KIBITZER: A Wearable System for Eye-Gaze-based Mobile Urban Exploration

Matthias Baldauf
FTW
Donau-City-Strasse 1
1220 Vienna, Austria
+43-1-5052830-47
baldauf@ftw.at

Peter Fröhlich
FTW
Donau-City-Strasse 1
1220 Vienna, Austria
+43-1-5052830-85
froehlich@ftw.at

Siegfried Hutter
FTW
Donau-City-Strasse 1
1220 Vienna, Austria
+43-1-5052830-10
shutter@ftw.at

ABSTRACT

Due to the vast amount of available georeferenced information novel techniques to more intuitively and efficiently interact with such content are increasingly required. In this paper, we introduce *KIBITZER*, a lightweight wearable system that enables the browsing of urban surroundings for annotated digital information. *KIBITZER* exploits its user's eye-gaze as natural indicator of attention to identify objects-of-interest and offers speech- and non-speech auditory feedback. Thus, it provides the user with a 6th sense for digital georeferenced information. We present a description of our system's architecture and the interaction technique and outline experiences from first functional trials.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *Interaction styles*

General Terms

Experimentation, Human Factors

Keywords

Wearable Computing, Mobile Spatial Interaction, Eye-gaze

1. INTRODUCTION

Computers are becoming a pervasive part of our everyday life, and they increasingly provide us with information about the ambient environment. Smartphones guide us through unfamiliar areas, revealing information about the surroundings, and helping us share media with others about certain places. While such location-based information has traditionally been accessed with a very limited set of input devices, usually just a keyboard and audio, multimodal interaction paradigms are now emerging that take better advantage of the user's interactions with space.

The research field of Mobile Spatial Interaction [7] breaks with the conventional paradigm of displaying nearby points-of-interest (POIs) icons on 2D maps. Instead, MSI research aims to develop new forms of sensing the users' bodily position in space and to envision new interactions with the surrounding world through gesture and movement.

Key interaction metaphors for recently implemented MSI systems are the 'magic wand' (literally pointing the handheld at objects of interest to access further information [16]**),** the 'smart lens' (superimposing digital content directly on top of the recognized real-world object [14]), or the 'virtual peephole' (virtual views aligned to the current physical background, e.g. displaying a "window to the past" [15]). These interaction techniques have been successfully evaluated in empirical field studies, and first applications on mass market handhelds have attracted much end-user demand on the market [20].

Metaphors like the magic wand, smart lens and virtual peephole incorporate the handheld as the gateway for selecting physical objects in the user's surroundings and to view related digital content. However, it may not always be preferable to put all attention on the mobile device, such as when walking along a crowded pedestrian walkway or when having both hands busy.

We present *KIBITZER*, a gaze-directed MSI system that enables the hands-free exploration of the environment. The system enables users to select nearby spatial objects 'with the blink of an eye' and presents related information by using speech- and non-speech auditory output.

In the following section, we provide an overview of relevant previous work in the area of eye-based interaction. We then describe *KIBITZER*'s system architecture and the realized user interaction technique, and we demonstrate its operation with photos and screenshots. We conclude with experiences from first functional trials and plans for further research.

2. EYE-BASED APPLICATIONS

Applications making use of eye movement or eye gaze patterns can be broadly categorized as diagnostic and interactive applications [4].

In diagnostic use cases, an observant's eye movements are captured to assess her attentional processes over a given stimulus. User tests including eye tracking techniques are a popular method in HCI e.g. to evaluate the usability of user interfaces and improve product design. In the 1940s, pioneering work in this field was done by Fitts et al [6] who were the first ones to use cameras to capture and analyze observants' eye movements. They

Augmented Human Conference, April 2–3, 2010, Megève, France.

collected data from pilots during landing to propose a more efficient arrangement and design of the cockpit instruments.

The investigation of interactive eye tracking applications started in the 1980s. In contrast to diagnostic systems where the recorded data mostly is evaluated after the test, eye movements in interactive scenarios are exploited as input parameters for a computer interface in real-time. Bolt [2] first introduced the idea of eye movements acting as a complementary input possibility for computers. First studies investigating the usage of eye movements for common desktop computer tasks were presented by Ware [19] and Jacob [9]. Ware identified eye input as fast technique for object selection; Jacob found it to be effective for additional tasks such as moving windows. Recently, eye gaze as natural input has been revisited in the context of so-called attentive user interfaces [18] that generally consider intuitive user reactions such as gestures and body postures to facilitate human-computer-interaction. Commercial interactive eye tracking applications are currently focused on military use cases and tools for people with severe physical disabilities [11]. A comprehensive survey of eye tracking applications can be found in [4].

Mobile applications in the field of Augmented or Virtual Reality that allow the visual exploration of real or virtual worlds are mainly restricted to the awareness of head movement. E.g. Feiner et al [5] and, more recently, e.g. Kooper et al [10] and Reitmayr et al [13] presented wearable systems that support object selection by gaze only estimated from head pose. Bringing the object-of-interest within the center of a head-worn see-through display would select or trigger another specified action. The exploitation of a user's detailed eye gaze through suitable trackers is a rarely considered aspect in Augmented Reality. Very recent examples include the integration of eye-tracking in a stationary AR videoconferencing system [1] and a wearable AR system combining a head-mounted display and an eye-tracker to virtually interact with a real-world art gallery [12].

Dependent on the underlying technology, two basic types of eye tracking systems can be distinguished. Systems based on so-called electro-oculography exploit the electrostatic field around a human's eyes. Electrodes placed next to the observant's eyes measure this field's changes as the eyes move. As the eye's position can only be estimated using this technique, electro-oculography is rather applied for activity recognition than for gaze detection [3].

In contrast, video-based eye tracking approaches make use of one or several cameras recording one or both eyes of a user. By analyzing reflections in the captured eye using computer vision techniques the eye's position and thus, the eye gaze can be determined. The cameras can either be placed near the object-of-interest (usually a computer display) to remotely record the user's head in a non-intrusive way or mounted on a headpiece worn by the user. Though this approach's disadvantage of intrusiveness, head-mounted eye trackers offer a higher accuracy than non-intrusive systems and can also be applied in mobile use cases [11] such as our exploration scenario.

3. SYSTEM SETUP

This section introduces the technical setup for the realization of our *KIBITZER* system. First, we present the used hardware, and then we explain the involved software components and their functionality.

3.1 Mobile Equipment

The core hardware component of our setup is an *iView X HED* system, a latest generation mobile eye tracker from Sensomotoric Instruments GmbH (SMI). It includes two cameras to record both an eye's movement and the current scene from the user's perspective. For best possible stability the equipment is mounted at a bicycle helmet (Figure 1). Via USB the tracker is connected to a laptop computer (worn in a backpack) where the video stream is processed and the current eye-gaze is calculated.

To augment a user's relative gaze direction with her global position as well as her head's orientation and tilt, we use a G1 phone powered by Android (Figure 2). This smartphone contains all necessary sensors such as a built-in GPS receiver, a compass and accelerometers. With a custom-made fixation the G1 device is mounted on top of the bicycle helmet (Figure 3).

Figure 1. iView X HED, a mobile eye-gaze-tracker [17]

Figure 2. G1 phone powered by Android

Figure 3. The combined head-mounted device worn during a functional test.

3.2 Software Components

Figure 4 gives an overview of the software components of our system architecture and their communication.

The aforementioned eye tracker system comes with a video analyzer application that is installed on the laptop computer. This *iView X HED* application offers a socket-based API interface via Ethernet to inform other applications about calculated eye data. For our scenario, we implemented a small component that connects to this interface and forwards the fetched gaze position in pixels (with regard to the scene camera's picture) via Bluetooth.

Our *mobile application* installed on the attached smartphone receives the gaze position and (based on a prior calibration) is able to convert these pixel values in corresponding horizontal and vertical deviations in degrees with regard to a straightforward gaze. These values are continuously written to a local log file together with the current location, the head's orientation and tilt. Adding the horizontal gaze deviation to the head's orientation and, respectively, adding the vertical gaze deviation to the head's tilt results in the global eye gaze vector. To provide auditory feedback the mobile application makes use of an integrated *text-to-speech engine.*

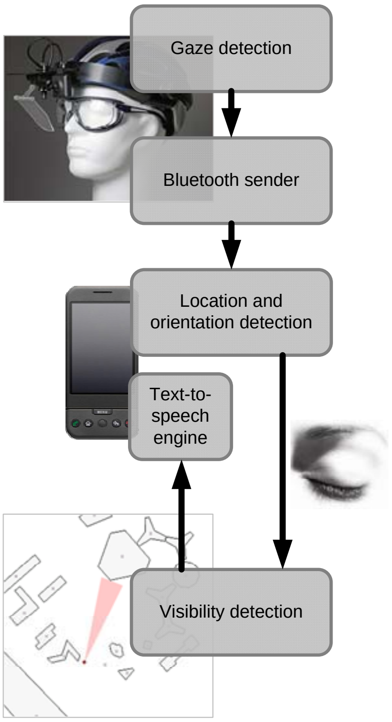

Figure 4. Overview of our system's software components and their communication.

The mobile application may invoke a remote *visibility detection service* via a 3G network. This service takes the user's current view into account: By passing a location and an orientation (in our case the global eye gaze vector) to this HTTP service, a list of currently visible POIs in this direction is returned. The engine makes use of a 2.5D block model, i.e. each building in the model is represented by a two-dimensional footprint polygon, which is extruded by a height value. Based on this model, POIs with a clear line-of-sight to the user and POIs located inside visible buildings can be determined. The resulting list contains the matching POIs' names and locations as well as the relative angles and distances with regard to the passed user position and orientation. More details about the used visibility detection engine can be found in [16].

3.3 Initial Calibration

Before the equipment can be used it needs to be calibrated. This procedure starts with a *nine point calibration* for the eye-tracker using the *iViewX HED* application. For this purpose, nine markers must be arranged at a nearby wall in a 3x3 grid, whereas the marker in the center should be placed at the user's eye height. The calibration points then can be setup in the application via the delivered scene video and mapped point by point to the corresponding gaze direction. The standard procedure is extended with a *custom calibration* to later map gaze positions in pixels to gaze deviations in degrees. By turning and tilting the head towards the calibration points while now keeping eye gaze straight forward, conversion factors for horizontal and vertical gaze positions can be determined based on the fetched compass and accelerometer data.

4. USER INTERACTION

After the calibration process, the presented equipment is ready for outdoor usage. Previously mentioned, our goal is the gaze-sensitive exploration of an urban environment providing the user with a 6th sense for georeferenced information.

When to trigger which suitable action in an eye-gaze-based system is a commonly investigated and discussed issue known as the 'Midas Touch' problem. A good solution must not render void the intuitive interaction approach of such an attentive interface by increasing the user's cognitive load or disturbing her gaze-pattern. At the same time, the unintended invocation of an action must be avoided.

The task of object selection on a computer screen investigated by Jacob [9] might seem related to our scenario of mobile urban exploration where we want to select real-world objects to learn more about annotated POIs. Jacob suggests either to use a keyboard to explicitly execute the selection of a viewed item via a key press or, preferably, apply a dwell time to detect a focused gaze and fire the action thereafter. In Jacob's experiment, users were provided with visual feedback about the current selection and therefore, were able to easily correct errors.

Due to our mobile scenario, we want to keep the involved equipment as lightweight as possible sparing an additional keyboard or screen. Therefore, we rely on an explicit eye-based action to trigger a query for the currently object. As though the user would memorize the desired object, closing her eyes for two seconds triggers the selection. In technical terms, the spatial query is executed for the last known global gaze direction if the user's tracked eye could not be detected during the last two seconds. An invocation of the query engine is marked in the log file with a special status flag.

The names of the POIs returned by the visibility detection service are then extracted and fed into the text-to-speech engine for voice output. If a new query is triggered during the output, the text-to-speech engine is interrupted and restarted with the new results. The auditory output is either possible via the mobile's built-in loudspeakers or attached earphones.

5. TOUR ANALYSIS

During the usage of our *KIBITZER* all sensor values are continuously recorded to a log file. These datasets annotated with corresponding time stamps enable a complete reconstruction of the user’s tour for later analysis.

To efficiently visualize a log file’s content we implemented a converter tool that generates a KML file from a passed log file. KML is a XML-based format for geographic annotations and visualizations with the support of animations. The resulting tour video can be played using Google Earth [8] and shows the user’s orientation and gaze from an exocentric (‘third person’) perspective (Figure 5). The displayed human model is orientated according to the captured compass values; its gaze ray is corrected by the calculated gaze deviations. The invocation of the visibility detection service, i.e. the gaze-based selection of an object, is marked by a different-colored gaze ray.

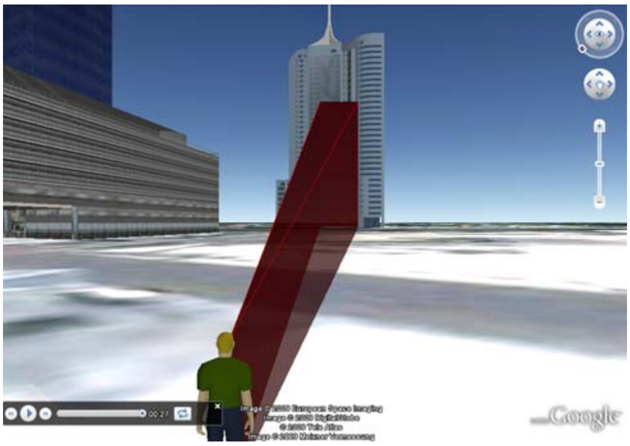

Figure 5. Screenshot of a KML animation reconstructed from the logged tour data.

Figure 6. Screenshot of the video taken by the helmet-mounted scene camera. The red cross represents the current eye gaze.

As the scene camera’s video stream can be recorded via the *iViewX HED* application, the reconstructed tour animation can be compared to the actually captured scene video (Figure 6). The scene video is overlaid with a red cross representing the user’s current gaze and thus, can be used to evaluate our system’s accuracy. Furthermore, when combined with the visibility detection engine, the tour reconstruction can be used to automatically identify areas of interest or compile further statistics.

6. CONCLUSIONS AND OUTLOOK

In this paper, we introduced *KIBITZER*, a wearable gaze-sensitive system for the exploration of urban surroundings, and presented related work in the field of eye-based applications. Wearing our proposed headpiece, the user’s eye-gaze is analyzed to implicitly scan her visible surroundings for georeferenced digital information. Offering speech-auditory feedback via loudspeakers or earphones, the user is unobtrusively informed about POIs in their current gaze direction. Additionally, we offer tools to reconstruct a user’s recorded tour visualizing her eye-gaze. These animations are not only useful for accuracy tests during development but rather aim at a later automated tour analysis, e.g. to identify areas of interest.

Experiences from first functional tests and reconstructed tour videos showed that the proposed system’s overall accuracy is sufficient for determining POIs in the user’s gaze. However, in some trials the built-in compass was heavily influenced by magnetic fields resulting in wrong POI selections. This problem could be solved by complementing the system with a more robust external compass.

During these tests we observed some minor limitations of the chosen vision-based gaze tracking approach and the blinking interaction. In rare cases, unfavorable reflections caused by direct sunlight prevented a correct detection of the user’s pupil and therefore, interfered the gaze tracking. Obviously, at night the usage of such a vision-based system is not feasible without any artificial light source.

Our proposed research prototype is a first step towards the exploitation of a user’s eye-gaze in mobile urban exploration scenarios and therefore, it is deliberately designed for experimentation. The current system built from off-the-shelf hardware components provides a complete framework to study possible gaze-based interaction techniques. With the future arrival of smart glasses or even intelligent contact lenses, the required equipment is supposed to become more comfortable to wear, if not almost unnoticeable.

Applying the presented system, we will evaluate the usability and effectiveness of eye-gaze-based mobile urban exploration in upcoming user tests. We will set special focus on the acceptance of the currently implemented ‘blinking’ action and the investigation of alternative interaction techniques, respectively. Inspired by ‘mouse-over’ events known from Web sites such as switching an image when moving the mouse cursor over a sensitive area, implicit gaze feedback is conceivable. When a user glances at an object, she might be notified about the availability of annotated digital information by a beep or tactile feedback. The combination of our gaze-based system with a brain-computer-interface to estimate a gaze’s intention and thus, trigger an according action is another promising direction for future research.

7. ACKNOWLEDGMENTS

This work has been carried out within the projects WikiVienna and U0, which are financed in parts by Vienna's WWTF funding program, by the Austrian Government and by the City of Vienna within the competence center program COMET.

8. REFERENCES

[1] Barakonyi, I., Prendinger, H., Schmalstieg, D., and Ishizuka, M. 2007. Cascading Hand and Eye Movement for Augmented Reality Videoconferencing. In Proc. of 3D User Interfaces, 71-78.

[2] Bolt, R.A. 1982. Eyes at the Interface. In Proc. of Human Factors in Computer Systems Conference, 360-362.

[3] Bulling, A., Ward, J.A., Gellersen, H., and Tröster, G. 2009. Eye Movement Analysis for Activity Recognition. In Proc. of the 11th International Conference on Ubiquitous Computing, 41-50.

[4] Duchowski, A.T. 2002. A breadth-first survey of eye tracking applications. In Behavior Research Methods, Instruments, & Computers (BRMIC), 34(4), 455-470.

[5] Feiner, S., MacIntyre, B., Höllerer T., and Webster, A. 1997. A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment. In Personal and Ubiquitous Computing, Vol. 1, No. 4, 208-217.

[6] Fitts, P. M., Jones, R. E., and Milton, J. L. 1950. Eye movements of aircraft pilots during instrument-landing approaches. In Aeronautical Engineering Review 9(2), 24–29.

[7] Fröhlich, P., Simon, R., and Baillie, L. 2009. Mobile Spatial Interaction. Personal and Ubiquitous Computing, Vol. 13, No. 4, 251-253.

[8] Google Earth. http://earth.google.com. Accessed January 7 2010.

[9] Jacob, R.J.K. 1990. What you look at is what you get: eye movement-based interaction techniques. In Proc. of the SIGCHI conference on Human factors in computing systems, 11-18.

[10] Kooper, R., and MacIntyre, B. 2003. Browsing the Real-World Wide Web: Maintaining Awareness of Virtual Information in an AR Information Space. In International Journal of Human-Computer Interaction, Vol. 16, No. 3, 425-446.

[11] Morimoto, C.H., and Mimica, M.R.M. 2005. Eye gaze tracking techniques for interactive applications. In Computer Vision and Image Understanding, Vol. 98, No. 1, 4-24.

[12] Park, H.M., Lee, S.H., and Choi, J.S. 2008. Wearable Augmented Reality System using Gaze Interaction. In Proc. of the 7th IEEE/ACM international Symposium on Mixed and Augmented Reality, 175-176.

[13] Reitmayr, G., and D. Schmalstieg, D. 2004. Collaborative Augmented Reality for Outdoor Navigation and Information Browsing. In Proc. of Symposium on Location Based Services and TeleCartography, 31-41.

[14] Schmalstieg, D., and Wagner, D. 2007. The World as a User Interface: Augmented Reality for Ubiquitous Computing. In Proc. of Symposium on Location Based Services and TeleCartography, 369-391.

[15] Simon, R. 2006. The Creative Histories Mobile Explorer - Implementing a 3D Multimedia Tourist Guide for Mass-Market Mobile Phones. In Proc. of EVA.

[16] Simon, R., and Fröhlich, P. 2007. A Mobile Application Framework for the Geo-spatial Web. In Proc. of the 16th International World Wide Web Conference, 381-390.

[17] SMI iView X™ HED. http://www.smivision.com/en/eye-gaze-tracking-systems/products/iview-x-hed.html. Accessed January 07 2010.

[18] Vertegaal, R. 2002. Designing Attentive Interfaces. In Proc. of the 2002 Symposium on Eye Tracking Research & Applications, 23-30.

[19] Ware, C., and Mikaelian, H.T. 1987. An evaluation of an eye tracker as a device for computer input. In Proc. of the ACM CHI + GI-87 Human Factors in Computing Systems Conference, 183-188.

[20] Wikitude. http://www.mobilizy.com/wikitude.php. Accessed January 07 2010.