Towards a Mobile Eye-Gaze-based Exploration of Urban Environments

Matthias Baldauf, Peter Fröhlich, and Siegfried Hutter

FTW - Telecommunications Research Center Vienna
Donau-City-Strasse 1, 1220 Vienna, Austria
{baldauf, froehlich, shutter}@ftw.at

Abstract. As the amount of available georeferenced information rapidly increases, novel intuitive techniques to interact with such content on the move are needed. In this paper, we introduce our efforts towards a lightweight wearable system that enables the browsing of urban surroundings for annotated digital information. The proposed system exploits its user's eye-gaze as natural indicator of attention to identify objects-of-interest and provides speech- and non-speech auditory feedback. We present a description of our system's architecture and the interaction technique. In future, this framework will help us to evaluate eye-gaze-based techniques for mobile urban exploration.

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 (MSI) [1] 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.

We present a gaze-directed MSI system that allows 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. Whereas the usage of eye movements for common desktop computer tasks was investigated in early work, e.g. [2], it is only rarely considered for interaction with real-world objects. One very recent example is a wearable Augmented Reality system combining a head-mounted display and an eye-tracker to virtually interact with an art gallery [3].

In the following sections, we describe our system architecture and the realized user interaction technique. We conclude with experiences from first functional trials and outline our further research.

Fig. 1. The used eye tracker includes an eye camera and a scene camera.

Fig. 2. A smartphone with all needed sensors is mounted at the helmet.

2 System Architecture

The core hardware component of our setup is an iView X HED system, a latest generation mobile eye tracker from Sensomotoric Instruments GmbH. It includes two cameras to record both an eye's movement and the current scene from the user's perspective (Figure 1). For best possible stability the equipment is mounted at a bicycle helmet. 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 smartphone. This device 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 2).

The aforementioned eye tracker system comes with a video analyzer application that is installed on the laptop computer. This 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 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. A custom tool allows the reconstruction from the logged tour data for visualization in Google Earth (Figure 3) and comparison with the recorded scene camera video (Figure 4).

Additionally, the mobile application may invoke a remote visibility detection service via the 3G network. This service takes the user's current view into account: By passing a location and an orientation 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. More details about the used visibility detection engine can be found in [4].

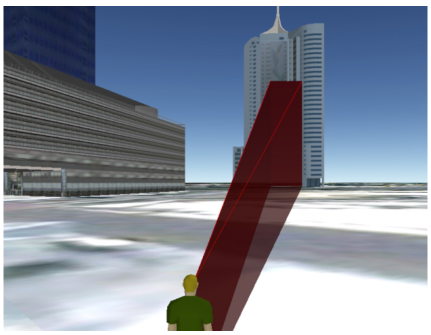

Fig. 3. Screenshot of a KML animation reconstructed from the logged tour data.

Fig. 4. Screenshot of the scene video with eye-gaze overlay.

3 User Interaction

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 [2] 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 display. 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. 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. Wearing one earphone, the user is unobtrusively informed about the spotted object. Inspired by 'mouse-over' effects known from Web pages, we plan to investigate non-speech auditory feedback such as beeps to notify about the availability of content while gazing.

4 Conclusions and Outlook

We introduced our ongoing work towards a lightweight gaze-sensitive system for the exploration of urban surroundings. 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.

Enhancing and applying the presented system, we will evaluate the usability and effectiveness of eye-gaze-based mobile urban exploration in future user tests. We will set special focus on the acceptance of the currently implemented 'blinking' action and the investigation of alternative interaction techniques, respectively.

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.

References

1. Peter Fröhlich, Rainer Simon, and Lynne Baillie. Mobile spatial interaction. *Personal and Ubiquitous Computing*, 13(4):251–253, 2009.
2. Robert J. K. Jacob. The use of eye movements in human-computer interaction techniques: What you look at is what you get. *ACM Trans. Inf. Syst.*, 9(2):152–169, 1991.
3. Park Hyung Min, Seok Han Lee, and Jong Soo Choi. Wearable augmented reality system using gaze interaction. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 175–176, Washington, DC, USA, 2008. IEEE Computer Society.
4. Rainer Simon and Peter Fröhlich. A mobile application framework for the geospatial web. In *Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 8-12, 2007*, pages 381–390, 2007.