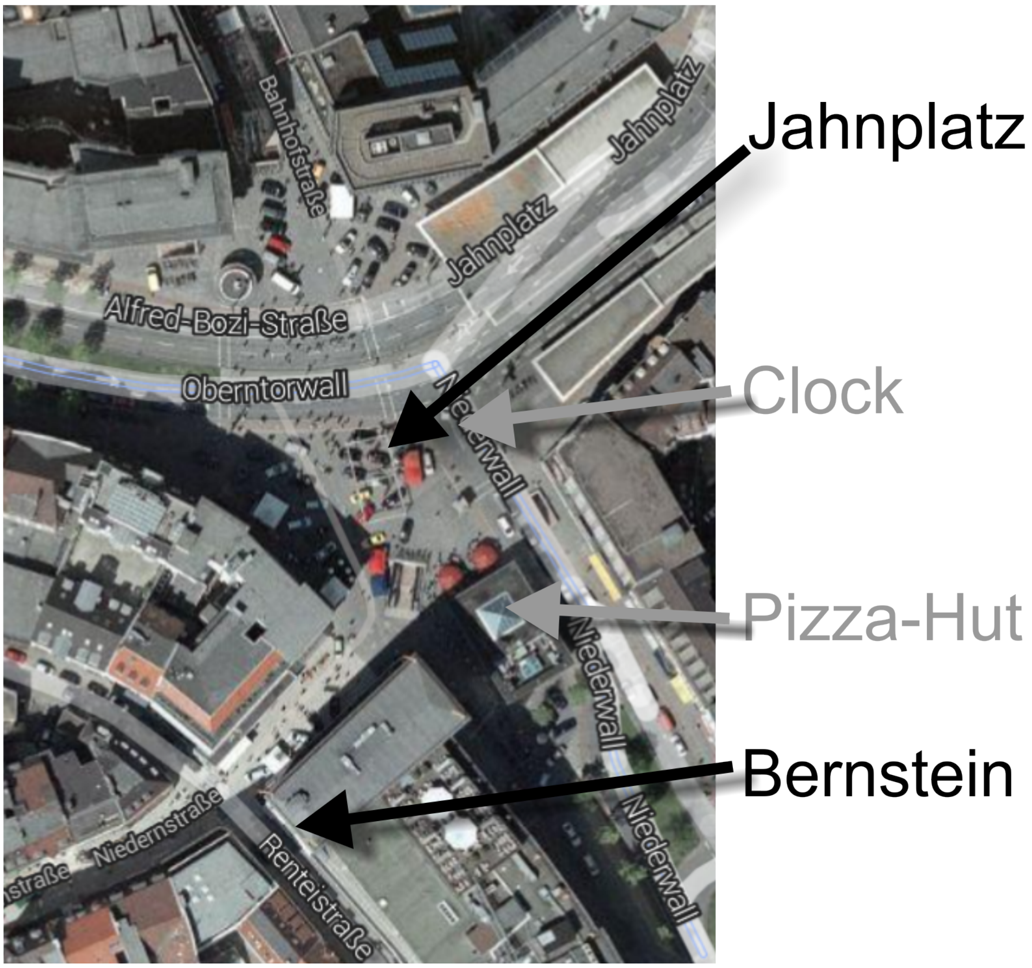


Figure 2: Map of the local area around Bernstein, the restaurant for the social event on Thursday.