ADAMAAS - Towards Smart Glasses for Mobile and Personalized Action Assistance

Kai Essig
CITEC - Center of Excellence Cognitive Interaction Technology
Bielefeld University
Inspiration 1
33619 Bielefeld, Germany
kai.essig@uni-bielefeld.de

Benjamin Strenge
CITEC - Center of Excellence Cognitive Interaction Technology
Bielefeld University
Inspiration 1
33619 Bielefeld, Germany
benjamin.strenge@uni-bielefeld.de

Thomas Schack
CITEC - Center of Excellence Cognitive Interaction Technology
Bielefeld University
Inspiration 1
33619 Bielefeld, Germany
thomas.schack@uni-bielefeld.de

ABSTRACT

In this paper, we describe the assistive system ADAMAAS *(Adaptive and Mobile Action Assistance)* introducing a new advanced smartglasses technology. The aim of ADAMAAS is to move from stationary status diagnosis systems to a mobile and adaptive action support and monitoring system, which is able to dynamically react in a context sensitive way to human error (slips and mistakes) and to provide individualized feedback on a transparent virtual plane superimposed on user's field of view. For this purpose ADAMAAS uses advanced technologies like augmented reality (AR), eye tracking, object recognition, and systematic analysis of users' mental representations in long term memory. Preliminary user tests with disabled participants at an early prototype stage revealed no substantial physical restrictions in the execution of their activities, positive feedback regarding the assistive hints, and that participants could imagine wearing the glasses for long periods of time.

Figure 1: ADAMAAS will provide situation-, user-, and action-specific feedback on a transparent virtual plane superimposed on the user's field of view.

CCS Concepts

•**Human-centered computing** → **Mixed / augmented reality;** *Ubiquitous and mobile computing;* HCI design and evaluation methods; •**Applied computing** → Computer-assisted instruction; •**Computing methodologies** → Activity recognition and understanding;

Keywords

Assistive Systems; Individualized Feedback; Augmented Reality (AR); Scene and Action Understanding; Eye Tracking

1. INTRODUCTION

Most people have encountered situations like the following: You want to bake a specific cake, but cannot remember the recipe, or you do not know how to use the new high-tech kitchen stove to prepare a meal or how to repair a bicycle. In such everyday situations it may be helpful to receive unobtrusive and intuitive support from an adaptive technical system that operates in a largely unnoticed and restriction-free manner. ADAMAAS focuses on the development and evaluation of intelligent glasses for this purpose (Fig. 1). It combines techniques from human memory research, eye tracking and vital parameter measurement (such as pulse or heart rate variability), object and action recognition (computer vision), as well as AR with modern diagnostics and corrective intervention techniques. The system will be able to identify problems during action processes, to react when mistakes are made, as well as to display situation- and context-dependent assistance by superimposing helpful information on a transparent virtual plane in users' fields of view. The mobile ADAMAAS assistance system aims to provide support for people to be able to live an independent life in an age-appropriate way, according to their mental and physical capabilities. Thus, the system will be able to suggest new action options and to support learning processes. All in all, the aim of ADAMAAS is to evolve from stationary status diagnostic systems to a mobile and dynamically adaptive monitoring system which is able to react to hu-

PETRA '16, June 29 - July 1, 2016, Corfu Island, Greece

DOI: xxxx/xxxxxxxxx

man failures and to provide individualized prompting and feedback for action support.

2. ASSISTIVE SYSTEMS

The earliest attention-based assistance systems were those for gaze-typing or mimicking mouse movements, which allowed people with restricted (manual) interaction capabilities to operate a computer by eye movements. An overview of different gaze-based assistive technologies for desktop computers, 3D Virtual Reality and natural environments can be found in [1]. One problem for designing such systems is that eye movements often occur involuntarily and thus fixations do not necessarily indicate a an intent to interact with a particular element at the screen [2]. This is described as the Midas Touch problem [3]. The authors in [2] address this problem by using extended fixation duration or blinking to select keys on a virtual keyboard in a text typing task. They found that participants rate fixations as quicker, easier to use and less tiring than blinking. However, blinking is more suitable for sparse and non-continuous input.

A current trend, enhanced by the development of more robust and powerful mobile eye tracking systems, is to support environmental control by gaze, mobile applications and gaze-based mobility, and control of physical objects and human-robot interaction [4]. Recent application fields for mobile assistive systems include, among others, chess assistance [5] and the display of nutrition information when customers look at products in a supermarket [6]. Wearable systems such as Google Glass, Microsoft HoloLens, the Motorola HC1, Golden-i and the AITT system float 2D/3D images on a Prisma or Waveguide display in front of the eyes and overlay them on top of the real-world. The HoloLens provides also head-movement based gaze control and gesture recognition via depth sensing cameras, while Google Glass has a touch pad and speech. These new technologies enable new forms of interaction with the user and the environment. These systems work already quite robust and can be used in several mobile environments. Currently, these systems mostly display previously annotated data (such as nutrition information stored in a database) when objects are recognized. They do not show a particular sensitivity to users' needs or to the actual action process by determining what the wearer is currently trying to do, e.g. cooking a meal. Our goal is to allow users to continue working on a task while receiving instructions on how to execute it better and prevent or correct mistakes as they happen. Therefore, ADAMAAS is a different approach in the wearable space, taking diagnostic systems that used to be stationary and making them mobile while also integrating monitoring technology that enables reaction to and individualized support for the user's actions in real time. Thus, it provides support particularly for the elderly and disabled that struggle with everyday tasks. To the best of our knowledge, this approach to interactive wearables with an on-the-spot diagnosis and feedback is unique.

3. THE ADAMAAS SYSTEM

In the following we describe the technical components of the ADAMAAS system and idiosyncrasies of our human-centered development approach.

3.1 Hardware

Figure 2: The ADAMAAS AR-Eyetracking System designed by SensoMotoric Instruments (SMI)

The hardware of the ADAMAAS system consists of the world's first eye tracking integration for AR, specifically designed for this project by our project partner SensoMotoric Instruments (SMI). It combines the Epson Moverio BT-200 see-through head-mounted display and the mobile binocular eye tracking platform from SMI (Fig. 2). The system does not only allow the gaze movements in the real environment to be tracked, but also those on the transparent virtual plane superimposed on the user's field of view (Fig. 1). The SMI SDK provides access to real-time gaze data synchronized with the AR display content. By continuously recording participants' eye movements, the content on the AR-plane can be dynamically updated on demand. The system has a sampling rate of 30 Hertz with an accuracy of $<0.6^\circ$ on the AR display and 0.5° otherwise over all distances and a weight of approximately 110 grams. The system uses a 1 to 5-point calibration process on the virtual plane. The Epson Moverio BT-200 provides a 0.42" LCD size with a resolution of 960 x 540 pixels, a field of vision of 23°, and a projection size of 40" at 2.5m. The ADAMAAS glasses and the Epson Moverio controller are connected via a USB 3.0 port to a Microsoft Surface Pro 4 Computer with a Intel Core i5 Processor, a 128 GB Hard Disk and 4 GB RAM using Windows 8.1. The eye tracking server and the ADAMAAS remote control component are running on the Surface computer. The application server and host are running on the Epson controller on Android 4.0.4. The scene image and the eye tracking coordinates are sent from the ADAMAAS glasses to the Surface computer. There the analysis (i.e. gaze coordinates, computer vision) is running and a corresponding update for the AR content is then sent to the Epson controller. The design choice to use the Surface computer in addition to the mobile Epson controller was made under the assumption that more powerful mobile units will be available in the near future, superseding the Surface computer.

3.2 Eye tracking

The ADAMAAS system provides real-time access to the scene video and the eye tracking data. If the eyes can be seen as an indicator for the brain's performance ("eyes are a window to the mind" [7]), the fixation duration can be considered as a measure of the effort of information processing. The longer visual attention is directed toward an object location, the longer it presumably takes us to deal with the information presented at that particular area of the scene. This relationship, known as the "eye-mind" hypothesis, is strongly supported by results from reading experiments ([8]). Furthermore, the number of fixations and the distribution of fixations reflect the degree of cognitive processing required for the understanding of particular scene regions. Long fixation durations and short saccade lengths signify a fine and deep processing of a particular image region, indicating that the understanding of its visual information is quite difficult. In contrast, long saccade lengths and short fixation durations indicate a fast and coarse scanning of a particular image region, signaling that the information content of that particular image region is easy to process or less important for the current task.

We use the number of fixations, cumulative fixation duration and attention shifts between objects or regions of interest as measures to identify the relevant cognitive processes. We have to take care that cognitive deficits, attention disorders or uncontrolled behavior may lead to misinterpretations of the recorded eye movements [9]. Therefore, we will perform extensive evaluation studies with healthy and disabled people in order to get deep insights into the perceptual differences and the informative value of the recorded eye tracking parameters. To the best of our knowledge, this topic has not yet been studied in detail in the literature. Most of the eye tracking systems available on the market are used for research and diagnostic purposes, but only a relatively small number are specifically aimed at people with disabilities [10]. This is particularly true for system providing action support for elderly or disabled people in mobile everyday scenarios.

3.3 Mental Representation Structures

In order to allow for age-appropriate, individual assistance in everyday situations, task representations in users' long-term memory will be assessed in a first step. Therefore, we use an approach based on the structural-dimensional analysis of mental representations [11] which is well established in cognitive psychology and movement science. It captures mental representation structures using a special splitting procedure and subsequent hierarchical clustering of actions. Dedicated software tools for desktop and mobile platforms are available for this purpose. The ADAMAAS system is then able to identify user-specific problems in task execution by comparing their mental representation structures to predefined feasible sequences of action steps. If action steps which must be executed consecutively (according to the feasible workflow sequences) are unassociated in users' long-term memory, the transition between those steps is expected to be error-prone. Based on this information, the current visual focus, and task-related objects detected in the environment, the system shall provide adequate feedback, i.e. by displaying helpful hints and illustrations regarding the pending action via AR. Laboratory studies are currently conducted to assess how accurately this approach predicts human errors.

At a later stage the ADAMAAS system shall also keep track of users' learning processes by storing timestamped information about their mistakes. The system can then extrapolate from the gathered data to project the current point on the individual learning curve. This shall lead to fewer and more accurate system interventions, thereby reducing the amount of distraction and facilitating the independent execution of activities by the user. Otherwise, consistent usage could provoke unwanted dependency and reliance on the assistive system.

3.4 Object and Action Recognition

In order to understand the current (action) environment and to provide adequate feedback, the ADAMAAS system needs to be able to recognize robustly the scene objects as well as users' actions using computer vision and image processing algorithms. In the ADAMAAS system, objects and hand-object contacts are detected in real-time from the video provided by the scene camera. Therefore in a first step we use a combination of segmentation and classification from RGBD camera data [12]. The segmentation is done by either Euclidean or feature-graph segmentation depending on the scene complexity. The segmentation results are classified, and a smoothing operation followed by automatic annotation with a label, bounding box and classification score is applied. Combining bottom-up segmentation with top-down classification, the algorithm is able to improve the segmentation results in complex scenes by autonomously finding the correct grouping level. The algorithm does not only allow for a simple interactive relearning of objects, but also an extension of the classification structure with unknown objects. Using the hand tracking feature of a Kinect we can detect if people establish or lose hand-object contact by checking if the detected hand coordinates falls within the objects' bounding box [13]. Furthermore, the approach allows for a task-dependent focus on the grouping hierarchy in order to identify task-relevant object parts (e.g. a mug composed from the handle and its inner or outer surfaces). As part of the project, we will substitute the Kinect part with a solution which is suitable for the intended mobile usage scenarios of the ADAMAAS system.

3.5 Augmented Reality (AR) Component

The AR component allows adaptive feedback prompting in textual, visual or virtual agent-based format on the transparent virtual plane in users' fields of view to be displayed (Fig. 1). This includes the proper adjustment and placing of the visuals when users move their eyes, head and body. For the rendering, the Unity3D software framework is used. An audio channel is used as an additional feedback modality. A marker-based approach to anchor the glasses in the real world scenario will allow for a proper adjustment of the AR feedback under natural head movements. A Wizard-of-Oz interface was implemented which allows the experimenter to show different content on the transparent plane. To gather preliminary user feedback on the general system concept we used static images as a first step (Fig. 3) which are not yet embedded in the scene and did not adjust to participants' head movements.

4. TEST SCENARIOS AND USABILITY ENGINEERING

In general, the ADAMAAS system shall be able to efficiently support arbitrary tasks and activities of bounded complexity. However, the research project is primarily focused on the requirements of elderly and cognitively impaired users. Therefore, two scenarios have been picked for evaluation: Assembly of a birdhouse from wooden pieces in a workshop for the handicapped, and usage of a modern

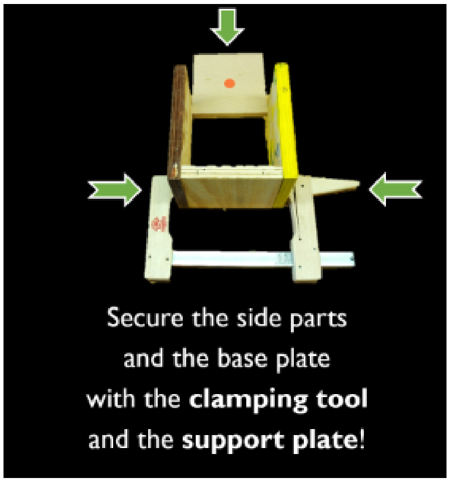

Figure 3: Action assistance for assembling a birdhouse displayed via AR (black background = transparent)

washing machine in a residential complex for the elderly. Apart from these, baking and cooking, and industrial assembly have also been taken into consideration as prospective test cases.

Our human-centered system design approach follows ISO 9241-210 [14], i.e. potential end users from the target groups are continually involved in the development. We determined the appropriate usability methods using a semi-automated approach based on ISO/TR 16982 [15] which implements propositions from [16]. Analyses of the contexts of use in the workshop and the retirement home, as well as task analyses were conducted in order to identify feasible sequences of actions (cp. section 3.3).

Furthermore, we collected requirements from relevant system stakeholder groups and derived a set of contingent system features, i.e. functions and properties. Potential users and other stakeholders were then asked to rate the importance of these features. The ratings were combined with demographic data and the respondents' roles. This data allows for an algorithmic inference of characteristic properties for the definition of appropriate stakeholder and user models, e.g. personas (cp. [17,18]). These will subsequently guide the prioritization of system features during development. Thereby, the role of the "system owner" or "on-site customer" known from agile development methodologies like SCRUM [19] or Extreme Programming [20] is replaced by a surrogate more focused on the needs of the system's potential end users and other stakeholders.

In a first evaluation study handicapped and elderly people wore the glasses and were shown exemplary assistive hints superimposed on the transparent plane in user's field of view (Fig. 3). Using a Wi-Fi network, the glasses were connected to a notebook so that the experimenter could change the displayed hint depending on the current activity. Participants wore the glasses during the whole process which lasted approximately 15 minutes. Preliminary qualitative results regarding the physical comfort of wearing the glasses, user feedback concerning the suitability of graphical and textual action representations superimposed via AR, as well as general acceptance, are promising. Participants were (physically) not substantially restricted by the glasses in the execution of their activities, gave positive feedback regarding the legibility and usefulness of the displayed information and could imagine wearing the glasses for even a longer period of time. However, some participants criticized the relatively high weight of the prototype glasses. Therefore, further technological progress is required to manufacture lighter and more comfortable ADAMAAS glasses in the future.

5. CONCLUSION

The ADAMAAS system is a new approach to providing mobile and adaptive cognitive action monitoring and assistance. ADAMAAS determines what the wearer is doing in order to automatically generate in real time context-appropriate individualized assistance in form of text or visuals. The goal is to allow users to continue working on a task while receiving instructions on how to do it better and correct mistakes as they happen. ADAMAAS is particularly targeted at elderly and disabled people, and will ultimately provide support for everyday basic tasks to allow for a more independent life despite reduced mental and/or physical capabilities.

The system is currently in a very early prototype stage. Future development will concentrate on embedding object and action recognition while eliminating dependence on depth data, real-time interpretation of eye movements with respect to cognitive processes and object classification, implementation of dynamic AR feedback, as well as further evaluation of our approach to human error prediction. We will also integrate new technological advances, e.g. more powerful AR rendering hardware to allow for a smoother and faster adaptation of the AR feedback prompting.

6. ACKNOWLEDGMENTS

This research is funded by the German Federal Ministry of Education and Research (BMBF) in the frame of the project "ADAMAAS" (Adaptive and Mobile Action Assistance in Daily Living Activities) for the BMBF call on adaptive and learning systems âĂŞ for an intuitive interaction between humans and complex technologies. The work was also supported by the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG).

7. REFERENCES

1 Pfeiffer, T. (2013). Gaze-based assistive technologies. In G. Kouroupetroglou (Ed.), Assistive Technologies and Computer Access for Motor Disabilities (pp. 90-109). Hershey:IGI Global.

2 Koesling, H., Zoellner, M., Sichelschmidt, M.; & Ritter, H. (2009). With a flick of the eye: Assessing gaze-controlled human-computer interaction. In: H. Ritter, G. Sagerer, R. Dillmann, and M. Buss (Eds.), Cognitive Systems Monographs, Human Centered Robot Systems: Cognition, Interaction, Technology (pp. 83-92). Berlin: Springer Verlag

3 Jacob, R.J.K. (1993). What you look at is what you get. Computer, 26(7), pp. 65-66.

4 Dybdal, M.L., San Agustin, J., & Hansen J.P. (2012). Gaze input for mobile devices by dwell and gestures. In: Proceedings of the symposium on eye tracking research and applications, ETRA 12. ACM, New York, pp 225-228.

5 Losing, V., Rottkamp, L., Zeunert, M., & Pfeiffer, T. (2014). Guiding Visual Search Tasks Using Gaze-Contingent Auditory Feedback. UbiComp'14 Adjunct: The 2014 ACM Conference on Ubiquitous Computing Adjunct Publication, pp. 1093-1102.

6 Renner, P. and Pfeiffer, T. (2015). Online Visual Attention Monitoring for Mobile Assistive Systems. In: Pfeiffer T, Essig K, (Eds.), SAGA 2015: 2nd International Workshop on Solutions for Automatic Gaze Data Analysis (pp. 14-15). eCollections Bielefeld University 2015.

7 Holmqvist, K. et al. (2011). Eye tracking - A comprehensive guide to methods and measures. New York: Oxford University Press.

8 Just, M.A. and Carpenter, P.A. (1987). The Psychology of Reading and Language. Newton: Allyn and Bacon.

9 Reinert, G. (1993). Augenbewegungen bei geistig behinderten Kindern. Lehrstuhlbericht. Arbeitsgruppe fÃijr Umwelt und Kognitionspsychologie, Ruhr-Universitaet Bochum, Germany.

10 Majaranta, P. et al. 2012. Gaze Interaction and Applications of Eye Tracking - Advances in Assistive Technologies. Hershey: IGI Global.

11 Schack, T. (2012). Measuring mental representations. In Measurement in sport and exercise psychology (pp. 203-214). Champaign, IL: Human Kinetics.

12 Ueckermann, A. et al. (2014). Real-Time Hierarchical Scene Segmentation and Classification. 14th IEEE-RAS International Conference on Humanoid Robots (Humanoids), November 18-20, 2014, Madrid, Spain.

13 Schroeder, M., Maycock, J., Ritter,H., & Botsch, M. (2014). Real-Time Hand Tracking using Synergistic Inverse Kinematics.IEEE International Conference on Robotics and Automation (ICRA), pp. 5447-5454.

14 ISO 9241-210 (2008). Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems.

15 ISO/TR 16982 (2002). Ergonomics of human-system interaction - Usability methods supporting human-centred design.

16 Fischer, H., Strenge, B., & Nebe, K. (2013). Towards a holistic tool for the selection and validation of usability method sets supporting human-centered design. In Design, User Experience, and Usability. Design Philosophy, Methods, and Tools (pp. 252-261). Springer Berlin Heidelberg.

17 Sinha, R. (2003). Persona development for information-rich domains. In CHI'03 extended abstracts on Human factors in computing systems (pp. 830-831). ACM.

18 Tu, N., Dong, X., Rau, P., & Zhang, T. (2010). Using cluster analysis in persona development. In International Conference on Supply Chain Management and Information Systems.

19 Schwaber, K., & Sutherland, J. (2013). The Scrum GuideâĂŞThe Definitive Guide to Scrum: The Rules of the Game. URL: http://www.scrumguides.org

20 Beck, K. (2000). Extreme programming explained: embrace change. Addison-Wesley.