MARA – A Mobile Augmented Reality-Based Virtual Assistant

Andreas Schmeil*, Wolfgang Broll†

Collaborative Virtual and Augmented Environments Department

Fraunhofer FIT, Sankt Augustin, Germany

ABSTRACT

High-end mobile devices are becoming increasing popular in every day life. Augmented Reality (AR) builds on this trend by combining mobile computing with connectivity and location-awareness. In doing so, AR can provide a very rich user experience. In this paper we discuss the approach and development of an AR-based personal assistant, combining the familiar interface of a human person with the functionality of a location-aware digital information system. The paper discusses the main components of the system, including the anthropomorphic user interface as well as the results of an initial prototype evaluation.

CR Categories and Subject Descriptors: D.2.14.a [Software Engineering]: Human Factors in Software Design—User Interfaces, H.1.2.b [Models and Principles]: User/Machine Systems—Human-centered computing, H.2.8.o [Database Management]: Database Applications—Spatial databases and GIS, H.5.1.b [Information Interfaces and Representation (HCI)]: Multimedia Information Systems—Artificial, augmented and virtual realities, H.5.2 [Information Interfaces and Representation (HCI)]: User Interfaces—Design for wearability, Evaluation/methodology, Graphical User Interfaces, Voice I/O, J.9 [Mobile Applications]: Location-dependent and sensitive, pervasive computing.

Additional Keywords: Augmented Reality, Mobile Computing, Virtual Humans, Anthropomorphic User Interfaces, Digital Assistants, Environment Model, Location Based Systems

1 INTRODUCTION

Today people make increasing use of electronic equipment. While mobile phones are already widespread, PDAs and laptop computers now replace more traditional media and tools. Smartphones combine PIM (Personal Information Management) functions of PDAs and the communication functionality of mobile phones. With providers having started cell based services on the one hand and more and more devices supporting GPS on the other hand, we can further observe an increasing availability of location-based services. The popularity of such services is also increasing with new connectivity technologies such as WiFi, WiMax and UMTS. However, the available hardware, protocols and services often lack some important features, such as:

- Many tasks are cumbersome and time-consuming, due to the limited user interfaces provided by the devices, especially compared to having a real assistant or secretary.
- The devices and displays used in them often place limits on what the users can undertake at any given time, resulting in a limited range of applications and services.
- Existing systems are difficult to use if you need your hands free. Everyone knows the situation where one's mobile phone signals some event while both hands are in use.
- Existing electronic organizer functionality is not integrated with location information. The users mostly are not made aware of the link and hence do not benefit from the information, even when it is available.

The work presented in this paper aims to overcome these difficulties by using an alternative user interface: a virtual human. In theory, interacting with such an anthropomorphic user interface should be similar to communicating with a real assistant. To strengthen this experience they have to be part of the real world. Mobile Augmented Reality (AR) is a technology that enables us to use anthropomorphic virtual assistants as part of the user's local environment almost anywhere and anytime (see figure 1).

Figure 1. The virtual anthropomorphic assistant in a mobile AR environment

In this paper we describe the development of an intuitive multi-modal user interface that combines mobile AR technology with a personal assistant. The second part provides a brief overview of related work. In section 3 we present our approach in detail. Section 4 discusses the results of our initial field trials and section 5 provides conclusions and further directions.

2 RELATED WORK

Our system presents a first step to integrate an anthropomorphic assistant with an AR information and navigation system. Related work so far concentrates on one of these technologies.

Mobile augmented reality systems such as the Touring Machine [6] and the applications developed within the OCAR project [11] provide information about the environment to a mobile user by superimposing text, pictures and other data onto the user's view. The user interfaces are mostly based on two-dimensional GUIs operated by touchpads, which according to our experience are often considered as somewhat awkward and rather

*e-mail: andreas.schmeil@googlemail.com

†e-mail: wolfgang.broll@fit.fraunhofer.de

difficult to use. Furthermore, reading text in the display requires the user's complete attention.

Virtual characters were used in augmented reality systems for digital storytelling, to play AR film sequences [13] or guide children to outdoor locations [12]. When a user requires additional information and further interactivity the situation becomes more complex. Virtual assistants have mainly been used in virtual reality environments so far. The research mostly focuses on how to interact with the characters rather than the functions they provide. Thus, virtual assistants are in most cases restricted to one specific task. Hamilton [14] was designed to arrange furniture in a virtual environment, MAX [9] demonstrates how to assemble a construction kit. Most assistants understand voice commands and interpret hand gestures. Welbo [1] was the first anthropomorphic assistant in AR and took the form of a little robot. He was able to point at and rearrange virtual objects in a living room upon voice commands. Other virtual characters in AR function as game characters [3], test subjects [2] or behavior-driven presenters [4]. However, an interactive anthropomorphic assistant that fulfils tasks for everyday life, as our assistant does, has not been developed yet.

Since most AR systems work with manually entered data sets, only little research has been done in the field of databases for mobile AR applications. The Nexus project [7] for example explores the creation, definition and development of environment models from heterogenic data sources. A similar approach was pursued within the OCAR project [11]. The problem with many of the environment models created so far is that they only contain geometric and physical data, and semantic information has to be added manually. In contrast, as our work is focuses on the development of an anthropomorphic assistant there was need for an environment model containing information about the objects and how they can be presented.

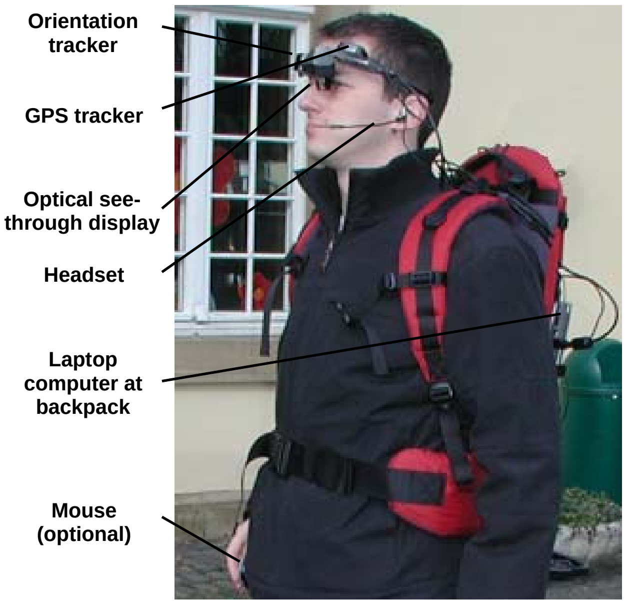

Figure 2. Technical overview of the system hardware

3 Our Approach

Our approach includes the following major components, which will be presented in detail: the hardware setup, the virtual assistant, our developed environment model, the software system, and the multimodal user interface.

3.1 Mobile AR System

Our mobile AR setup uses a notebook computer with a 2 GHz processor and an ATI Mobility FireGL V3100 graphics accelerator operating under Windows XP. The user's position is tracked by a Holux GPSlim236 GPS receiver in outdoor settings and by ARToolKit [15] tracking in indoor settings. For accessing the tracking devices and for combining the provided data into stable 6-DOF tracking information we use the MORGAN AR/VR framework [10], which also provides support for the seamless transition between outdoor and indoor tracking. The user can thus roam freely between outdoor and indoor environments. The notebook computer and the GPS receiver are mounted on a backpack worn by the user.

The virtual assistant is superimposed onto the user's view by using a Shimadzu DataGlass 2/A display. We decided for a monoscopic optical see-through head-worn display for safety reasons (especially when using the system outdoors). The InertiaCube3 and an external active GPS antenna are mounted to the display. For speech input and output the user wears a headset (a one-ear mono headset, to reduce magnetic interferences). The alternative input device, a common PC wheel-mouse, is either carried in one hand or in a (jacket) pocket.

Unlike most other mobile AR configurations, our setup works perfectly without a helmet (see figure 2). Components were chosen that are light-weight and small in size. This layout is a prototype and we expect to make some modifications over time so that it is more suitable for daily use.

Figure 3. The virtual assistant with the calendar

3.2 Virtual Character

In order to render the anthropomorphic character we extended the Morgan framework by the open source character animation library Cal3D [16]. As the anthropomorphic assistant uses different objects for different tasks it was necessary to develop a system to support dynamic hiding and showing of virtual geometries. The system enables the assistant to take out objects like a calendar (for scheduling appointments, as shown in figure 3), a notebook (for writing down notes) and a pen. Further, all animation and show/hide functions can be scheduled, facilitating the creation of animation sequences.

To bear a human impression, an anthropomorphic assistant also requires a natural voice. We extended our framework with a component for speech synthesis and output based on Loquendo TTS. Equipped with a voice corresponding to the character's appearance that is synchronized with the character's gestures, it goes from being just a moving body to becoming a virtual person.

Other than basic gestures like nodding, shaking the head or pointing in a specific direction, the assistant is equipped with elementary movements like walking or running. The assistant is able to move naturally at any desired speed. Furthermore, we created waypoint-specific and object-specific animations that are executed (synchronized with speech output) at the presentation of waypoints, or objects, respectively. However, to ensure flexibility, the synchronization of presentations cannot be controlled by the application. Rather, it has to be specified in a modifiable database where all the other information about the world is defined as well.

3.3 World Knowledge

For the guidance and navigation functions of the virtual assistant, extensive information about the real world is required. Therefore, we developed a lightweight, yet powerful environment model, based on the widespread format GPX [17]. This XML schema defines waypoints, routes and tracks. We extended the waypoints to utilize them as local environment models. A local environment is defined as a limited area that can contain objects and has its own Cartesian coordinate system. The sum of all specified local models, which are transformed unambiguously, defines a global environment model, as illustrated in figure 4.

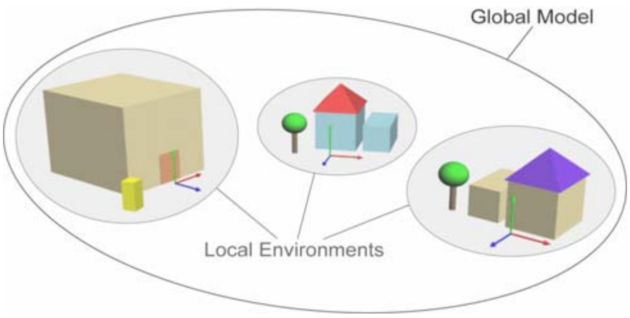

Figure 4. Local environments in the global model

A so-called augmented waypoint is composed of a name, a location in geodetic coordinates, an elevation, a description, a comment, a type, some GPS-related data and the following additional information:

- the orientation of its local coordinate system
- the extent of each contained object (a VRML scene)
- a name, a comment and a description of each object
- position and orientation of ARToolKit markers
- specific information for the assistant: animations to be executed, positions for presentations and labels in the text for the synchronization of speech and gestures

The assistant presents an object by reading out aloud its description. Animations are triggered by the speech synthesis. We decided to develop a flexible and easy-to-enlarge environment model, so it relies on multiple VRML scenes rather than one large 3D model. It is based on textual descriptions within XML data – so no complex object hierarchy as in existing models is needed.

3.4 The Software System

Figure 5 gives an overview of the components running within the MORGAN framework. The MARA application controls the assistant, manages the user's personal information and handles all aspects of interaction. The MORGAN Viewer's purpose is to render the scene graph with the virtual character and its objects of utility as the only visible geometry.

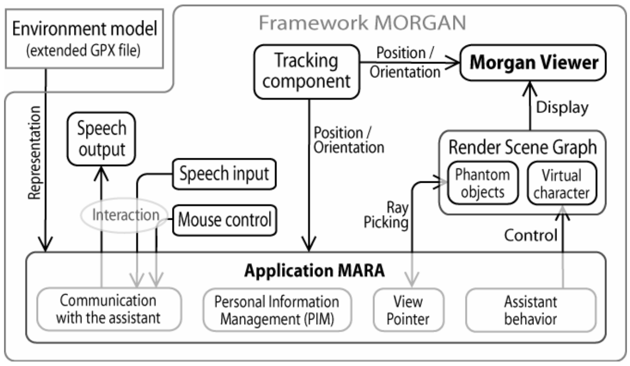

Figure 5. The modular system

Continuously updated tracking data, provided by the tracking component, is used to set the viewpoint for rendering, as well as to keep the assistant informed about the user's whereabouts. The assistant is thus able to walk towards the user while talking to him, to follow him and to turn to him while presenting an object. The data further provides the basis for the gaze-based selection mechanism.

For the personal information management (PIM) functionality of the system, an internal representation of a calendar (a set of appointments) and an ordered list of nodes are kept. Appointments can be associated with locations, i.e. with waypoints. For the navigation a route is planned using paths defined in the environment model.

We realized the assistant's behavior as a 2-level hierarchical state machine. The higher states represent the implemented functions (Idling, Calendar, Notebook, Introduction, Presentation and Navigation), the lower states correspond to several steps of the particular functions. The assistant gives feedback on the current states with distinct gestures and accompanying utterances. State transitions are mostly caused by user actions, but also by expiration of time intervals or the end of ongoing actions of the assistant.

3.5 Multimodal Interaction

The system control comprises of two distinct types of interaction: the communication with the assistant and the selection of (real world) objects.

We decided for speech input as the main user interface for communication, for it comes closest to communicating with a real person and thus promises a high user acceptance, allowing intuitive conversations. Our implementation uses IBM's ViaVoice speech recognition software, utilizing multiple grammars to increase recognition rates. The system switches automatically between grammars at high-level state transitions.

For situations that do not allow using speech input, we developed a secondary user interface, utilizing a common PC wheel-mouse. Defining the left mouse button as affirmative, the right button as negative, and using the wheel for jogging through choices we designed an intuitive mouse control. To keep it as natural as possible it functions without an on-screen menu. Speech input and the mouse interface can be used at the same time, no switching between input modes is necessary. However, for dictating notes speech input is mandatory.

The user selects objects with the view pointer, i.e. by keeping them centered in the field of view for longer than one second. The assistant then utters the name of the object and points at it. An easy accessible help function lets the assistant tell about the current state and give a hint on how the user can interact.

Figure 6. The assistant during the presentation of an object

4 EVALUATION

We evaluated the system in an outdoor setting with 13 volunteers (7 male, 6 female). The environment model for the evaluation contained several waypoints on our campus and paths between them. The waypoints contained separately selectable objects. First the subjects had to make appointments, take notes, make the assistant describe and present real world objects (as shown in figure 6) and use the navigation function. Each task had to be done once by speech interaction and once by mouse control. After that, a questionnaire had to be completed.

All participants considered such an assistant as useful and the use of anthropomorphic interfaces as seminal. While speech recognition in general needs to be improved, the applied kind of natural communication was considered as very intuitive, as was the view pointer for object selection. The mouse as an alternative input method was considered rather feasible. The users primarily complained about the limited field-of-view and brightness of the display. 7 users would like to have further feedback (due to the fact that the virtual assistant was often outside the field of view).

5 CONCLUSIONS AND FUTURE WORK

This paper introduced MARA – our omnipresent virtual assistant, representing a first step towards integrating PIM functionality with services that are usually offered by personal guides and secretaries and a familiar interface of a human person.

Based on our findings we would expect that such a system would enjoy wide user acceptance. In order to use anthropomorphic characters in mobile AR systems however, significant improvements in underlying technologies (as displays and speech recognition systems) are required. In times the assistant could not be seen by the user, the system degraded to an auditory assistance system and was found useful even so. On that score, we believe an evaluation against a similar system without a virtual character would bring interesting results.

Our intention is to extend the system to support additional awareness and feedback mechanisms and to make the system more flexible and scalable (as suggested in [5]), as well as to integrate online communication possibilities.

ACKNOWLEDGEMENTS

The authors thank their colleagues at Fraunhofer FIT, especially Jan Ohlenburg, for their support and contributions regarding the implementation of MARA. Further thanks belong to the participants of the evaluation for their valuable feedback.

REFERENCES

[1] Mahoro Anabuki, Hiroyuki Kabuka, Hiroyuki Yamamoto, and Hideyuki Tamura. Welbo: An Embodied Conversational Agent Living in Mixed Reality Space. In *CHI '2000 extended abstracts*, pages 10–11, 2000

[2] Selim Balcisoy, Marcelo Kallmann, Remy Torre, Pascal Fua, and Daniel Thalmann. Interaction Techniques with Virtual Humans in Mixed Environments. In *Proc. International Symposium on Mixed Reality*, Yokohama, Japan, 2001

[3] Istvan Barakonyi, Markus Weilguny, Thomas Psik, Dieter Schmalstieg. MonkeyBridge: Autonomous Agents in Augmented Reality Games. In *Proc. of ACM SIGCHI International Conference on Advances in Computer Entertainment Technology (ACE'05)*, Valencia, Spain, 2005

[4] Istvan Barakonyi, and Dieter Schmalstieg. Ubiquitous Animated Agents for Augmented Reality. In *Proc. of IEEE and ACM International Symposium on Mixed and Augmented Reality 2006 (ISMAR'06)*, Santa Barbara, CA, USA, 2006

[5] Wolfgang Broll, Léonie Schäfer, Tobias Höllerer, and Doug Bowman. Interface with Angles: The Future of VR and AR Interfaces, *IEEE Computer Graphics and Applications, 21 (6)*, 2001

[6] Steven Feiner, Blair MacIntyre, Tobias Höllerer, and Anthony Webster. A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment. In *Proc. ISCW '97*, pages 74–81, Cambridge, MA, USA, 1997

[7] Fritz Hohl, Uwe Kubach, Alexander Leonhardi, Kurt Rothermel, and Markus Schwehm. Nexus – An Open Global Infrastructure for Spatial-Aware Applications. In *Proc. of The Fifth Annual International Conference on Mobile Computing and Networking (MobiCom '99)*, Seattle, WA, USA, 1999

[8] Daniel Holweg and Oliver Schneider. GEIST – Mobile Outdoor AR-Informationsystem for historical education with digital storytelling. In *Federal Ministry of Education and Research (BMBF), Virtual and Augmented Reality Status Conference*, 2004

[9] Stefan Kopp, Bernhard Jung, Nadine Leßmann, and Ipke Wachsmuth. Max – A Multimodal Assistant in Virtual Reality Construction. *KI – Künstliche Intelligenz 4/03, Issue on Embodied Conversational Agents*. Pages 11–17, Bremen, Germany, 2003

[10] Jan Ohlenburg, Iris Herbst, Irma Lindt, Thorsten Fröhlich, and Wolfgang Broll. The MORGAN Framework: Enabling Dynamic Multi–User AR and VR Projects. In *Proc. of the ACM Symposium on Virtual Reality Software and Technology (VRST 2004)*, pages 166–169, ACM, 2004

[11] Gerhard Reitmayr, and Dieter Schmalstieg. Collaborative Augmented Reality for Outdoor Navigation and Information Browsing. In *Proc. of Symposium Location Based Services and TeleCartography*, vol. 66, pages 53–59, 2004

[12] Tom Sephton. Teaching agents for wearable augmented reality systems. In *Proc. ISWC '03*, pages 250–251, 2003

[13] Vassilios Vlahakis, Thomas Pliakas, Athanassios Demiris and Nikolaos Ioannidis. Design and Application of an Augmented Reality System for continuous, context-sensitive guided tours of indoor and outdoor cultural sites and museums. In *Proc. VAST '03*, pages 155–164, Brighton, UK, November 5–7, 2003

[14] Ipke Wachsmuth, Britta Lenzmann, T. Jörding, Bernhard Jung, Marc Latoschik, and Martin Fröhlich. A virtual interface agent and its agency. In *Proc. of the First International Conference on Autonomous Agents*, pages 516–517, ACM Press, 1997

[15] ARToolKit website: http://www.hitl.washington.edu/artoolkit

[16] Cal3D website: http://cal3d.sourceforge.net

[17] GPX website: http://www.topografix.com/gpx