A Multimodal User Interface Component for an Augmented Reality Mobile User Guidance System

Marco Thum
University of Koblenz-Landau
Working Group Computer Graphics
Campus Koblenz, Germany
mthum@uni-koblenz.de

Thanos Demiris
Content Delivery Systems
Intracom S.A.
Peania, Athens, Greece
dema@intracom.gr

Stefan Müller
University of Koblenz-Landau
Working Group Computer Graphics
Campus Koblenz, Germany
stefanm@uni-koblenz.de

ABSTRACT

In general, user interfaces should be intuitive, self-explanatory and adaptive to various user skills. Especially in augmented reality systems with complex interaction possibilities simple concepts are necessary to guide the user. Therefore, we extend the ideas of the traditional WIMP metaphor (Windows, Icons, Menus, Pointer) with the capabilities of multimodal interfaces where multiple ”human” communication channels are used as input data for navigation, orientation and interaction.
In this work we will present the adaption of the user interface of an existing mobile augmented reality system for cultural heritage to multimodal interaction. We present the concepts that led us to the decision for the use of speech and capturing of hand movements by means of an inertial tracker, as well as the implementation aspects of the initial prototype. User evaluation trials will prove our approach.

Keywords: Augmented reality, mixed reality, graphical user interfaces, multimodal user interfaces, human computer interaction, software ergonomics, cultural heritage.

1 INTRODUCTION

The system is embedded as a client in the architecture of the intCulture [2] solution for the management and dissemination of multimedia content related to cultural heritage, featuring an extensive back-office and a set of supported client platforms. One of them is the augmented reality mobile guide, for which we implemented the multimodal interaction system presented herein. It integrates the ARBrowser developed at Fraunhofer IGD and featuring markerless tracking [8]. Because of less training necessity we decided for speech as the primary input device incorporating a speaker-dependent recognition engine. A sub-notebook, a video-see-through head mounted display resp. binoculars, an one-ear headset, a GPS sensor and an electronic compass are forming the hardware part of the system. An ordinary desktop mouse resp. a wireless remote control with inertial tracker have been chosen as secondary input devices.

2 RELATED WORK

We examined the usage of interaction metaphors and input devices of several kinds of AR user interfaces addressing various application domains, tasks and varying user skills. Guidelines and hypotheses have been derived from [1, 4, 9] et al.

Posters proceedings, ISBN 80–86943–04–6
WSCG'2006, January 30 – February 3, 2006
Plzen, Czech Republic.

3 METHODOLOGY

We continued defining hypotheses proving the general user dialogue and guidance principles laid down by [3]. Characteristics of multimodal interfaces [7] were also evaluated in the user trials. Several use cases followed which fit possible user interactions and needs observed by [5].

3.1 Application Scenario

Visible elements can be dynamically arranged all over the screen. A thumbnail view presents strictly location-sensitive guiding content like contained, adjacent or related site elements. We integrated object related information access in augmented menus aligned as translucent overlays on top of the augmented image as fig. 1 demonstrates. The inertial tracked input device was additionally used for recording and analysis of hand movements to deliver appropriate feedback to the user's actions. The speech dictionary is extended dynamically with keywords of recently visited exhibits.

3.2 System Architecture

For the detailed system design we extracted basic functionality needs out of the scenario and sorted them into patterns and components in accordance to [6]. Almost every component of the architecture is entirely configured via XML. Each input device has to be initialized using a separate configuration file and setting event filters. Events are processed as shown in fig. 2. System messages are logged via a centralized

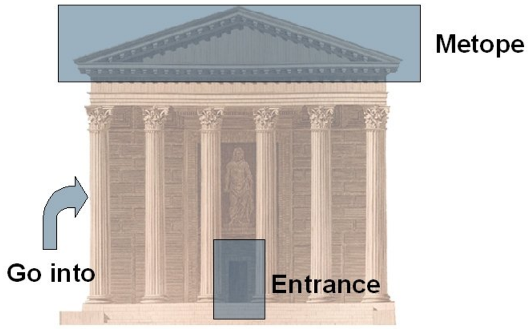

Figure 1: Translucent rectangles indicate additional information on exhibit parts. Handlers like pointers lead to manipulations on the exhibit (concept).

global mediator system and appropriate feedback is presented as audio or text.

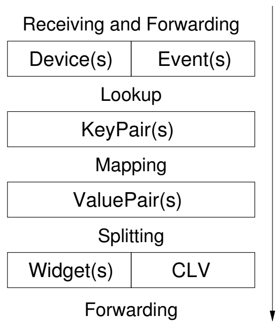

Figure 2: Events are dispatched to widgets by comparing tuples to a key map. Widgets interpret events as Common Layer Values (CLV) which are independent from the event description sent by the device. This leads to a highly configurable and extensible event management.

3.3 User evaluation

For the initial trials we separated 14 users into 2 groups. The first group was requested to use a head-mounted display and an ordinary desktop mouse. The second one interacted with a wireless, inertial tracked remote control and used binoculars for the video output. The speech recognition engine was trained to a greek male and greek female voice. A german male profile has been set up for almost acentless speakers. The test system has been arranged locally. A questionnaire should prove the hypotheses deployed in the design. A short extract of the evaluation results is given in fig. 3.

4 CONCLUSION & FUTURE WORK

On the basis of the outcome of the evaluation a set of adaptations will take place, which will lead to the second prototype of the system. Randomly selected visitors of a selected site (either the archaeological site of ancient Pompeii in southern Italy or ancient Olympia in Greece) will evaluate the improved system.

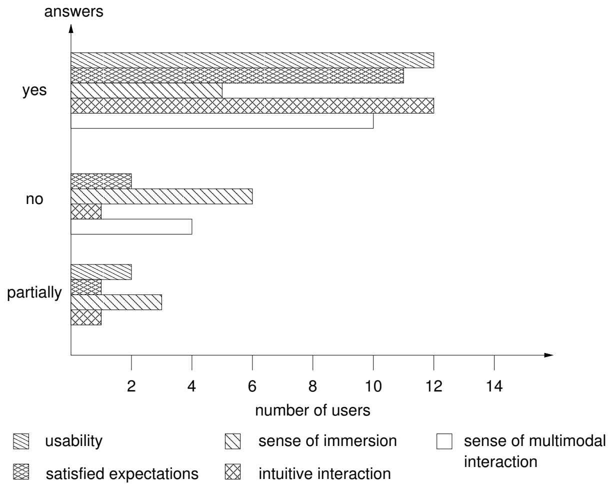

Figure 3: Evaluation results of guidelines and hypotheses.

5 ACKNOWLEDGEMENTS

We would like to thank Nikos Pronios and Nikos Ioannidis as representants of Intracom S.A., Peania, Greece, for the support and funding of this project.

REFERENCES

[1] O. Achenbach, D.-I. S. Göbel, and O. Schneider. GEIST – A Narrative AR Learning Application. *INIGRAPHICS – computer graphik topics*, 16(3), 2004.

[2] T. Demiris, V. Vlahakis, and N. Ioannidis. intCulture: Location-based Multiplatform Publishing of Cultural Heritage Information. In *Proceedings of the ICHIM 2004*, pages CD–ROM paper C2, 2004.

[3] European Committee for Standardization. Ergonomic requirements for office work with visual display terminals (EN ISO 9241-(10 &13)). Technical report.

[4] T. Höllerer, S. Feiner, T. Terauchi, G. Rashid, and D. Hallaway. Exploring MARS: Developing Indoor and Outdoor User Interfaces to a Mobile Augmented Reality System. *Computer & Graphics*, 23(6):779–785, 1999.

[5] S. Hsi. The Electronic Guidebook: A Study of User Experiences using Mobile Web Content in a Museum Setting . In *Proceedings of the IEEE International Workshop on Wireless and Mobile Technologies in Education* , 2002.

[6] A. MacWilliams, T. Reicher, G. Klinker, and B. Bruegge. Design Patterns for Augmented Reality Systems. In *IUI-CADUI Workshop on Exploring the Design and Engineering of Mixed Reality Systems*, volume 91, 2004.

[7] S. Oviatt. Multimodal Interfaces. In Julie A. Jacko and Andrew Sears, editor, *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, chapter 14, pages 286–304. 2002.

[8] D. Stricker. Tracking with Reference Images: A Real-Time and Markerless Tracking Solution for Out-Door Augmented Reality Applications. In *Proceedings of Conference on Virtual Reality, Archeology, and Cultural Heritage*, pages 77–82, Darmstadt, Germany, 2001.

[9] J. Yang, W. Yang, M. Denecke, and A. Waibel. Smart Sight: A Tourist Assistant System. In *ISWC*, pages 73–78, 1999.