

Proceedings of the Joint Workshop

Mobile Interaction with the Real World (MIRW 2007)

www.mimuc.de/mirw2007

5th Workshop on "HCI in Mobile Guides" (MGuides 2007)

www.mguides.info

in Conjunction with the

9th International Conference on
Human Computer Interaction with Mobile Devices and Services
(MobileHCI 2007)

Gregor Broll, Alexander De Luca, Enrico Rukzio,
Chie Noda, Paul Wisner (Eds. MIRW 2007)

Keith Cheverst, Barbara Schmidt-Belz (Eds. MGuides 2007)

Singapore, September 2007

Preface

Welcome to the joint MobileHCI 2007 workshop of “Mobile Interaction with the Real World” (MIRW 2007) and “HCI in Mobile Guides” (MGuides 2007). We accepted 13 papers that will be presented within the workshop. We hope that the union of these two individual yet complementary topics will provide interesting insights for the combined audiences and create a lively discussion.

We would like to thank the authors for their contributions and the organizers of MobileHCI 2007 for hosting this joint workshop. Furthermore we would like to thank the research project Perci (www.hcilab.org/projects/perci/) funded by NTT DoCoMo Euro-Labs for the support to publish this workshop record in a printed form.

We look forward to the workshop providing a rich environment for academia and industry to foster active collaboration in the areas of mobile interactions with the real world as well as HCI in mobile guides.

Singapore, September 9, 2007

Gregor Broll, Keith Cheverst, Alexander De Luca, Enrico Rukzio,
Barbara Schmidt-Belz, Paul Wisner, Chie Noda

Mobile Interaction with the Real World (MIRW 2007)

Theme of the Workshop

Mobile devices have become a part of our everyday lives as most people rely on mobile phones, smart phones or PDAs as personal and pervasive communication devices. So far, mobile interaction mostly takes place between users, their mobile devices and basic services like phone calls, text messages or organizer functionalities.

Over the last years, there has been an increasing interest in extending the interaction between users and mobile devices to the interaction with objects from the everyday world. This development has benefited from the pervasiveness of technologies for the augmentation of people, places and things with additional information. Complementary, mobile devices have been established as ubiquitous computing platforms that provide the technologies to capture, process and use this information. For example, people can use their mobile phones to take pictures of visual markers and have their codes recognized. The usage of RFID/NFC is gaining in popularity as it can reduce payment, identification or access control to simply swiping a mobile phone over a reader. Mobile interaction with places - using e.g. GPS or cell positioning - is the foundation for location based services. Other areas of application that could benefit from this new kind of mobile interaction are smart objects such as advertisement posters, vending machines, pervasive gaming, mobile services or information systems, e.g. in museums or at exhibitions.

Following the success of Mobile Interaction with the Real World; at MobileHCI 2006, we would like to continue this workshop as a forum that concentrates on mobile interactions with real world objects. Among the questions that come up in this area of research are:

- Which kinds of interactions with the real world exist?
- What technologies can be used to implement mobile interaction with the real world?
- How can real world objects and services be described?
- How should systems and services for this kind of mobile interactions be designed?
- What should these user interfaces look like?
- What does the interaction design and usability look like for mobile interaction with physical objects?
- Can these interfaces be generated automatically?
- Should real world services be defined in a standardized way (e.g. with semantic web services)?
- How can real world objects be associated with new services?
- Which issues concerning privacy and security arise from this new kind of mobile interaction?

Topics

Possible topics for the workshop include (but are not limited to):

- Interactions between mobile devices and objects from the real world
- Automatic user interface generation
- Semantic Web within mobile applications and interactions
- Security and privacy aspects
- Using mobile devices as user interfaces for terminals and vending machines
- Using objects from the real world as pervasive interfaces extending interfaces on mobile devices
- Guidelines for mobile interactions with the real world
- Authoring support for physical mobile applications
- Multimodal interaction taking mobile devices into account
- Usage of sensors of mobile devices (camera, microphone, GPS, etc.) for pervasive applications
- Interaction metaphors for pervasive applications and services
- Applications and scenarios

Goals

The main goal of the workshop is to develop an understanding of how mobile devices can be used when interacting with the real world. We will provide a forum to share information, results, and ideas on current research in this area. In addition we aim at the development and dissemination of new ideas on how mobile phones can be exploited for new ways of interacting with the environment. We will bring together researchers and practitioners who are concerned with design, development, and implementation of new applications and services using personal mobile devices as user interfaces.

Webpage

All information about the workshop, the papers and the proceedings are available at the website of the workshop <http://www.mimuc.de/mirw2007>

Organizers

Gregor Broll
University of Munich (Germany)
gregor.broll@ifi.lmu.de

Alexander De Luca
University of Munich (Germany)
alexander.de.luca@ifi.lmu.de

Enrico Rukzio
Lancaster University (UK)
rukzio@comp.lancs.ac.uk

Chie Noda
NTT DoCoMo, Inc. (Japan)
noda@nttdocomo.co.jp

Paul Wisner
Nokia Research Center (USA)
Paul.Wisner@nokia.com

Additional Reviewers

Michael Rohs (Deutsche Telekom Laboratories, Germany)
Jukka Riekkı (University of Oulu, Finland)
Albrecht Schmidt (University of Bonn, Germany)
Pasi Valkkynen (VTT, Finland)
Massimo Paolucci (NTT DoCoMo Euro-Labs, Germany)
Christof Roduner (ETH Zurich, Switzerland)
John Hamard (NTT DoCoMo Euro-Labs, Germany)

5th Workshop on "HCI in Mobile Guides" (MGuides 2007)

Theme of the Workshop

Today's mobile user demands easy access to relevant information services from a variety of devices (both personal and situated/public), whenever and wherever they need them. Example applications for mobile guides include: mobile tourism services, indoor and outdoor museum/exhibition/event guides and context-aware directory services. Although the latest mobile devices and information services offer new and enhanced ways to support nomadic users, they also raise challenges concerning interaction modalities, usability, accessibility and trustworthiness.

The need to consider the HCI implications of mobile guides, i.e. systems designed to guide a user who is moving in a physical environment by giving directions and supplying relevant information and access to services via some form of mobile device, is clear. Specific research in this area is required in order to ensure that the usability of mobile guides does not get overlooked in favour of the exciting technological advances in this area, e.g. the current marketing of mobile phones with built in GPS and other sensor technologies and integrated high-speed wireless networking capabilities. Indeed, the huge gap between technical feasibility and visions, on the one hand, and useful and usable applications on the other hand needs to be bridged by empirical research in a human-centred way. This workshop shall be a forum for this research.

The first workshop in the series was held in conjunction with the Mobile HCI 2002 conference in Pisa, Italy. Since then "HCI in Mobile Guides" workshops have been held in Udine (2003), Glasgow (2004) and Salzburg (2005). To date the workshops in this series have brought together approximately 120 researchers and practitioners and has provided a research/paper archive for many more.

Topics

A range of topics are relevant to a discussion on the human computer interaction issues relating to mobile guide systems. In this workshop, the following topics are of particular relevance:

- Appropriate techniques for supporting the User-centred and/or Participatory Design of mobile guides.
- Accessibility for particular groups, e.g. older users, the visually impaired etc.

- Suitability of different interaction modes for mobile guides. e.g. utilising multimodal input/output, anthropomorphic approaches, etc.
- Fostering user-understanding of adaptive behaviour, e.g. location awareness, and the uncertainties associated with such behaviour, e.g. due to inaccuracies in GPS readings etc.
- Visualization of the spatial environment, Augmented Reality. 2D/3D maps etc.
- Conveying dynamic information, e.g. changes to available services, changes to the underlying physical model etc. and supporting information retrieval whilst faced with changing infrastructure conditions.
- Leisure/entertainment use of mobile guides (e.g. by games on treasure hunts or to support spontaneous social gatherings).
- User Interface techniques to facilitate access to heterogeneous and/or distributed services.
- Group/community support provided by mobile guides, e.g. social navigation, facilitating group rendezvous, use/sharing of content produced by the community/other visitors etc.
- Personalization of services, e.g. use of user modelling techniques.
- User evaluation of mobile guides and understanding mobile guide use.
- Design principles or heuristics for mobile guides.
- Approaches to (and results of) requirements analysis for mobile guides, e.g. the use of ethnographic techniques etc.
- Design solutions for “baby interfaces” on mobile guides, i.e., small buttons, small screens and small interaction devices.
- Introducing the services to use; facilitating the out-of-box experience.
- Issues arising from the opportunities and challenges provided by multimodal user interfaces.
- Designing for usage “in the wild”: design and evaluation of mobile guides in natural environments, etc.
- The design and use of situated displays for supporting activities related to tourism/guides, e.g. supporting information retrieval, sharing of content etc.
- Issues arising from supporting user/community generated content for mobile guides.

Goals

The main goal of this workshop is to bring together experts who develop or evaluate mobile guides and wish to share and discuss their experiences in this workshop. Aspects of human-computer interaction are to be addressed explicitly, empirical user studies being considered the most relevant method of research here.

Webpage

All information about the workshop, and indeed all previous workshops in the series are available at the workshop web-site <http://www.mguides.info>

Organizers

Keith Cheverst
Lancaster University (England)
kc@comp.lancs.ac.uk

Barbara Schmidt-Belz
Fraunhofer FIT (Germany)
Barbara.Schmidt-Belz@fit.fraunhofer.de

Program Committee and Additional Reviewers

Lynne Baillie (FTW, Vienna)
Mark Dunlop (Strathclyde University, Scotland)
Connor Graham (Melbourne University, Australia)
Keith Mitchell (Lancaster University, England)
Elke-Maria Melchior (ACIT, Germany)
Jesper Kjeldskov (Aalborg University, Denmark)
Jeni Paay (Aalborg University, Denmark)
Eija Kaasinen (VTT Information Technology, Finland)
Chris Kray (Lancaster University, England)
Janna Hakkila (Nokia, Finland)
Tsvi kuflik. (University of Haifa, Israel)

Schedule and Table of Contents

09:00 – 09:15	Welcome
	Augmenting the City with Fiction: Fictional Requirements for Mobile Guides <i>Jesper Kjeldskov, Jeni Paay (Department of Computer Science, Aalborg University, Denmark)</i>
09:15 – 10:30	
Session 1	A Multimodal Mobile Museum Guide for All <i>Carmen Santoro, Fabio Paterno, Giulia Ricci, Barbara Leporini (ISTI-CNR, Italy)</i>
Chair: Keith Cheverst	Heritage & Habitus: Designing to Support Situated, Living Knowledge <i>Nicola J Bidwell (James Cook University, Australia), truna aka j.turner (Australasian CRC for Interaction, Design Pty Ltd. QUT, Brisbane, Australia), Jason Holdsworth, Colin Lemmon, Michael Shay (Discipline of IT, Cairns, Australia)</i>
10:30 – 11:00	Break
	Traffic Advisories Based on Route Prediction <i>Kari Torkkola, Keshu Zhang, Haifeng Li, Harry Zhang, Christopher Schreiner, Mike Gardner (Motorola, Intelligent Systems Lab, USA)</i>
11:00 – 12:30	
Session 2	A Gaze-based Guidance System based on a Real-world 3D Photo Logging System <i>Rieko Kadobayashi (National Institute of Information and Communications Technology, Japan)</i>
Chair: Enrico Rukzio	A Mock-up System for the Early Testing of Location Based Services <i>Florian Klompmaker, Claudius Stern (University of Paderborn, Germany), Christian Reimann, Holger Santelmann (Siemens IT Solutions and Services, Germany)</i>
	Demonstrating Coherent Interactions between Personal Mobile Devices and Situated Installations <i>Ben Bedwell, Boriana Koleva (Mixed Reality Laboratory, University of Nottingham, UK)</i>
12.30 - 14.00	Lunch
	Contextual Bookmarks <i>Niels Henze (OFFIS, Germany), Mingyu Lim (University of Geneva, Switzerland), Andreas Lorenz (FHG-FIT, Germany), Michael Mueller (Lancaster University, UK), Xavier Righetti (VRlab, Switzerland), Enrico Rukzio (Lancaster University, UK), Andreas Zimmermann (FHG-FIT, Germany), Nadia Magnenat-Thalmann (University of Geneva, Switzerland), Susanne Boll (University of Oldenburg, Germany), Daniel Thalmann (VRlab, Switzerland)</i>
14.00 - 15.30	
Session 3	Bluetooth Familiarity: Methods of Calculation, Applications and Limitations <i>Barry Lavelle, Daragh Byrne, Cathal Gurrin, Alan F. Smeaton, Gareth J.F. Jones (Centre for Digital Video Processing, Dublin City University, Ireland)</i>
Chair: Michael Rohs	Semantic Integration and Language Access to Mobile Data <i>Raimondas Lencevicius, Alexander Ran (Nokia Research Center Cambridge, USA)</i>
15:30 - 16.00	Break
	Koubachi: A Mobile Phone Widget to Enable Affective Communication with Indoor Plants <i>Philipp Bolliger, Benedikt Ostermaier (Institute for Pervasive Computing, ETH Zurich, Switzerland)</i>
16.00 - 17.30	
Session 4	Swiss Army Knife meets Camera Phone: Tool Selection and Interaction using Visual Markers <i>Christian Kray (Informatics Research Institute, Newcastle University, UK), Michael Rohs (Deutsche Telekom Laboratories, Germany)</i>
Chair: Gregor Broll	Mediating Multi-Party Negotiation Through Marker-Based Tracking of Mobile Phones <i>Michael Rohs (Deutsche Telekom Laboratories, Germany), Christian Kray (Informatics Research Institute, Newcastle University, UK)</i>
17.30 - 18.00	Closing Discussion / Closing / Wrap-up

Augmenting the City with Fiction: Fictional Requirements for Mobile Guides

Jesper Kjeldskov and Jeni Paay
Aalborg University
Department of Computer Science
Fredrik Bajers Vej 7E
DK-9220 Aalborg East, Denmark
{jesper, jeni}@cs.aau.dk

ABSTRACT

In this position paper, we discuss the user experience of a new genre of mobile guide services, which augments the users' physical surroundings with *fiction* rather than with facts. First, we outline sources of inspiration from fiction, storytelling in place and on the move, and related research within the area of interactive narratives and location-based games. We then present and discuss five overall types of fictional mobile guides embodied by the metaphors of 1) treasure hunts, 2) jig-saw puzzles, 3) playing dominos, 4) playing scrabble, and 5) collecting butterflies. Finally, we describe a preliminary design idea for a prototype system intended to explore the design, implementation, and user experience of fictional mobile guides.

Categories and Subject Descriptors

H5.2. [Information interfaces and presentation (e.g., HCI)]:
User Interfaces - *User-centred design, Graphical user interfaces.*

General Terms

Design, Human Factors.

Keywords

Location-Based Services, Interactive Narratives, Fiction

1. BACKGROUND

In our recent research, we have been exploring the design, implementation, and user experience of context-aware and location-based mobile information services, which “augment the city” with a digital layer of information about, for example, people, places and the users' physical surroundings [14]. Common for our research has been that the systems and services explored have provided users with factual information, that would sometimes otherwise be invisible, from which people could then decide what to do while socialising out on the town; where to go, how to get there, what to do there, who to meet up with etc. In many ways, this matches the majority of research and design being done within the area of mobile guides: providing users on

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MobileHCI '07, September, 2007, Singapore.

Copyright 2004 ACM 1-58113-000-0/00/0004...\$5.00.

the move with facts about the reality that they are in. However, let's face it, people cannot live by reality alone! Reality is everywhere. It is right there in our face when we wake up in the morning, and it continues to sneak up on us throughout our day, in the car, in our offices, in the tearoom, and at the dinner table. In recent years it has even invaded our TVs; not only through eternal streams of live news from wherever, but also through endlessly boring reality TV shows documenting minute-by-minute details of peoples' mundane lives. Perhaps this is why we spend quite a lot of effort (and money) on escaping reality. People need wonderment [19]. We daydream, we go to the cinema, we rent a video, or we read or listen to a book. We like to immerse ourselves in imaginary worlds and stories as way of escaping every day life. We like to immerse ourselves in *fiction*.

So why not let mobile guides provide people with fiction about their physical surroundings rather than just facts? Inspired by this question, we have engaged with a stream of research exploring the user experiences achieved by “augmenting the city with fiction” and looking at some fundamental questions related to doing this. How can mobile guides provide for peoples' desire for fictional content? How can quality storytelling benefit from its listeners being on the move or located in a particular place? How can we characterise different types of context-aware or location-based stories? How can we narrate engaging stories, which take into consideration the listener's location, movements over time, and perhaps even social context? How can we design and implement useable media for this genre of storytelling? In short, what are the fictional requirements of a mobile guide?

2. INSPIRATION AND RELATED WORK

Making interactive narratives and relating stories to locations are not new ideas on their own. However, combining the two into dynamic narratives that respond to the users/receiver's location and interactions in physical space over time is a relatively new phenomenon with key references in the research literature dating only a few years back.

Early context-aware mobile guides such as *Cyberguide* [1], the Lancaster *GUIDE* [7] and *Hippie* [18] provided visitors, city travellers and museum tourists with location-based experiences which included pertinent information and guidance about their current location, but did not focus on narrative issues. However they act as useful vehicles for exploring and understanding important interaction and technical issues about designing mobile guides. Location-based mobile games such as *Pirates!* [4], *ARQuake* [20], *Mindwarping* [21] and *Can You See Me Now?* [2] offer inspiration for potential new applications for location-based technologies, and offer us additional vehicles for studying

important aspects of interaction design for emerging context-aware technologies designed to operate ‘in the wild’. In particular, *Can You See Me Now?*, with its engaging artistic experience of seamlessly connecting online players in a virtual world to play against real players in the actual city, demonstrates new potentials for human interaction with computers and with each other.

Beyond these applications that provide users with a straightforward gaming experience, is the emerging area of location-based games, which incorporate stories into the mix to enhance the user experience with fiction. *Backseat Playground* [12] provides a narrated experience interweaving a crime mystery, visible physical elements of the environment and gaming for children sitting in the back seat of a car, while on a journey. *Uncle Roy All Around You* [3] is a programmed game that involves the surrounding city, live actors, online and street players to follow a set of pre-scripted clues to find Uncle Roy’s office.

The delivery of location-based stories using mobile technology is a new area of research, which includes applications such as *Urban Tapestries* [13], *Riot!* [5], *Hopstory* [17] and *Geist* [15]. *Urban Tapestries* is a collection of stories, histories, experiences and events of a community linked to familiar and related locations. *Riot!* delivers authored episodes triggered by specific locations about an historical riot that happened in and around Queens Square where the installation operates. *Hopstory* allows visitors to an historic brewery to collect location related video vignettes as they navigate the building, and then view their collected story at the end of the visit. *Geist* provides an augmented reality experience for visitors to the city of Heidelberg, which combines history and fiction to allow the viewer to experience how it felt to be in that city during the 30 Years’ War in the 17th century. These applications are the most inspiring for our current research as they provide experience and understanding of peoples responses to a city augmented with fiction.

Anecdotally, people’s desire for storytelling and important aspects of that fiction can be seen in popular stories of our time. Many best-selling novels illustrate the appeal of blurring of boundaries between fact and fiction, supporting the need for the “suspension of disbelief” in story telling. The *Da Vinci Code* [6] sold over 60.5 million copies, was translated into 44 languages world-wide, is the thirteenth best selling book of all time, and inspired a feature-length movie which was the second highest grossing movie of 2006, world-wide. The *Da Vinci Code* establishes its sense of intrigue and allurement from a clever interweaving of things that we know to be historical facts and real places with “believable” stories involving these elements. Crime fiction becomes all the more evocative when it locates its scenes in places that we are familiar with, and invites us to believe that this has happened in that location. For example, crime stories by the prolific Melbourne writer, Kerry Greenwood, about a Melbourne baker [9][10][11] walks us through the streets of Melbourne and possibilities of dark happenings that we are unaware might be occurring in the city we think we know so well. This technique is also brought into play in several location-based storytelling applications, for example, *Riot!* [5] *Hopstory* [17] and *Geist* [15], all bring this technique into play.

Further supporting our understanding that people are looking for aural input while on the move is the meteoric rise in popularity of the iPod, mp3 players generally, and now mobile phones with audio listening capabilities to supplement and even disconnect people from their daily realities. This has recently been extended beyond music to the downloading of pod-casts and e-books to add

an element of storytelling to our lives in situations where we have previously had to resort to the printed word (i.e. on the way to work in the train) or broadcast media and narratives, and therefore centrally controlled (i.e. listening to the radio while driving). Now we can decide what we want to listen to and where – so why not take this further and incorporate our surrounding environment into the story.

3. AUGMENTING THE CITY w. FICTION

Inspired by the works and research described above, we have explored conceptual ideas for mobile guide systems that augment the city with fiction related to the users’ location and immediate surroundings. So far, our work has consisted of literature research and a series of open-ended brainstorming sessions and design workshops during the first half of 2007. In these sessions and workshops we have particularly focussed on the interplay between stories, locations, time, and people’s social context, as well as exploring the interplay between interactivity and narrative structure. From this preliminary work, we have produced five metaphors for different types of fictional mobile guides:

- 1) Treasure hunts
- 2) Jig-saw puzzles
- 3) Dominos
- 4) Scrabble
- 5) Collecting butterflies

These are described in detail below.

3.1 Five Metaphors for Fictional Guides

3.1.1 Treasure hunts

“Treasure hunts” are linear stories characterized by a fixed series of parts or chapters, which the receiver has to read or listen to in correct sequence, as prescribed by the author, in order to get to the end. As when finding a hidden treasure from a map, they have to follow the correct path on the treasure map to get from a designated starting point to the place marked by the red X.



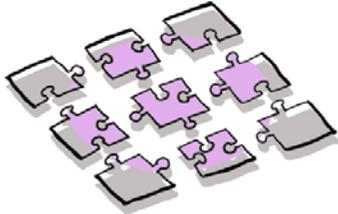
Applying the “treasure hunt” metaphor to the user experience of a fictional mobile guide, this means that users have to move physically from place to place, in the correct sequence, in order to get through the story as envisioned by the author. The story has a fixed starting location as well as a fixed end location. This provides the author with a large degree of control over the narrative in terms of classic principles of storytelling such as creating a plot, building up tension, deciding, introducing key characters at key times, etc. However, because each fragment of the “treasure hunt” story is linked to a specific physical location, the author furthermore gains an opportunity to use features of this location as a powerful backdrop for their story (e.g. visuals, sounds, smells, buildings, typical people etc.). At the same time, of course, this also makes the story sensitive to the dynamics of physical space, for example, throughout the day and the night and across the different seasons of the year. From a user experience point of view, “treasure hunt” types of fiction in the city require a

high level of mobility and interaction in the physical world, but provide a low level of interactivity with the actual narrative both in terms of its content and in terms of its composition.

Uncle Roy All Around You [3] is an example of this type of fictional mobile guide, where you need to follow pre-scripted clues to get to a specific end point, with clues attached to specific locations.

3.1.2 Jig-saw puzzles

“Jig-saw puzzles” are non-linear stories characterized by consisting of a number of parts or chapters that the receiver may read or listen to in any order they like in order to create the whole story. Just like assembling a jigsaw puzzle, the user has to work out which pieces fit together in order to create a meaningful whole. Any piece can be picked up at any time, but each piece only makes sense in relation to particular other pieces. The sequence of assembly does not affect the outcome in itself but may influence peoples’ experience of the story. All pieces have to be considered in order to complete the picture.



Applying the “jigsaw puzzle” metaphor to a fictional mobile guide, users are free to pick up the different pieces of a story located around the city in any sequence they like, but have to find all the pieces, and put them together themselves, in order to complete the story as thought of by the author. The story has no fixed starting location and no fixed end location. Obviously, this takes some of the author’s control away in respect to the flow of the narrative, and requires the author to create meaningful fragments that can be read or listened to individually and in random order. However, from a user experience point of view, “jigsaw puzzle” types of fiction in the city increase the level of interactivity in terms of composition, and have a built-in quality for exploration in terms of finding the next piece as well as putting the pieces together. For example, one story fragment may inspire what pieces to look for next, and contain clues about where in the city to find them.

Riot! [5] can be regarded as fictional mobile guide of this type, as it uses a collection of separate vignettes, all about one central story, which can be collected at controlled locations and pieced together by the listener to create a cohesive picture of the entire drama.

3.1.3 Dominos

“Dominos” are selections of self-contained short stories related to physical locations where the next possible story depends on the current one. There is no overall storyline, and not necessarily an end (unless all the pieces have been used), and the sum of the pieces does not necessarily make up a greater whole (as with the jigsaw puzzle). However, each step of the way there a link from one story to the next by means of some similarity between them. The meaning lies in the transition. Similar to playing dominos, each piece is a separate entity, but



can be paired by means of similarity from one to another. The first piece limits the number of possible subsequent pieces, and only through specific other pieces can some others now be fitted.

Applying the “dominos” metaphor to the user experience of a fictional mobile guide, when people pick up a story in a place, the next story given to them will have some sort of link with this one. As an example, entering a café may trigger a story about a murder in this place, which could then trigger a story about a related murder on a ship when subsequently going to the docks. Starting at the docks, however, could trigger a story about a sailor discharging from his ship, which could then trigger a story about a love affair between a sailor and a barmaid when subsequently going to the café. Each story is loosely related to the previous one but the author does not predetermine pairing them. From the author’s point of view, this gives complete control over each short story in terms of storytelling, and each story can be tailored to match the backdrop of a particular physical location. Through careful selection of keywords for pairing of stories, the author can to some extent influence the possible sequences by which stories can be put together. From a user experience perspective, “dominos” types of fiction in the city have limited interactivity in terms of controlling what story is presented where, but at the same time each user interaction along the way has a large impact on future discourse. Associations between each specific story provide the thread through the overall experience.

Hopstory [17] with its linear plot, delivered from different character viewpoints, provides a collection of scenes, which are interesting entities in themselves. They are collected at specific locations but in any order, and can be strung together at the end of the experience to create an ordered story, which is loosely controlled by time thereby exhibiting the characteristics of a “dominos” type fictional mobile guide. *Backseat Playground* [12] can also be regarded as this type of guide as its linear plot allows user interaction with the system to influence the story to branch in different directions.

3.1.4 Scrabble

“Scrabble” describes stories that have a high interactive complexity and, for example, consists of many small fragments, which can be put together in a large number of ways (with and without meaning). Similar to playing scrabble, the user is offered a number of random pieces, which may be put together to a larger whole. Each piece has very limited meaning on its own, but together they can make up meaningful constellations. As an example, a “scrabble” type of fiction in the city could take place on board a tram or bus and not only respond to the precincts moving past the windows but also to the people getting on and off during the journey sensed through, for example, Bluetooth phones and personal profiles. Every time a new passenger boards the tram, information from his or her profile is used to give the stories offered to the other passengers a new possible direction.



From the authors’ point of view, applying the “scrabble” metaphor makes it highly challenging to create meaningful stories by traditional means of storytelling because most of the control is left either in the hands of the users or in the hands of the serendipity of events taking place around them. As with all highly

dynamic and interactive narratives, this, of course, also makes “scrabble” type of fictional mobile guides highly challenging to implement. At the same time, however, it also opens up for highly immersive and engaging user experiences with a high level of interactivity as well as intriguing uncertainty about what and who it is in their surroundings that influences the course of a story.

A mobile fictional guide experience that exemplifies this metaphor is *Geist* [15]. From an availability of many different story pieces, the player uses pieces of stories collected from different locations to put together their own, personally unique story by the end of the experience. They also collect pieces of history and information along the way that do not necessarily form part of that final story, but are interesting pieces themselves.

3.1.5 Collecting Butterflies

“Collecting butterflies” are selections of self-contained short stories related to physical locations but with no predetermined relation to each other. Hence, as with “dominos”, there is not a greater storyline beyond the whole created by the receivers themselves. Similar to collecting butterflies, people are free to explore the physical space around them in pursuit of stories they like to collect. They may pick up an interesting story that they have spotted in their surroundings, take a closer look at it, keep it, let it go, or go looking for others of a



similar kind. The resulting collection is a highly individual patchwork of entities put together in a way that makes sense to each individual person.

The “collecting butterflies” metaphor provides a relatively simple approach to creating a fictional mobile guide. When a person enters a particular place, he or she is offered a number of stories that fit this location. When moving to another place, he or she is offered different stories. Staying at the same place for a longer period of time may cause more stories to appear or some stories to disappear. As for “dominos”, the “collecting butterflies” metaphor gives the author full narrative control over each short story in terms of sequence and relation to particular physical surroundings. However, control in terms of the creation of an overall picture is in the hands of the user. This results in a user experience that is highly interactive and exploratory on one side, but at the same time is also opaque and constrained.

Urban Tapestries [13] is a mobile fictional guide where the user can “collect butterflies” by roaming the city and collecting autonomous individual stories and histories which are overlaying the city in pertinent places.

4. A PRELIMINARY DESIGN IDEA

As a preliminary study, we are exploring the “collecting butterflies” approach to fiction in the city through the design of a simple prototype application. The idea of the system is to provide the reader with audio short stories related to their physical location, but at the same time to also provide the listener with some means of interaction influencing what story they are given

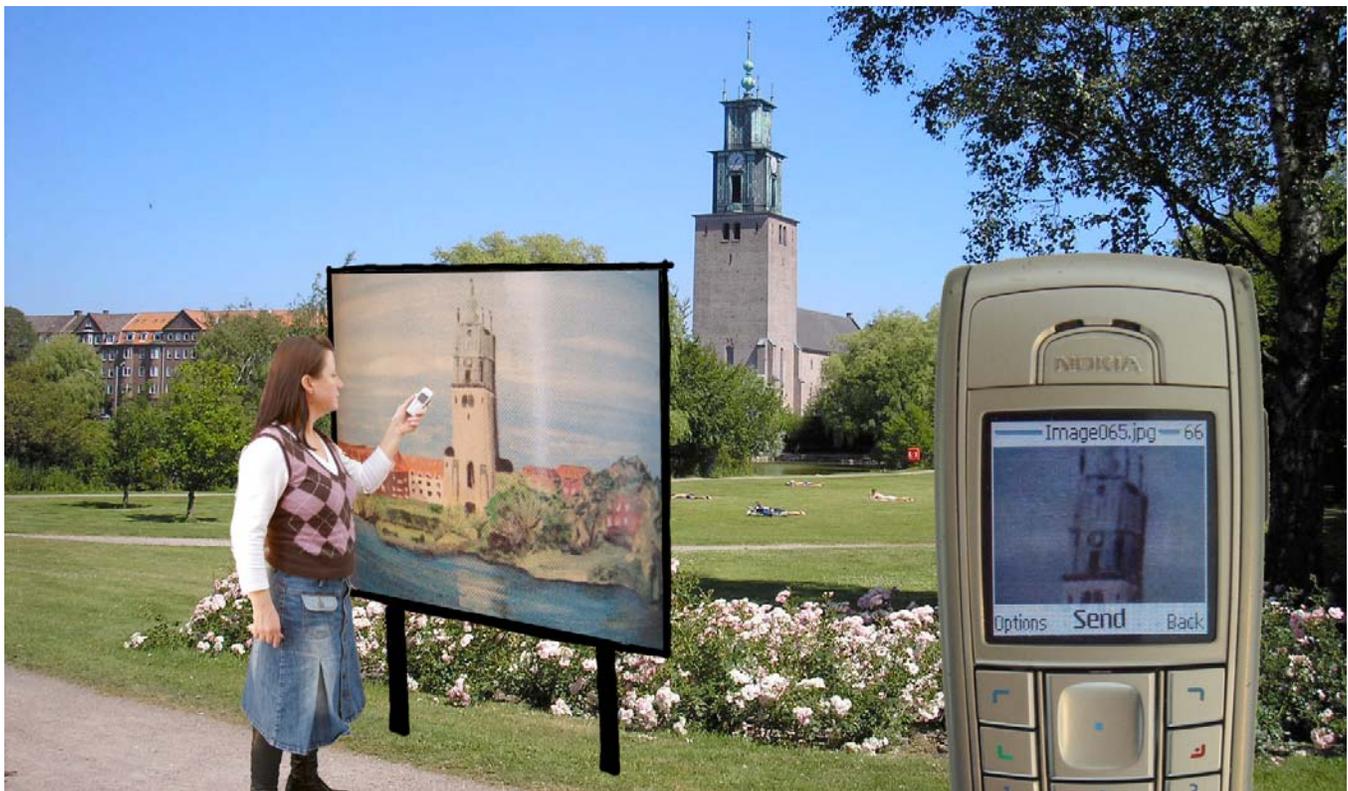


Figure 1. An example of “collecting butterflies” type of fiction in the city combined with ambiguous user interaction. By taking a picture with her mobile phone camera of a particular part of a painting situated in the park, the user will receive an audio short story related to this location and the particular motive captured. It could, for example, involve the old clock tower.

in a way that is playful, ambiguous [8] and opaque rather than goal-oriented and transparent. Technically, we wanted the design to be lightweight, use off-the-shelf mobile technologies (i.e. mobile phones with a built-in camera, MMS, and audio playback capabilities), and not require dedicated software to be installed.

Firstly, a number of salient locations around the city area of Aalborg are selected, and classified by location keywords (e.g. “waterfront”, “inner city”, “park”, “backstreets” etc.). Secondly, each physical location is captured in a painting with strong colours, which is digitised, printed as a large poster, and installed in a visible position at the depicted location. Each poster installation is given a unique ID number visible on the front along with a phone number for the storytelling service. Thirdly, around a hundred selected short stories (5-10 min) are recorded digitally, and classified by the same keywords used for classifying locations. Each story is also colour coded in terms of its general “mood”, after Marcus [16] (e.g. in Western culture, black = fear, red = passion, blue = calm, green = safety, yellow = happiness, white = purity). Thus, stories are related to locations and are also indirectly linked to the colours of the paintings.

The envisioned user experience of the system is as follows. When people wander around, they identify fictionally augmented areas by spotting the painting installations set up around the city. They can then walk up to one of these installations, take a picture of a given part of the painting with their phone-camera, and send this picture along with the ID number as an MMS to the phone number specified. The service then sorts out the location of the listener, and also analyses the picture for prominent colours (or a predefined “hot spot” in the painting). In response, a short story is selected that matches the location of the reader, and the colours (or hot spot) of the posted picture. This is then returned to the listener as audio attachment to an MMS message, who can now take the story with them and listen to it through their headset. At any time, the listener can repeat the process; take another picture, send it to the storytelling service, and receive a new short story. By logging the phone numbers used to post MMS messages, the service makes sure that the same story is not returned to a listener more than once.

As an alternative to MMS over GPRS, file exchange could be done via Bluetooth through access points embedded into the installations around the city. Audio file download could also be done over the 3G network or through WiFi.

5. SUMMARY

We have proposed the concept of “augmenting the city with fiction” as an approach to mobile guide services that presents their users with stories related to their physical surroundings rather than with facts about it. In an attempt to characterise different types of possible fictional mobile guides, we have proposed five metaphors describing different levels of narrative structure and possible interactivity. Embodying one of these metaphors, the metaphor of “collecting butterflies”, we have described a preliminary design idea for a prototype installation, which offers a collage of audio short stories through mobile phones in response to the listeners’ location and their “opaque” interaction with the service through picture-taking with their mobile phones.

Our future work involves the detailed design and implementation of this idea into a solid service that allows for longitudinal use by a large number of people. Future work also involves the development of design ideas and prototype fictional mobile guide

systems embodying other metaphors described in order to investigate their strengths and weaknesses, we also wish to investigate which different types and genres of stories fit well with each of the described metaphors.

REFERENCES

- [1] Abowd, G., Atkeson, C., Hong, J., Long, S., Kooper, R., and Pinkerton, M. Cyberguide: A mobile context-aware tour guide. *Wireless Networks*, 3 (1997), 421-433.
- [2] Benford, S., Crabtree, A., Flintham, M., Drozd, A., Anastasi, R., Paxton, M., Tandavanitj, N., Adams, M., and Row-Farr, J. Can You See Me Now? *ACM TOCHI*, 13, 1 (2006), 100-133.
- [3] Benford, S., Seager, W., Flintham, M., Anastasi, R., Rowland, D., Humble, J., Stanton, D., Bowers, J., Tandavanitj, N., Adams, M., Row-Farr, J., Oldroyd, A., and Sutton, J. (2004) The error of our ways: the experience of self-reported position in a location-based game, *Proceedings of UbiComp 2004*, Springer, London, 2004, 70–87.
- [4] Bjork, S., Falk, J., Hansson, R., Ljungstrand, P. Pirates! Using the Physical World as a Game Board. In *Proceedings of Interact 2001*. IOS Press, Netherlands, 2001.
- [5] Blythe, M., Reid, J., Wright, P., and Geelhoed, E. Interdisciplinary criticism: analysing the experience of riot! a location-sensitive digital narrative. *Behaviour & Information Technology*, 25, 2 (2006), 127-139.
- [6] Brown, D. *The Da Vinci Code*. Bantam Books, UK, 2003.
- [7] Cheverst, K., Davies, N., Mitchell, K., Friday, A., and Efstratiou, C. Developing a Context-aware Electronic Tourist Guide: Some Issues and Experiences. In *Proceedings of CHI 2000*. ACM Press, New York, 2000, 17-24.
- [8] Gaver, W., Beaver, J., and Benford, S. Ambiguity as a Resource for Design. In *Proceedings of CHI 2003*. ACM Press, New York, 233-240.
- [9] Greenwood, K. *Earthly Delights*. Allen & Unwin, Australia, 2004.
- [10] Greenwood, K. *Heavenly Pleasures*. Allen & Unwin, Australia, 2005.
- [11] Greenwood, K. *Devil's Food*. Allen & Unwin, Australia, 2006.
- [12] Gustafsson, A., Bichard, J., Brunnberg, L., Juhlin, O., and Combetto, M. Believeable environments – Generating interactive storytelling in vast location-based pervasive games. In *Proceedings of ACE 06*. ACM Press, New York, 2006, Article no. 24.
- [13] Jungnickel, K. Urban Tapestries: Sensing the City and other Stories. *Proboscis "Cultural Snapshot" no. 8*, available at <http://urbantapestries.net/> (2004).
- [14] Kjeldskov J. and Paay J. Public Pervasive Computing: Making the Invisible Visible. *IEEE Computer* 39, 9 (2006), 60-65.
- [15] Malaka, R., Schneider, K., and Kretschmer, U. Stage-Based Augmented Edutainment. In *LCNS 3031* (2004), 54-65.
- [16] Marcus, A. *Graphic Design for Electronic Documents and User Interfaces*. ACM Press, New York, 1991.

- [17] Nisi, V., Wood, A., Davenport, G., and Oakley, I. Hopstory: an Interactive, Location-based Narrative Distributed in Space and Time. In *Proceedings of TIDSE 2004*. Springer, Berlin, 2004, 132-141.
- [18] Oppermann, R. and Specht, M. A Context-sensitive Nomadic Information System as an Exhibition Guide. In *Proceedings of HUC2K*. Springer-Verlag, London, 2000, 127-142
- [19] Paulos, E and Beckman, C. Sashay: Designing for Wonderment. In *Proceedings of CHI 2006*. ACM Press, New York, 2006, 881-884.
- [20] Piekarski, W. and Thomas, B. ARQuake: The Outdoors Augmented Reality System. *Communications of the ACM*, 45, 1 (2002), 36-38.
- [21] Starner, T., Liebe, B., Singletary, B., and Pair, J. MIND-WARPING: Towards Creating a Compelling Collaborative Augmented Reality Game. In *Proceedings of IUI 2000*. ACM Press, New York, 2000, 256-259.

A Multimodal Mobile Museum Guide for All

Carmen Santoro, Fabio Paternò, Giulia Ricci, Barbara Leporini
ISTI-CNR
Via G. Moruzzi 1
Pisa, Italy

{carmen.santoro, fabio.paterno, giulia.ricci, barbara.leporini}@isti.cnr.it

ABSTRACT

Museums represent a particularly suitable context in which to experiment with new interaction techniques for guiding visitors and improving their experience. This is mainly due to the large amount of digital information available, the technological resources more and more adopted in such environments, and the range of different visitors museums always receive. In this paper we present a preliminary work of a portable, multimodal museum guide also able to offer diverse accessibility options to users, so as to fit and support the needs and preferences of different users, including the visually impaired. Our work combines multiple modalities –gestures, location, graphical and voice. In particular, we will focus on how tilt gestures can be used to control and navigate the user interface in order to enhance the user experience, including the case when the users are the blind.

Categories and Subject Descriptors

H.5.M [Information interfaces and presentation].

General Terms

Design, Experimentation, Human Factors, Languages

Keywords

Mobile guides, Accessibility, Gesture interaction, Tilt interfaces, RFID, Museum guides, Multi-modal user interfaces.

1. INTRODUCTION

Recent technological advances (including increasing availability of various types of interactive devices, sensors and communication technology) together with the growing availability of digital information, enable novel interactive software environments to support users in several contexts for different objectives. However, such a wealth of both information and devices might become at the same time a potential source of disorientation for users, if not adequately supported, and such issues become further critical when disabled visitors are considered. The support for museum visits is traditionally limited to audio guides and interactive kiosks, which have several limitations in different ways from various viewpoints, then, it is important to exploit new technologies to identify new solutions able to enhance user experience.

In recent years, there has been growing interest in the development of mobile guides, thanks to the spread and increase in the performance of PDAs and mobile phones, the progress of wireless communications and improvements in localization

technologies. The main advantage of these applications is the possibility, by following the context-aware computing paradigm [3], to provide users with context-dependent services, exploiting information such as the user's location, time, nearby people and devices, current task, etc.

One of the objectives of a museum guide is improving the effectiveness of the users' museum experience. For visually impaired users, the guide should not only integrate what the disabled cannot get by their own (e.g. by vocally providing the visual information in the museum), but it is also important that the users can easily interact with the application. For the others, a virtual guide should support and integrate the physical visit, not being too intrusive and then substituting or obstructing the fruition of the experience while in the museum.

In this paper, we present a preliminary work about how to provide different levels of accessibility in a museum guide through the exploitation of different modalities (vocal and tilt) in order to enhance the user experience in a museum.



Figure 1. User in the Marble Museum with mobile guide.

The organization of the paper is as follows. First, we discuss related work in the area of mobile guides; next, we provide some background information on the previous version of our guide. Then, after having presented some main features of the tilting modality, we present the different accessibility options we provided to fit the needs of disabled users. Lastly, some concluding remarks and indications for future work are provided.

2. RELATED WORK

The museum domain has raised an increasing interest regarding the support that can be provided to visitors through mobile devices. One of the first works in this area was the Hippie system [12], which located users via an IR system with beacons installed at the entrance of each section and emitters installed on the artworks. The GUIDE project [4] addresses visitors in outdoor environments supported through several WLANs.

Research on gesture interaction for mobile devices includes various types of interactions: tilt, discrete gesture interaction and continuous gesture interaction [10]. In [6] a continuous, control theory-based approach for tilt-based interaction for mobile terminals equipped with 6DOF accelerometers is presented. In [8] a study discussing the effects of ergonomics in tilt interaction is presented. An accelerometer-based tilt interaction system for scrolling and determining screen orientation is presented in [9]. Tilt interaction to be utilised in navigating menus and scrolling documents and maps is described in [13]. However, we think that it needs to be augmented with other techniques in order to make museum visitors' interaction more complete and natural. To this end, we selected the use of accelerometers able to detect tilt events, allowing users to easily select specific information regarding the artworks.

Physical browsing [1] allows users to select information through physical objects and can be implemented through a variety of tag-based techniques (e.g. RFID tags). An example of an application using RFID tags to guide the disabled in a museum is described in [2]. In this work, authors implemented a location-aware tour guide based on RFID localisation, specifically developed to guide visually impaired people in an exhibition.

3. BACKGROUND

Our interactive environment for museum visitors has been applied to a previously existing application for mobile devices: Cicero [5]. In its first version, it was one digital museum guides developed for a PDA platform and freely available to the visitors of the museum of Marble located in Carrara (Italy). It provides visitors with a rich variety of multimedia (graphical, video, audio, ...) information regarding the available artworks and related items. This application is also location-aware. This is implemented through a number of infrared beacons located on the entrance of each museum room. Each of them is composed of several infrared emitters and generates an identifier that can be automatically detected by the application, which thus knows what room the user is entering (see Figure 2) and immediately activates the corresponding map and vocal comments. In addition to information regarding artworks, sections and the museum, the application is able to support some services such as showing the itinerary to get to a specific artwork from the current location. Most information is provided mainly vocally in order to allow visitors to freely look around and the visual interface is mainly used to show related videos, maps at different levels (museum, sections, rooms), and specific pieces of information.

A further extension was developed, based on a new interaction paradigm we called Scan & Tilt [11]. The concepts that were at the basis of its design came out from some considerations derived from previous experiences about users visiting a museum with a virtual digital guide. First of all, the fact that the virtual guide should be able to somehow directly interact with the available physical objects and it should not be intrusive on the user experience, by leaving the visual channel open to enjoy the

artwork. Indeed, visitors in a museum would not be interested in spending much time understanding how the electronic guide works, since they will probably not visit the museum again. Therefore, it is important to improve as much as possible the naturalness of the interaction supported by the guide. In addition, a good degree of freedom should be provided to users to move around and obtaining information only when they want.



Figure 2. Two screenshots of the Cicero guide.

To this end, we looked at exploiting RFID-based solutions. They are composed of two main parts: the set of tags, or radio transponders, and the tag reader, or transceiver. Tags basically have a static identification number (ID), but can also store different type of information such as sensed data (e.g., environmental temperature). The reader scans for available tags, and, depending on their features, may interrogate them for additional information stored on their embedded memory. To make our mobile guide as small and as light as possible, we opted for a totally handheld-based solution consisting of Compact Flash (CF) RFID reader with small-sized antenna. The PDA does not need any additional expansion or adapter because the reader plugs directly into the CF slot.

In order to fulfil the above requirements, we developed the scan and tilt paradigm, which combines multiple modalities – gestures, physical selection, location, graphical and voice. The PDA was equipped with an RFID reader able to read passive tags and a 2D accelerometer able to recognise small movements of the handheld device. Physical selection is obtained by scanning RFID tags associated with the artworks, and single handed tilt gestures are used to control and navigate the user interface and multimedia information. By pointing at the artwork of interest and controlling audio information with small single hand gestures, the visual channel is not too overloaded, resulting in a less intrusive interaction technique. We performed a first empirical evaluation on this prototype. The test showed an overall good acceptance among users but, at the same time, highlighted some limitations. Indeed, the passive RFID tags used in this prototype forced the users to stand in very close proximity to the artworks, which is not very natural in museum environments. Moreover, the visitors found the tilting interaction technique not so easy, especially in their first trials.

In the new version we overcome the problem of passive RFID tags by using active RFID tags, which can reach a larger range and at the same time provide a localisation of the current position of the user without forcing them to stand in close

proximity with the artwork. In addition, we planned to investigate whether the tilt modality might be better exploited to make the digital museum guide accessible to blind and/or sight-impaired users. In the next sections, we will provide further details about the different levels of accessibility that we achieved by exploiting the features of the tilt modality also in combination with the vocal modality.

4. TILT MODALITY

The gesture modality in our approach utilizes 2D acceleration sensor hardware from Ecertech. The sensor hardware is attached to an iPAQ PDA with Pocket PC operating system and can be used also in other Pocket PC PDAs and SmartPhones. The sensor produces signals that are interpreted as events (e.g.: TiltLeft, TiltRight, TiltBackward, TiltForward, ..) by the tilt manager-data processing module of the mobile device. The movements are detected by an accelerometer and, depending on the direction and speed of such movements they are translated into suitable actions/events onto the application (selection, navigation or activation) according to the user interface at hand.



Figure 3. The Museum Mobile Guide equipped with accelerometer

In general, the tilt interface follows a simple to learn pattern: horizontal tilts are used to navigate through different pieces of information at the same level or to start/stop some activity, vertical tilt down events are used to go down in the information hierarchy and access more detailed information, whereas vertical

tilt up events are used to get up in the information hierarchy. Since there are different levels of information supported (the museum, the thematic sections, the artworks, and the information associated to specific artworks and its rendering), when a specific artwork is accessed, it is still possible to navigate by horizontal tilting to access voice control (to decrease/increase the volume), control the associated video (start/stop), and access information regarding the author. The first version of the software prototype used a tilt monitoring algorithm based on static angle thresholds and taking into account the initial tilt angle of the device when the application starts. The tilt of both horizontal and vertical axes were measured every 1/10 second. These values are then compared to the original tilt measurement performed at application start-up time, and if a 15 degree threshold was exceeded for over 500ms in one of the axes, this was interpreted as the appropriate tilt gesture for that axis. At that time the possible tilt events considered were just 'forward', 'backward', 'left' or 'right'. In the new version of the prototype we have performed some improvements to the tilt algorithm, also introducing the possibility to handle further tilt events (for instance, also a *diagonal* tilt is considered).

5. ALLOWING DIFFERENT ACCESSIBILITY LEVELS

In order to enable different users to use the software application according to their needs and preferences, the new version of the application has been designed so as to support different levels of accessibility. In this context we mean with "accessibility" the set of graphical and vocal features that are provided to the user for interacting with the application.

We planned to support different levels of accessibility in our tool. Depending on the level selected, a different application configuration is provided to the users, who then interact differently with the software prototype.

- *Basic Navigation:* this option is aimed at allowing user to use the application through the classical buttons/pen of the PDA
- *Navigation with Audio Feedback:* in this option, which represents the minimal level of accessibility and can be considered helpful for both partially-sighted and blind users, the user's navigation is facilitated by a vocal feedback that is activated when necessary. For instance as soon as a button is selected, the corresponding label is vocally reproduced by means of a TTS (Text To Speech) engine. The Loquendo Embedded TTS (www.loquendo.com) has been used as vocal engine in our software prototype.
- *Navigation using Tilt:* in this modality only the tilt modality is active (the voice is not used). Therefore, there might be tilt movements to which no action is associated anymore. The navigation is carried out by using small inclination movements of the PDA, which substitute the use of PDA pen and buttons.
- *Navigation with Audio and Tilt:* in this option the tilt is used together with the voice. This type of navigation allows the users to use the museum guide and navigate within it using the tilting options and, at the same time,

to have a vocal feedback from the application. This configuration option is especially aimed at supporting the blind.

The accessibility level is selected just after the application starts (see Figure 4, left) by displaying a form allowing the user to select which is the accessibility level preferred by the user. However, depending on the current configuration of the device it may happen that not all the possibilities are really available, because e.g. the accelerometer is not connected to the PDA, or it is connected by it does not work properly, etc.. In this case, the application automatically detects the concerned situation and does not make available the corresponding options.

As soon as the accessibility level has been selected, the user accesses the proper application (see Figure 4, right). If in the first form the user selected an accessibility level including one (at least) between voice and tilt modality, the “Accessibility” link will be available, allowing to configure and use the tilting and voice options.

In the Main Menu there are four links for accessing the different information sections. They provide respectively general information, describing the meaning of the different icons used in the software application, information about the infrared emitters in the museum, and a link providing the possibility of selecting the specific level of accessibility to be provided. The user is supposed to visit all the sections in order to get the related information, otherwise a pop up message will alert about it.

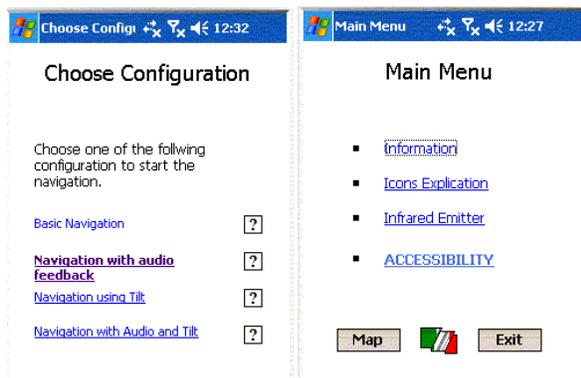


Figure 4. The Form for Selecting a Specific Type of Navigation (left) and the Main Menu of the guide (right)

5.1 Tilt and Voice

In this section we provide an overview of the tilting features when both tilt and voice are selected. Within the various forms, three main sections can be identified: the title, the “body” of the form and the toolbar. Using the vertical tilting it is possible to navigate through the different elements: depending on the element currently selected, further actions are available. For instance, if the title is selected, it is automatically rendered using the vocal modality; if the body is selected using a vertical tilt, further horizontal tilting possibilities are available for navigating within the body.

In addition, we judged useful to use a combined axis movement (i.e. diagonal tilt) to further improve the interaction richness between the user and the application. For instance, in the initial welcome form, a diagonal movement (right-up) allows the user to execute the action associated to the currently selected button. Conversely, a diagonal tilt movement (left-up) might be used to interrupt the vocal summary, as well as a left-down tilt movement might be used to change the language of the guide.

In the form showing the map of the museum (Map form), apart from the tilt functionalities already indicated before, more advanced tilting functionalities are provided, in order to interact more completely with the form:

- *Vertical tilt*: it is possible to move among the different subsections of a form. As soon as a new subsection is reached, it is vocally rendered to the user.
- *Horizontal tilt*: this movement is associated with several actions. Indeed, if the audio menu has been selected, with an horizontal tilting it is possible to control the audio volume, or moving within the map or the toolbar.
- *Diagonal tilt*: different options are provided depending on the movement selected. For instance, a *right-up* movement allow the users to execute an action; a *Left-up* movement interrupts the vocal summary or stops the automatic advance, while a *left-down* movement allows the user to change of the language of the guide.
- *Central PDA Button*: depending on the current state of the software prototype, it can be used in different ways. For instance, it works as Play/Pause button if a video is currently being played. Alternatively, it enables the user to deactivate/re-activate the tilting modality, also allowing to reset the initial coordinates.

An empirical testing session with real users has been already planned and it will be carried out in the next future. Through it, we hope to collect useful insights regarding the usability and the accessibility of the museum guide.

6. CONCLUSIONS AND FUTURE WORK

We have proposed a new interaction paradigm for mobile museum guides aiming at enabling more effective interactions through the combination of multiple modalities – gesture, location, graphical and voice.

Our solution for a mobile museum guide considerably extends interaction towards more natural ways of interacting with the environment, also enabling users with visual impairments to benefit from the guide. Future work is planned to consolidate and improve the prototype which has been described here in its preliminary stage.

7. REFERENCES

- [1] Ailisto, H., Plomp, J., Pohjanheimo, L., Strömmer E.: A Physical Selection Paradigm for Ubiquitous Computing. Proceedings EUSAI 2003: 372-383.
- [2] Bellotti, F., Berta, R., De Gloria, A., and Margarone, M., 2003, Supporting Efficient Design of Mobile HCI. Conference on Human Computer Interaction with Mobile

Devices and Services (MobileHCI), September 2003 (London: Springer), pp. 241-255.

- [3] Chen G., and D. Kotz: 2000, A survey of context-aware mobile computing research. Technical Report TR2000-381, Dept. of Computer Science, Dartmouth College, USA.
- [4] Cheverst K, Davies N, Mitchell K, Friday A, Efstratiou C (2000) Developing a context - aware electronic tourist guide: some issues and experiences. In: Proceedings of CHI 2000, ACM Press, The Hague, The Netherlands, pp 17-24.
- [5] Ciavarella, C., Paterno, F., The design of a handheld, location-aware guide for indoor environments Personal Ubiquitous Computing (2004) 8: 82-91.
- [6] Crossan, A., Murray-Smith, R., Variability in Wrist-Tilt Accelerometer-based Gesture Interfaces, Proc. of MobileHCI'04, Glasgow, September 13-16, 2004. LNCS 3160, Springer-Verlag, pp. 144-155, 2004.
- [7] Dini, R., Paternò, F., Santoro, C., An Environment to Support Multi-User Interaction and Cooperation for Improving Museum Visits through Games, Proceedings Mobile HCI 2007, 9-12 September 2007, To appear.
- [8] Eslambolchilar, P., Murray-Smith, R., Tilt-based Automatic Zooming and Scaling in Mobile Devices - a state-space implementation, Proc. of MobileHCI'04, Glasgow, September 13-16, 2004. LNCS 3160, Springer-Verlag, p120-131, 2004.
- [9] Hinckley, K., Pierce, J., Sinclair, M., Horvitz, E., Sensing Techniques for Mobile Interaction, ACM UIST 2000 CHI Letters 2 (2), pp. 91-100.
- [10] Mäntyjärvi, J., Kallio, S., Korpipää, P., Kela, J., Plomp, J., Gesture interaction for small handheld devices to support multimedia applications, In Journal of Mobile Multimedia, Rinton Press, Vol.1(2), pp. 92 - 112, 2005.
- [11] Mäntyjärvi, J., Paternò, F., Santoro, C. :2006, Incorporating Tilt-based Interaction in Multimodal User Interfaces for Mobile Devices, Proceedings TAMODIA 2006, LNCS Springer Verlag 4385, Hasselt, pp. 230-244.
- [12] Oppermann R, Specht M (2000) A context-sensitive nomadic exhibition guide. In: The proceedings of symposium on handheld and ubiquitous computing, LNCS 1927, Springer, pp 127-142.
- [13] Rekimoto J., Tilting operations for small screen interfaces, ACM UIST 1996, 1996., pp. 167-168.

Heritage & Habitus: Designing to Support Situated, Living Knowledge

Nicola J Bidwell
James Cook University
Discipline of IT (Cairns)
Australia
nic@it.jcu.edu.au

truna aka j.turner
Australasian CRC for Interaction Design Pty Ltd.
QUT (Brisbane)
truna@acid.net.au

Jason Holdsworth Colin Lemmon Michael Shay
James Cook University
Discipline of IT (Cairns)
Australia

ABSTRACT

We refer to an ongoing endeavour aimed to assist Indigenous communities in Australian in persisting their personal and cultural memories linked to temporally dynamic interactions *in situ*. The design enables Indigenous users to upload items they collect themselves (e.g. photographs, audio, video) using mobile phones, in their traditional lands into a topographical simulation; and, then to associate these items with their own hand-drawn markings in the simulation. The design responds to the rich interconnectedness between Indigenous culture and the land and the need to converge spatial information technologies with practices that are not, inherently, conditioned by the geometries of the West. We propose that the design approach contributes to thinking about ways that mobile guides can respond to multiple realities and corporeal and affective phenomena.

Categories and Subject Descriptors

J.5 [Arts and Humanities]; I.3.7 Three-Dimensional Graphics and Realism, I.6.8; Types of Simulation; H5.2[Information Interfaces & Presentation]:User Interfaces; K4.2[Computing Milieux]:Social Issues

General Terms

Design, Human Factors, Theory

Keywords

Representation, Mobile devices, cultural knowledge, Spatial Practices, Dialogic, Indexicality

1. INTRODUCTION

More Indigenous territory has been claimed by maps than by guns. This assertion has its corollary: more Indigenous territory can be reclaimed and defended by maps than by gun” Nietschmann 1995 [28]

In this paper we refer to an ongoing endeavour to address Australian Indigenous communities’ requirements to transmit their knowledge and memories embedded in their land. The design enables Indigenous users to upload items they collect themselves (e.g. photographs, audio, video), using mobile phones, in their traditional lands into a topographical simulation; and, then to as-

Mobile HCI 07, September 2007, Singapore.

Copyright 2007 ACM 1-58113-000-0/00/0004...\$5.00.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

sociate these items with their own hand-drawn markings in the simulation (Fig 1). We intend this venture to empower Indigenous people in persisting their personal and cultural memories linked to temporally dynamic interactions *in situ*. Given the rich culture of interconnectedness that has evolved for a people inhabiting a land for over two Ice Ages, we propose that the approach contributes ways that mobile guide design can respond to multiple realities more generally.

We proceed by outlining the need to better converge spatial information technologies with cultural practices that are not, inherently, conditioned by the geometries of the West. Then we summarise insights from various projects, with which we have been involved, that contribute conceptual appreciations informing our current interaction design. These include the achievements of a partnership to develop simulations of Indigenous traditional lands (e.g. [32][33][37]); our research on the role of natural features by people in cultural transmission to support a sense of belonging; and, our research into corporeal and affective dimensions in people’s experience while traversing natural places. We conclude by describing our progress in addressing these insights in our design.

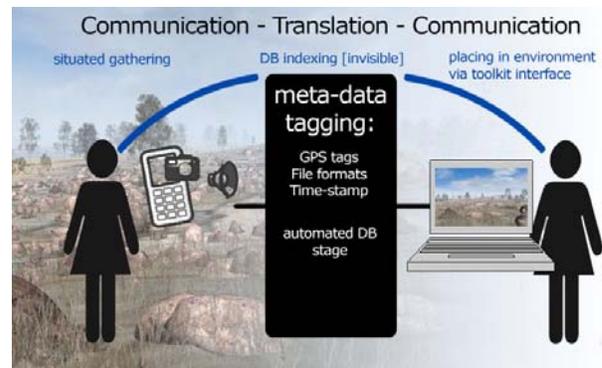


Fig. 1 Indigenous users upload items they collect *in situ* into the simulation associated with their own markings (see [36])

The way people represent local knowledge using spatial information technologies can reveal as much about their sociopolitical conditions as it can the meanings emerging from their own situated experience of places. Increased participation in Geographic Information Systems (GIS) enables diverse communities to interact with economic forces and governing authorities by describing the places in which they reside. In an increasing range of initiatives formerly marginalized people use Global Positioning Systems (GPS), data on the Internet, GIS, aerial photographs, satellite or remotely sensed imagery to promote their issues and concerns. For example, they support Australian Indigenous groups’ Native Title Claims to recover their nations’ traditional lands. Participatory GIS (PGIS) can integrate diverse forms of information to enable culturally-suited, peer-to-peer dialogue; however,

empowerment across socio-economic contexts, sectors and locations depends on interactions with certain representational conventions. So, usually geo-referencing and visualising two or three-dimensional maps are major conduits for participation.

Geo-spatial mapping in PGIS is a vital vehicle of advocacy for those formerly disenfranchised in decision-making but implicitly absorbs power relations. These uniform abstractions may help manage the indeterminacy of everyday experience of places in cultural practices for communication but they are neither neutral (e.g. [10][21]) nor compatible with situated usage (e.g. [7]). Rather, geo-spatial models are an ontological legacy of the enlightenment; for example, consider the bindings between the prime meridian of latitude and global exploration and empire. Contrast this with Amazonian Indians view where besides longitude there is "latitude, altitude, historical context, sacred sites, and spiritual or mythological sites, where invisible creatures mark watersheds and areas of high biodiversity as off-limits to exploitation" [30]. So patterns of adopting mapping applications portray only aspects of people's spatial relationships with places. This aspect, while important, allows the models of space underlying technologies to obscure specific cultural, historical and topological spatial "infrastructures" ([16]) which shape everyday experience and govern spatial praxis in the world. It also displaces people further (see: [6]) by constraining their technology mediated interactions with the world (e.g. [11])

The interdependency between physical terrain and "ways of knowing, being and doing" [e.g. 23] in Indigenous Australian culture sharply contrast with dominant spatial practices in design. Accounts of the way Indigenous (also known as Aboriginal) culture refers to physical terrain and landscape apply the term 'country' to describe a view of life invested with rich ecological and symbolic interconnections between people and places through many generations. Country is simultaneously lived in and is a system of living for physiological, social and spiritual nourishment. In literally living on the land Traditional Owner groups have accumulated a depth of knowledge about diverse aspects of the ecosystems they occupied to manage abundant resources over 40,000 years. The topological, cultural and historical infrastructures of country, through which an Indigenous person's spatial encounters have meaning, include relationships between flora, fauna, land, water, fire, meteorology and cosmology and interactions between their movements and cultural and personal memories embedded in the terrain. Thus, an Indigenous person has access to a socio-cultural and ecological memory that is embedded in the natural terrain [31]. GPS can record Indigenous people's movements in natural places; but its geo-spatiality flattens the very enlivenings felt in their experience and the meanings in the landscape that guide them as they live and exercise their Traditional Knowledge (TK).

2. FORMATIVE CONTEXTS

Our current venture is informed by several facets of our empirical and theoretical research in designing simulations representing natural places and mobile guides to support experiences in natural places. Here, we indicate conceptual issues shaping the design.

2.1 Digital Songlines Environments

Our endeavour emerged from the achievements of a partnership between Indigenous communities, Indigenous-owned Cyber-Dreaming [14] and the Indigenous Communities project stream of the Australasian CRC for Interaction Design (ACID). Over the

last 3 years the partnership has developed the Digital Songlines Environment (DSE) (e.g. [19], [23], [37]) a set of simulations of traditional lands which translate the schemata of Indigenous communities (e.g. [32],[33]). The DSE approach is named to reflect the complex web of symbolic motifs related to topographical features which Indigenous people use to organize and exercise their TK. The Indigenous conceptual landscape is criss-crossed with invisible storylines or songlines linking nations across the continent. These lines are inextricably linked to spirituality, which encompasses both individual and collective ritual responsibility for the land and practical access to and management of natural resources (e.g. a basic songline connects sources of water and sacred places). The DSEs respond to both the vulnerability of Indigenous culture, in the face of ongoing colonisation, and the "problematic disjunction between the structured information to be found on a computer, and the integrated, holistic, lived and performed knowledges of Aboriginal people on country" [12].

The DSEs use a commercial middle-ware game engine (GarageGames' Torque) to create inhabited topographical representations of the cultural landscape of individual Indigenous Australian communities. The simulation is constructed using Digital Elevation Models (DEM) and topographic surfaces generated from GPS co-ordinates collected in actual traditional homelands and represented via linear coordinates. Some liberties are taken with this data to create a sense of being in a world (e.g. edges are elevated to hide the game map's boundary and create a frame for the representation). The map is textured and populated with local flora based on photographic data and known ecology. The DSEs' basic topographic landscapes are elaborated with the flora and fauna remembered by Indigenous people who once lived in the area. For example, trees, bushes and grasses lace into the map and carefully modeled and animated fauna sometimes has artificial personality.

The DSEs represent traditional homelands [19] as a cultural heritage activity by recognising heterogeneity in cultural groups [37]. Each DSE iteration is developed through participatory relationships with the Traditional Owners of the area reconstructed who provide the visions, stories and other TK (e.g. bush food and medicine). This is important since Western approaches all too often form knowledge of 'Others' that are discontinuous with Indigenous historical contexts [28] and dichotomise TK systems [1] against those of the Canonical Western tradition. So the DSE approach represents the difference between tribal groups via place. For example, the Gunggari iteration aims to support a sense of 'walking the country' with Irene Ryder one of the few remaining speakers of the Gunggari language of regions in Queensland (the Maranoa, Warrego, Condamine and Balonne regions). Another, DSE recreates the landscape of pre-colonial Sydney harbour area in New South Wales.

To prioritise the communities' deeply cultural meanings and support a sense of being in a particular place the DSEs give prominence to country and stories. The stories told by Indigenous communities belong to the land and country has primacy in cultural transmission [31]. So the DSEs forefront land as a protagonist in spatial stories brought to life by cultural artifacts and activities. The environment is populated by different groups of people (women, men and children) engaging in appropriate activities and animals (e.g. eagles fly overhead, emus scatter in response to movement) and is further enriched with ambient audio (e.g. frogs chorus). In many instances, cultural information is embedded in the DSEs via pop-up text, video and audio, including recording as much of the community language words as survive (see [38]).

The DSEs significantly contribute to archiving aspects of TK and addressing the need for Indigenous communities to demonstrate to outsiders the meaning of their encounters with country. With the ongoing disappearance of Indigenous oral languages preserving cultural heritage is urgent and includes supporting communication between cultures over contested places (see: [32]). The DSEs communicate Indigenous cultural meanings to outsiders by exploiting Western emblematic spatiality and the interactional conventions of computer games. Although the DSEs conform with some concepts of space important to the communities (e.g. tracks left by animals and markings outlining people's camps that represent their interconnection) they use Western cartography rather than Australian Indigenous mapping traditions. The DSEs harness established game conventions to render topological data in linear perspective, where the world's geometry is inherently at the mercy of the interactant's position in the world. Some of the iterations also introduce other extrinsic views by using devices to support navigating the terrain which simultaneously inform the visitor that they are a passive "tourist" [9] able to affect the world only in the way designed. For example, a 'Heads Up Display' (HUD) imposes an interpretation of moving in the space in a way that is detached from the world; indeed it encouraged non-Indigenous focus groups to expect further game conventions (see [38]).

2.2 Dialogical Mechanisms of Belonging

We have been considering strategies to design simulations that support a sense of belonging by linking the places simulated with temporally evolving, individual and shared meanings. The term "habitus" refers to the bodily anchoring of culture, a person's daily practices and their sense of homeground [35]. Places become part of habitus via a person's continuous, ongoing proxemic interactions with those places. Certain locations become personally significant when they have some function in the routines of everyday life; have specific physical or social value; satisfy specific desires; or, are associated with meaningful events (In: [35]). Many of these proxemic interactions occur in an immediate, unreflective state of "being in the world" [18] which differs from states that can be accessed via Cartesian abstractions which separate the mind from the body and its surroundings. By privileging Cartesianism photo-real simulations coerce an interactant's relationship with a simulated place (see [39]). Their material properties as surrogate objects offer few, if any, of the minutiae of interactions by which people embody meaning in the physical world (e.g. [8]). They rarely allow interactants to shape landscapes, besides deforming environments with explosions, and tend to render interactants passive in physically "depositing" indicators of meaning in the terrain. For instance, interactant's footprints, if any, fade; the patten of multiple "feet" across terrains do not abrade paths; and, camp-fires leave no trace. This has provoked us to explore ways that simulations can harness mechanisms by which people share understandings about features in physical settings.

Data depicting people's egocentric experiences as they traverse natural places provides insights into processes by which meanings embedded in places contribute to habitus. We recorded data on visual, audio, and tangible interactions while people traversed natural terrain using a panoramic head-mounted video camera system [3]. This revealed a dialogical process [27] as people draw together the site's spatial resources, their past experiences and immediate interactions to render meaning. When people described what made the site personally significant the memories they mentioned (e.g. social relationships, activities and concerns; affects; aesthetics; motives for visiting; and, the rhythm of a visit) coupled

with features of the setting. They used features *in situ* as aide mémoire, to prompt meanings about places or similar features; as mnemonics to retain information in the flow of stories that articulated their experiences; and, indexically [21], to communicate with others without explicating salient referents. For example, seeing a rock besides a water-hole triggered a participant to recollect visits with her young children to collect tadpoles and to swim. We have found a dialogical approach [25] useful in understanding how meanings emerge from experiences and evolve and sustain constructs about places. One aspect of this is the tendency for people to interact at any instance with a place as if it is a finalized concept [25]. That is despite our own continuous, reciprocal shaping of past, current and future interactions with places we treat them in the moment as if we know everything that can be known about them. We maintain our sense of belonging via this finalisation fantasy by cross-referencing our immediate and recollected experiences to incorporate changes to spatial infrastructures.

Habitus is closely linked to people's sense of identity and tendency to form mutually supportive social groups [34]. The perspective that a person's identity is defined by their relationship with a place is acute for nations with ancestral inhabitant across thousands of years. For example, in focus sessions Indigenous communities were eager for the animated characters in the DSEs to reveal who they were in relation to the represented place.

The historical contexts of displacement in Australia and the efforts of Indigenous communities to pass on their TK even when they are displaced from their traditional homelands motivates us to develop ways to afford interactions with simulations of traditional lands that contribute to ongoing sense of habitus. Choices on which aspect of landscape are predicated by the cultural emphasis of the representation maker. Indigenous people depict sites by conveying their mythological, spiritual and/or ancestral relevance orally in stories and songs, visually in art, bodily in dance, and aurally in music (e.g. [22]). These forms help guide people's actions and interactions with the land. Rather than measuring, parceling or objectifying the land Indigenous symbolism refers to inter-connectedness (e.g. concentric circles form a design representing the interaction of eternal patterns of spiritual, ancestral and geographical significance). So a cluster of hand-prints under a boulder signifies a clan's sense of their personal and collective relationships and responsibilities for country rather than possession. Ancient meanings become embodied by people and, before colonisation, by the ecosystems in which they lived. For instance, Indigenous people interrelate the TK, contained within their stories and songs, to practicing traditional fire regimes and the influences of their fires on the vegetation and associated faunal population ecology (e.g. [40]). It is a tragic paradox that this, most concrete of inscription mediums, which reflects an equitable dialogue between human and nature, is also the most contentious.

2.3 Corporeal & Affective Orientations

Integrating Indigenous spatial praxis is essential to conserving Australia's unique biodiversity (e.g. [34]). The movements of Indigenous people in their country are a rich language conveying a 'Living Knowledge' *in situ*, tightly coupled with multi-sensory indicators of meaning in natural ecologies. To maintain their TK Indigenous people must continue to walk the land to affirm their songlines and storylines. This raises challenges for linking simulations with the dynamics of exercising knowledge on country, particularly given constraints on Indigenous people's access to their homelands and increasing urbanization.

Contemplating a person's direct contact with an environment can inform insights into how the body, mind and surroundings are integrated in situated ways of being, knowing and doing. The data we gathered using a panoramic head-mounted video camera system while non-Indigenous people traversed natural sites [3] revealed couplings between people's memories and bodily movement. When we analysed the dynamics of this engagement we noticed a proprioceptive connection with the land (e.g. [39]). To retain posture while walking participants spent an average of 25% of the time looking at the ground, depending on the unevenness of the terrain traversed. Such a connection is not as banal as might first seem. For example, when stepping into a muddy area along a dirt track a participant was pleased to discover the wet; later, recalling the muddiness and where the incident occurred provoked memories of his youth when, unlike recent years, the site had been wet and densely inhabited by waterfowl.

Corporeal interactions are not easily abstracted into software design, with its legacy of rationalism in elevating the mind over the body. Indeed, Coyne suggests: "*Paradoxically, even virtual-reality systems deny the importance of engaging the senses in the physical world*" [13]. Game-based visual simulations do not offer the minutiae of interactions by which meaning emerges during actions with the physical world and rarely demand interactants to minutely adapt their behaviours to the terrain (see: [5]). For example, avatars do not travel slower because they have mud on their boots or make detours because rain has changed the terrain. Indeed, the terrain rather than being an essential a partner in the process tends to be subservient to other functions (e.g. gameplay). Even when interactions are grounded in the physics of the terrain they emphasise the player's operations rather than empowering the represented landscape in creating meaning.

Representation of places, be that in story, song or simulation, become invested with corporeal meaning only when we relate them to situations in the world *in situ*. For example, one Gunggari community Elder said of a DSE: "*I can almost feel the dust between my toes*" because she associated the simulation with her own experience on country. However, for those who have yet to experience country as deeply as the Elder, who have not lived her life, understanding the terrain, traditional cultural forms and the DSE will develop reciprocally each absorbing meanings generated through the other. For example, consider the emergence of meaning from a wayfinding artifact such as a map to guide a cross-country run: "*Over time looking at the map evokes those [] bodily sensations, and running the terrain evokes that convergence of lines [on the map].... It is as if our sense of the situation and our sense of the map co-evolve.*" [28]. This means devising interactions with the simulation that enable appropriate meanings to emerge from, and be absorbed by, people's corporeal experience in the terrain.

Representation of places may also become invested with affective meaning by our relationship with the represented places *in situ*. Through our work with the Traditional Knowledge Retrieval Pathway [36] we have begun to appreciate the deeply affective connection between Indigenous people and their land and ecosystems. For example, a Kuku Thaypan Elder laments "*The place is now ruined Look how the magpie geese, the pelicans are all sick. It makes me sorry.*" [36]. We propose that this means devising interactions with the simulation that engage people's sense of felt-life "*at the level of their personhood*" [28] that emerges from, and is absorbed by, experiences *in situ*.

We have proposed that to support the range of couplings between a people, their setting and a representation of that setting we need to recognise a mosaic of information encounters. This is informed by situated field "experiments" on wayfinding (e.g. [2]) and reveals patterns of deliberate information seeking behaviour and serendipitous information discovery (e.g. [4]). We have used this to create two metaphors to inform the design of a mobile guide. "Daisies" are items that are salient to a sequence of landmarks along a route. "Berries" are items that are notable for other reasons, for example landmarks that are evocative of specific emotions or occur at points when corporeal sensations were noted (e.g. hunger, tiredness, heat). These metaphors are adaptable to designing for Indigenous conceptions of space and a mobile interface to the DSE. Daisies can be used to include items within an inherited songline (e.g. a chain of lakes, boulders and mountains) and berries can be used by an Indigenous adult to record items arising in their experience of the terrain as it exists today.

3. DESIGNING THE INTERFACES

Our current endeavour extends on the way the DSEs support a memory space for community members to relate a sense of habitus to the simulation and link the simulation to 'Living Knowledge' *in situ* in the land. Our design aims to empower users in Indigenous communities who have no technical or design expertise to control the evolution of their own DSE rather than to communicate cultural memory to outsiders. This enables us, as designers, to respect the rights of Indigenous-Australians to own and control their cultures and to maintain confidentiality about their personal and cultural knowledge and affairs. Sacred and secret material refers to information that is restricted under customary law (e.g. some information may only be learned or viewed by men or by women, or only after initiation). We seek to enable Indigenous people to transmit their personal and cultural knowledge memory through the DSE by linking markings they make on the digital landscape to items they have collected and, when these are collected *in situ*, correlate them with an appropriate location in the physical world.

Our iterative interaction design process focuses on users in Indigenous community groups, particularly in those situations where young people can no longer easily "walk the country" with Elder TK holders. Given the importance of maintaining TK by ensuring songlines and storylines are transferred to future generations in context, this connection between the physicality of country and the shape of cultural knowledge is critical. Here, we describe our progress to date in developing the technical components.

3.1 Graffiti: Interface & Database

We are in the process of "play-testing" the first prototype of our "Graffiti Engine" which enables Indigenous communities to make their own markers or mnemonics in a DSE and connect these with their own items or stories (e.g. metaphorical "berries"). For example, they can use representations of sharp stones to carve and organic dyes to paint trees and boulders and sticks to scratch sketches in the sand. Our design adds on to any existing DSE without requiring technical expertise in installing it. Ironically, we have modified the game-engine's default gun to create an invisible tool, that when the user is proximal to a drawable object, leaves a persistent trail of markings. The user can select from a wide range of terrain features (e.g. boulders, trees, sand) on which to draw by transferring a graphic overlay, or decal, onto the appropriate terrain surface and fading the decal to transparent (Fig. 2).

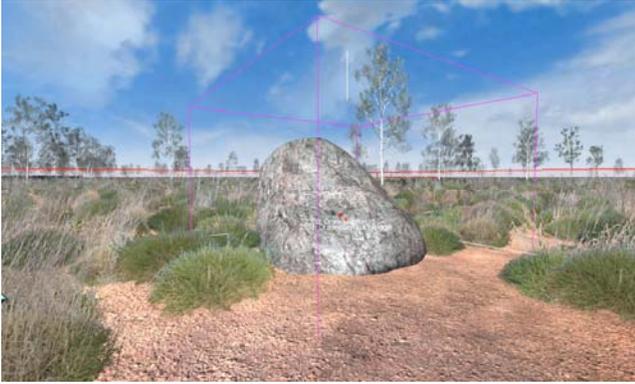


Fig. 2 Prototype drawable object within a DSE

The play-testing phase looks to the efficacy of the Graffiti concept in terms of treatment of virtual landscape as re-presentation of the real [10] where the purpose is not the construction of those second earths where the only limit is imagination. Rather, indications of success arise when the virtual is reconfigured by its resonance with the real, those moments of interaction with the landscape that are the intimate stuff of TK. Such intangible knowledge should remain intimate to its holders, the Graffiti engine endeavours to allow them to say: “Here, in this place.” and to then make further connection with other media items that they might deem important, exploiting a naturalistic interaction device in the same way that such ‘scratchings’ have given voice to those outside official histories since ancient times. This scratching mnemonic marker option might appear simplistic, pandering to the constraints of the real in an environment where such constraints have no requirement, however, it is essential to such oral knowledge that it is not open to individual construction of meaning that the second earth virtual worlds encourage [see: 38]. A recent event where a large company reconstructed an iconic area of Indigenous country as part of their second earth world offers an extraordinary example. The Traditional Owners raised objection and the iconic area within the second earth is now closed to trespass in the same manner as its real counterpart is carefully cared for.

The Graffiti Instances Database persists Indigenous users’ inscriptions to enable sharing between members of their community. It will involve database functionality to enable people to associate their own markings in the DSE with items they upload (e.g. photographs, audio, video) and store these relationships without interpretation by a technical expert. When “impacted” by the drawing tool (or on mouse release) the drawable object sends the 3D position of the decals to create a bitmap. This is applied to an invisible clone of the drawable object.

3.2 Mobile Mnemonics & Future Interface

The Mobile Mnemonics Interface enables Indigenous users to upload into the DSE items (“daisies” and “berries”) collected from places represented in the DSE. Our technical design is informed by experiences in creating location-based games using mobile phones [5]. We are in the process of testing mobile camera phones which we have programmed using Bluetooth and Java MIDlet technology to access GPS data, and store and catalogue photographs, using Java servlets, on a central server with their GPS positions. We will then start on functionality to pipe these items (e.g. photographs) from a mobile device via the Graffiti Engine into the DSE.

Our final intention is to enable information to be downloaded from the DSE onto mobile devices. This will enable an Indigenous person to download items relevant to a songline or storyline archived by a more senior community member (e.g. an Elder) to support persisting that motif *in situ*.

4. CONCLUSION

We believe that efforts to converge spatial information technologies with cultural practices that are not, inherently, shaped by the geometries of the West will be rewarded in diverse ways. Most importantly they can empower previously disenfranchised knowledge, such as that belonging to Indigenous Australia. However, designing interactions that facilitate this empowerment yields insights for designing mobile guides and simulations of place for non-indigenous peoples, particularly with respect to responding to the ephemeral phenomenon of embodied interactions and felt-life experiences in places beyond the city (see: [15]).

5. ACKNOWLEDGMENTS

We thank CyberDreaming and The Indigenous Communities Project for The Digital Songlines Project supported by ACID (the Australasian CRC for Interaction Design). See: We also express our respects to our Indigenous guides and collaborators

6. REFERENCES

- [1] Agrawal, A. (1995). Dismantling the Divide Between Indigenous and Scientific Knowledge. *Development & Change*, 26(3), 413-439.
- [2] Bidwell, NJ & Axup, J (2005) The Territory is the Map: Exploring the Use of Landmarks in Situ to Inform Mobile Guide Design *Proc. Xth Int. Conf. on Human-Computer Interaction (INTERACT05) IFIP & Springer-Verlag: Lecture Notes in Computer Science*
- [3] Bidwell, NJ & Browning D (2006) Making There: Methods to Uncover Egocentric Experience in a Dialogic of Natural Places *Proc. OzChi06, Sydney Australia (2006)*
- [4] Bidwell, NJ & Graham C (2005) Daisy-Chaining & Berry-Picking: Trajectory Oriented Design for Mobile Guides in the Wild *MobileHCI05 International Workshop on Mobile Guides*
- [5] Bidwell, NJ, & Holdsworth, J (2006) Battleship by Foot: Learning by Designing a Mixed Reality Game. *Proc. Joint International Conference on CyberGames & Interactive Entertainment, 2006 Perth Australia*
- [6] Bidwell, NJ, Radoll, P & truna aka j.turner (2007) Re-displacement by design. *<Interactions> March-April*
- [7] Brown, B. and Chalmers, M. (2003). Tourism and Mobile Technology. *Proc. ECSCW, 335-354, Kluwer.*
- [8] Browning, D, Stanley S, Friar M & Bidwell, NJ (2006) Emplacing Experience *Proc. of Joint International Conference on CyberGames & Interactive Entertainment, Perth 2007*
- [9] Casey, ES (1993). *Getting Back Into Place: Toward a Renewed Understanding of the Place-world.* Indiana University Pr
- [10] Casey, ES. (2002). *Representing Place: Landscape Painting and Maps.* University of Minnesota Press.

- [11] Chalmers, M. (2004). A Historical View of Context. *Comput. Supported Coop. Work*, 13(3-4), 223-247
- [12] Christie, M. (2004). Computer Databases and Aboriginal Knowledge. *Int. J. of Learning in Social Contexts*, 1, 4-12.
- [13] Coyne, R. (1995). *Designing Information Technology in the Postmodern Age: From Method to Metaphor*. MIT Press.
- [14] Cyberdreaming (<http://www.cyberdreaming.com.au/>),
- [15] Dourish P, & Bell, G (2004). Getting Out of the City: Meaning and Structure in Everyday Encounters with Space. Workshop on Ubiquitous Computing on the Urban Frontier. *Proc. Ubicomp 2004, Nottingham, UK*.
- [16] Dourish P, & Bell, G (2004) The Infrastructure of Experience and the Experience of Infrastructure: Meaning and Structure in Everyday Encounters with Space. *Environment & Planning B: Planning & Design*
- [17] Dourish, P. (2006). Re-space-ing Place:" Place" and" Space" Ten Years on. *Proc. ACM Conf. Computer-Supported Cooperative Work CSCW 2006 Banff, Alberta (2006)*
- [18] Dreyfus HL (1991). *Being-in-the-World: A Commentary on Heidegger's Being and Time* MIT Press
- [19] Gard, MS., & Bucolo, MS., (2005) Capturing Australian Indigenous Perception of the Landscape: Virtual environments with cultural meanings. *Proc. 11th International Conference on Virtual Systems and Multimedia, Ghent, Belgium*,
- [20] Garfinkel, H (1967) *Studies in Ethnomethodology, Englewood Cliffs: Prentice-Hall*
- [21] Harley, J. B. (1988). Maps, Knowledge, and Power. *The Iconography of Landscape*, 277-312.
- [22] Isaacs, J., & Morrison, R. (1984). *Australia's Living Heritage: Arts of the Dreaming*. JB Books.
- [23] Leavy, B., Barker, C., Hills, J. and Carrol, J. (2005) Digital Songlines - Digitising the Arts, Culture and Heritage Landscape of Aboriginal Australia *Proc. 11th International Conference on Virtual Systems and Multimedia, Ghent, Belgium*,
- [24] Martin K-BM Ways of knowing, being and doing: a theoretical framework and methods for Indigenous and indigenist research (1). (Reconciling Knowledges). *J. of Australian Studies (2003) p203(15)*
- [25] McCarthy J & Wright, P. (2005) Putting 'felt-life' at the centre of human-computer interaction (HCI)," *Cogn Tech Work*, vol.7, pp. 262-271,
- [26] McCarthy J & Wright, P (2005). Technology in Place: Dialogics of Technology, Place and Self. *Proc. INTERACT 2005 Rome, Italy*
- [27] McCarthy J & Wright, P (2006). Dialogical Approach to Experience: Uncovering Critical Potential. *Proc. The Virtual Conference M3: Man Medium Machine Research Platform*.
- [28] Nakata, M. (2002). Indigenous Knowledge and the Cultural Interface: Underlying Issues At the Intersection of Knowledge and Information Systems. *IFLA journal*, 28(5-6), 218-221.
- [29] Nietschmann, B. (1995). Defending the Miskito Reefs With Maps and GPS: Mapping With Sail, Scuba, and Satellite. *Cultural Survival Quarterly*, 18(4), 34-37.
- [30] Plotkin (2006). Cited in Butler, R. 2006, *Amazon Conservation Team Puts Indians on Google Earth to Save the Amazon*. *Mongabay News*
- [31] Rose, DB. (1996). Nourishing Terrains: Australian Aboriginal Views of Landscape and Wilderness. *Australian Heritage Commission*.
- [32] <http://songlines.interactiondesign.com.au/>
- [33] http://songlines.interactiondesign.com.au/media_gallery.html
- [34] Steffensen V (2004) Traditional knowledge recording project case study: Kuku Thyapan fire management practices and aspirations. In: *Bushfire in a changing environment – New Directions in management*. Sydney University, Australia
- [35] Thwaites, K (2001) Experiential Landscape Place: an exploration of space and experience in neighbourhood landscape architecture *Landscape Research, Vol. 26, No. 3, 245-255*
- [36] Traditional Knowledge Revival Pathway "Water We Know" Available from: http://tkrp.com.au/index.php?option=com_content&task=blogcategory&id=31&Itemid=47
- [37] truna aka j.turner & Adkins, B (2006) Mobile Mnemonics, *ACID Quarterly Review Research Documentations*
- [38] truna aka j.turner (2006) Destination Space: Experiential Spatiality and Stories. *Proc. Joint International Conference on CyberGames & Interactive Entertainment 2006 Perth Australia*
- [39] truna aka j.turner, Browning D & Bidwell NJ (2007) Wanderer Beyond Gameworlds. *Proc. of Digital Arts & Culture DAC2007. In press*.
- [40] Yibarbuk D, Whitehead PJ, Russell-Smith J, Jackson D, Godjuwa C, Fisher A, Cooke P, Choquenot D, Bowman DMJS (2001) Fire ecology and Indigenous land management in central Arnhem Land, northern Australia: a tradition of ecosystem management. *J. of Biogeography* 28:325-343

Traffic Advisories Based on Route Prediction

Kari Torkkola, Keshu Zhang, Haifeng Li, Harry Zhang, Christopher Schreiner, Mike Gardner

Motorola, Intelligent Systems Lab, Tempe, AZ, USA

Phone: (602)-438-3467

harryzhang@motorola.com

ABSTRACT

Traffic advisories can be delivered to assist travelers in avoiding congested areas and reaching their destinations in a timely manner. Instead of relying on a user to manually enter travel routes and times, we propose a prototype application that learns from the user's travel history and automatically predicts the user's destinations and routes. Traffic flow information is then accessed to determine whether congestions will occur along the predicted routes to warrant the delivery of a traffic advisory alert. User acceptance and trust are evaluated in a field study.

Categories and Subject Descriptors

H.5.2 User Interfaces: *User-centered design*

General Terms

Algorithms, Design, Human Factors.

Keywords

Context aware, user modeling, mobile assistance.

1. INTRODUCTION

Highway congestion has become a way of life in many parts of the world. This problem stems from several factors. First, the limited highway capacity simply cannot keep up with the population growth in many metropolitan areas. In U.S.A., driving is still the dominant method of transportation, despite the availability of public transportation. Second, urban sprawling has resulted in a rush hour nightmare when driving to metropolitan areas from the suburban areas during the morning commute and when driving in an opposite direction during the afternoon commute. The use of the high occupancy vehicle (HOV) lane can mitigate the congestion problem, but carpooling is not always available or convenient to everyone. Finally, highway constructions often lead to lane restrictions, and highway crashes can bring the traffic to a standstill. In U.S.A., there are over 6.4 million police-reported highway crashes every year.

Several commercial companies provide traffic flow information that is gathered by the department of transportation at the state level, the local police department, and their own helicopter fleet. Roadway sensors and cameras have been the prevalent method of acquiring the traffic flow information. The traffic flow information is made available to the public via radio or TV broadcast, Internet portals (e.g., www.traffic.com, maps.yahoo.com, maps.google.com), RSS feeds, email messages, or SMS messages on mobile phones. Radio or TV broadcast appeals to a large-scale audience, whereas a user can manually set up a traffic monitoring service on Internet portals that will send a custom report (via an email, an SMS message, or an RSS feed)

when the traffic congestion exceeds a threshold for a particular route during a time window.

A traffic report that is customized for a user is desirable because it tends to provide information that is more pertinent to the user. However, the initial manual set-up using Internet portals is tedious. If travel routes and times are changed, they will need to be modified manually on the Internet portals. The manual set-up and modification process is difficult to manage when the number of travel routes and times grows to accommodate mobile lifestyles. This process can be simplified or largely bypassed by using an automatic algorithm that predicts the routes that a user will take at a particular time. This paper describes one such algorithm that learns from a user's travel history, predicts likely routes and destinations using context information such as the current location and time, determines traffic conditions along the predicted routes, and delivers timely advisories and alerts to the user when there are unexpected slowdowns.

Krumm and Horvitz and their colleagues [3,4] have attempted to predict a user's routes from the user travel history. We adopt a similar approach to learn a user's usual locations and routes from the usage history and predict likely routes based on the location and time information. Unlike Krumm and Horvitz, we do not attempt to handle destinations that the user has not previously traveled to. The resulting traffic advisory system is a functional prototype which uses time and location information in conjunction with learned patterns of user behavior to determine when a user is leaving a location for a learned destination. The application reviews past routes that the user has taken to traverse from the current location to the predicted destination and checks available traffic monitoring services to determine if there are any unusual alerts.

2. DATA COLLECTION

Location and route data was collected from fourteen participants, representing 6 undergraduate students, 5 office employees, and 3 independents (stay-at-home parents or businessmen such as realtors who did not work consistently in an office environment). Each participant was provided with a commercially available mobile phone equipped with Bluetooth, along with a separate Bluetooth GPS receiver to acquire location data. Data logging software was installed on the mobile phone. Each participant used the phone and the GPS receiver for a two to three-month period. Context data logging software is "ContextPhone", public domain software from University of Helsinki, Finland [6]. The logged context variables consist of the following:

- GPS data from a Bluetooth GPS receiver,
- Current GSM cell ID,
- Bluetooth devices around the phone,
- Phone profile,

- Active phone application,
- Phone idle/active time,
- Battery status and charger status,
- Incoming/Outgoing calls, SMS or multimedia-messages,
- User interaction with the Phonebook and recent call log,
- Media captured with the device (e.g., photo, video).

The data is uploaded over the air to a web server, transparent to the participants. Logged data consists of events, a time stamp followed by the event. GPS and Bluetooth environment are scanned every 30 seconds. Other variables are logged as their status changes.

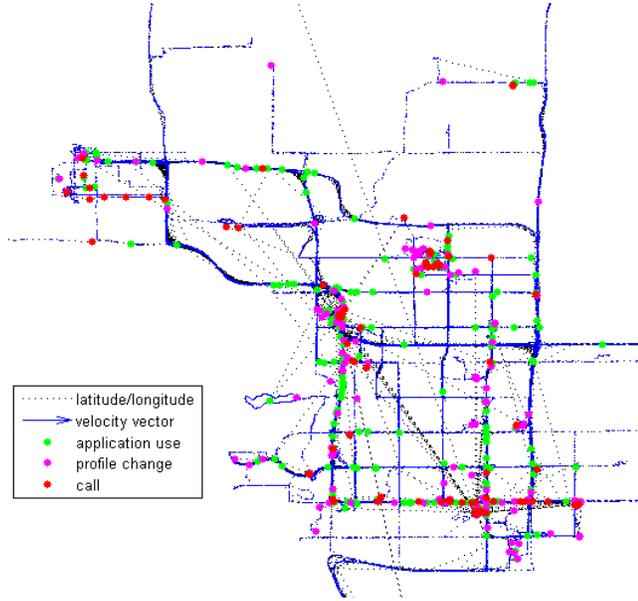


Figure 1. GPS coordinates are plotted together with selected phone application events for a typical participant.

As our focus in this paper is on a user's locations and routes, GPS coordinates can be plotted together with the use of phone applications to visualize the roads to which the participant has traveled and where the mobile phone is used (Figure 1). Figure 1 is cropped to cover only the local metropolitan area. It indicates that there are a small number of seemingly important locations where the phone has been used. It also indicates that the user's daily commute routes are regular and predictable.

3. ADAPTIVE LOCATION LEARNING

3.1 Adaptation of location clusters

Having alluded to the existence of regularities in the user behavior data, we describe our experimentation in discovering those regularities. Based on the collected context data, we employ machine learning techniques to determine the important locations where the user spends most of his/her time or most often is engaged in some activities. Location learning is mainly based on the collected GPS data. The GPS signal may be uncertain, however, when the user enters a building, a concrete canyon in an urban area, or when the GPS device is powered off [7]. To mitigate this problem, we use cell IDs as a secondary source to resample GPS data when GPS data loss occurs. The void GPS data is made up by repeating the last valid GPS reading, if the

area in which a GPS signal disappeared and reappeared, is covered by the same group of cell IDs.

Although a batch (non-adaptive) clustering method can automatically identify important locations, it is computationally expensive and requires processing available data as a whole [3,4]. It cannot adjust to changes in a user's lifestyle pattern. In order to track changes, adaptive (or incremental/sequential) algorithms are adopted. An adaptive algorithm is trained with a small amount of data initially, and adapts continuously as new data is entered. There are several advantages of using adaptive algorithms over batch algorithms. Adaptive algorithms are computationally efficient. They require less storage because a large set of previously used training data do not need to be kept in memory. The algorithm parameters can vary continuously to accommodate lifestyle changes.

For our traffic advisory system, a Gaussian mixture model is used to identify important locations. Assume for the moment that the true (but unknown) density of the GPS value g_t is of the form

$$p(g_t; \theta_t) = \sum_{\ell=1}^{N_t} \varpi_t^\ell \phi(g_t; \mu_t^\ell, \Sigma_t^\ell)$$

Where the number of components $N_t < \infty$ is known, parameter vector is $\theta_t = \{\varpi_t^1, \mu_t^1, \Sigma_t^1, \dots, \varpi_t^{N_t}, \mu_t^{N_t}, \Sigma_t^{N_t}\}$, the nonnegative mixing coefficients ϖ_t^ℓ sum to unity, and $\phi(g_t; \mu_t^\ell, \Sigma_t^\ell)$ is the normal probability density function with 2 dimensional mean μ_t^ℓ and standard deviation Σ_t^ℓ . The goal is to estimate the parameter vector θ_t , which consists of $3N_t$ components.

At time t , the number of Gaussians N_t in the mixture model represents the number of significant location candidates, the weight of Gaussians ϖ_t^ℓ denotes the significance of the learned locations, and the mean μ_t^ℓ and variance value Σ_t^ℓ in each Gaussian specifies the important location center and size.

In order to make model parameters adaptive, Titterton [8] has provided a recursive update formula based on maximum likelihood estimation. With Priebe's adaptive mixture estimator [5], we have designed an adaptive location clustering method which includes the update rule, creation rule, pruning rule and a model group switch rule. The details of the adaptation procedure will be described in a forthcoming paper. The adaptive location clustering method can recursively adjust not only the location parameters, such as the center and size, but also the number of location clusters needed to fit the data. This adaptive mixtures approach does not require intensive computational power and storage capacity and produces a consistent small sample estimator.

3.2 Evaluation of adaptive location learning

To evaluate the location learning feature in the traffic advisory system, we compare our adaptive location learning algorithm to a batch learning method. As the batch method we employ the DBSCAN (density-based clustering) algorithm, because it can automatically find the number of clusters, identify outliers, work well for arbitrary-shaped clusters and efficiently for large datasets [1]. We first evaluate the algorithms on one office employee who worked in regular business days at a fixed location. Both batch and adaptive algorithms are applied to the re-sampled GPS data,

the comparison results for important location discovery are generated for the first day, the first week, and the first month. Figure 2 presents the visualization results of the discovered significant locations after one week.

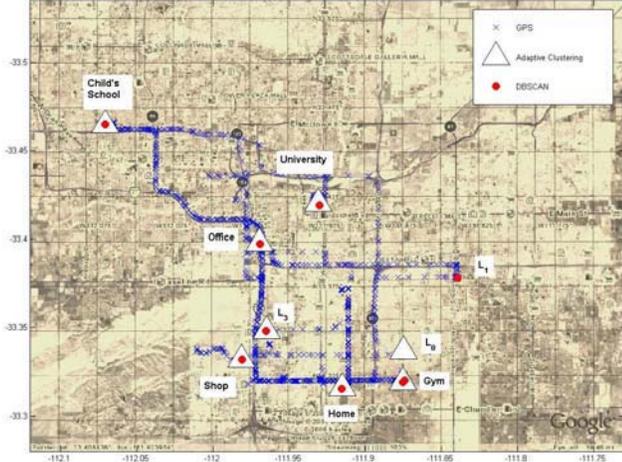


Figure 2. Location identification after one week

After the first day, the discovered locations by both batch and adaptive methods match very well. After the first week, compared with the batch method, the adaptive method discovered location L0 instead of L1 (Figure 2). Since the adaptive method adjusts the parameters online, location L1 which was visited early in the week is already discarded from the significant location candidates due to the short visit time and low visit frequency (only one visit for that week). Instead, location L0 which has been visited most recently had higher initial weight and then becomes more important than location L1. Meanwhile, both algorithms can discover several nearby locations in certain large areas, i.e. the “University” campus and the “Gym”. This comparison shows that the proposed location learning algorithm is effective in location identification and can adapt to the user’s daily life.

4. ROUTE LEARNING AND PREDICTION

After learning the important locations, the next step is to learn routes between a pair of locations. Currently, most transportation route learning methods are based on geographic information systems (GIS). Because our system runs on cell phones with limited computing and storage resources, we instead use raw GPS data sequences to represent routes [3,4]. A GPS string between two locations is denoted as a route. The starting point and ending point of the route involves detecting of the user leaving one important location boundary and entering into another important location boundary.

Because travelers often take different routes from location A to location B, GPS data strings between these two locations may not represent one single route. In order to determine if two GPS data strings represent the same route, the data strings must be compared. If the difference is small (with a MinMax criterion), then the two GPS strings belong to the same route. Here is the mathematical formula. Suppose two GPS data strings are

$$S_1 = \{g_1^1, g_2^1, \dots, g_n^1\} \text{ and } S_2 = \{g_1^2, g_2^2, \dots, g_m^2\}$$

The best match of S_1 in terms of S_2 is

$$\hat{S}_1 = \{\hat{g}_1^1, \hat{g}_2^1, \dots, \hat{g}_n^1\}, \text{ where } \hat{g}_i^1 = \arg \min_{g_j^2} (\|g_j^2 - g_i^1\|).$$

Similarly, $\hat{S}_2 = \{\hat{g}_1^2, \hat{g}_2^2, \dots, \hat{g}_m^2\}$ is the best match of S_1 in terms of S_2 , where $\hat{g}_j^2 = \arg \min_{g_i^1} (\|g_i^1 - g_j^2\|)$.

Two GPS strings are declared to belong to the same route if

$$\max(\|g_i^1 - \hat{g}_i^1\|) < d_1 \text{ and } \max(\|g_j^2 - \hat{g}_j^2\|) < d_2,$$

where d_1 and d_2 are the largest sampling intervals of R1 and R2, respectively.

With the adaptive location clustering, the significant locations are gradually discovered and updated. Once an important location pair is identified, the commute routes between them are recorded and learned. Using the preceding comparison process, unique routes are identified and saved in the database. The unique routes are non-directional. That is to say, the same route may be traveled from location A to location B, or from location B to location A. Frequencies of the unique routes are tallied in an hourly basis. After two month learning, 47 unique routes are obtained for one user. The route frequency is derived by considering all different routes starting from Location A to Location B. Table 1 presents the frequency of routes from "home" to "office" for one user. Note that routes 4, 8, and 26 are the most frequently traveled routes.

Table 1. Frequency of travel routes from home to office

Route ID	4	8	25	26	27	28	29	30
Frequency	0.6	0.1429	0.0286	0.1143	0.0286	0.0286	0.0286	0.0286

5. ROUTE PREDICTION-BASED TRAFFIC ADVISORY SYSTEM

Once routes are predicted, a traffic advisory system can be built. We designed the system to ensure that the user experience was unobtrusive and well accepted by the users. Ideally, the system is designed to assist users during transition times – as the user is transitioning out of one location and likely into a vehicle environment. Previous research has shown that messages from context aware devices are better received during transitions [2], and in our case it is during transitions that the information provided by the application is of the greatest relevance.

The application is designed to automatically detect when the user is leaving a location and predict to where they are likely headed, and the most likely routes the user will take. When this occurs, a pop-up window will appear on the visual display informing the user that that the application has detected the transition and is searching for traffic information. This will usually occur before driving. Soft keys will be allocated to allow the user to edit or cancel this function. No audio or haptic feedback will be provided to alert the user in case the user is not interacting with the mobile device during this time. To adopt a user-centric approach and minimize the obtrusiveness, traffic history may be studied to derive traffic averages and variances and traffic alerts are only provided when current or predicted traffic flow exceeds the average by a certain margin. If a traffic alert is detected that is out of the ordinary for that route and time of day, then the user will receive an audio or haptic alert, depending upon the phone’s current profile settings. In addition, a pop up window such as

Figure 3 will appear alerting the user that there is an alert, and offering the user the choice of ignoring the alert or obtaining more information. If the user selects the latter, more detailed information on the alert (e.g. type of incident and location) will be provided with an option for the application to recommend a new route to the predicted location. Because of driver distraction concerns, alerts are issued only when it is necessary to avoid traffic delays.



Figure 3. Screen shots of the application. The left side depicts a textual pop-up alert, and the right side a map view of the alert.

The developed prototype is a client-server system. The client software is running on a cell phone. The client collects GPS data and cell IDs and learns semantic locations and routes between them with our adaptive algorithm. An easy-to-use user interface is developed to alert users about heavy traffic situations. The server side includes a set of machines to provide map, real-time traffic, and other information to clients based on Java EE. The communication between clients and the server side is based on standard HTTP protocol.

All specific aspects of the implementation of the design of the traffic advisory system, as well as the overall acceptance of the application are validated by real users from specific demographic areas of interest in a large-scale field evaluation of the application.

6. CONCLUSION

Context knowledge such as the user's needs, locations, and travel routes is a major enabler for providing the user with unobtrusive assistance for mobile device. In this paper, the focus has been on a location-based traffic advisory system that is based on learning and predicting the user's patterns using data available from the mobile devices. We used a data-driven approach in the

development of the traffic advisory system. We collected usage data from multiple users on phones equipped with sensors over an extended period of time, mined the data to identify various lifestyle patterns across different forms of context, and attempted to predict the patterns. The predicted patterns were used to design applications that satisfy user needs.

We have demonstrated a successful implementation of a traffic condition assistant that learns user's important locations, routes between those, and learns to predict those routes. In order to maximize the system effectiveness and user acceptance, advisories are provided to the user only when there is a real need (e.g., a traffic problem on the predicted route). We envision that context aware features like this including a user modeling component will become essential in supporting new applications and concepts that connect consumers with media and services through multiple devices in the office, car, or home.

7. REFERENCES

- [1]. Ester, M. et al. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. KDD, 1996.
- [2]. Ho, J., Intille, S. Using Context-Aware Computing to Reduce the Perceived Burden of Interruptions from Mobile Devices. CHI 2005. April 2-7, 2005. Portland, OR.
- [3]. Krumm, J., Horvitz, E., Predestination: Inferring Destinations from Partial Trajectories. UbiComp 2006: The Eighth International Conference on Ubiquitous Computing, September 17-21 (2006), Orange County, CA.
- [4]. Letchner, J., Krumm, J., Horvitz, E. Trip Router with Individualized Preferences (TRIP): Incorporating Personalization into Route Planning. Eighteenth Conference on Innovative Applications of Artificial Intelligence (IAAI-06), July 2006.
- [5]. Priebe, C. E. Adaptive mixtures. *Journal of American Statistical Association*, 1994, pp. 796-806.
- [6]. Raento, M., Oulasvirta, A., Petit, R., Toivonen, H., ContextPhone: A prototyping platform for context-aware mobile applications. *IEEE Pervasive Computing* 4:51-59, 2005.
- [7]. Rukzio, E., Hamard, J., Noda, C., De Luca, A. Visualization of uncertainty in context aware mobile applications. *MobileHCI'06*, September 12-15, 2006, Helsinki, Finland.
- [8]. Titterton, D. M. Recursive parameter estimation using incomplete data. *Journal of Royal Statistical Society*, 1984, pp. 257-267.

A Gaze-based Guidance System based on a Real-world 3D Photo Logging System

Rieko Kadobayashi
National Institute of Information and Communications Technology
3-5 Hikaridai, Seika-cho, Soraku-gun
Kyoto 619-0289, Japan
rieiko@nict.go.jp

ABSTRACT

To achieve easy access to the information that users need in the real world, we are developing a gaze-based guidance system by extending the 3D photo-logging system we have previously developed for recording the real world. The prototype system consists of several servers running on a single machine and user terminals. The user terminals are equipped with mobile phones to which motion sensors are attached, a GPS sensor, and a notebook PC for logging the position and orientation of the mobile phone. Collecting photographs with 3D viewpoint information (based on the photographer's position and direction of gaze) by using the 3D photo-logging system, we can use these photographs and the text explanations associated with them as guide information. Users can access the information simply by pointing and clicking their mobile phones in the real world.

Categories and Subject Descriptors

H.4.3 [Information Systems Applications]: Communications Applications—*Information browsers*;
H.5.2 [Information Interfaces and Presentation]: User Interfaces—*Interaction styles, Prototyping*

General Terms

Design, Experimentation, Human Factor

Keywords

Gaze-based services, Navigation, mobile applications, photo browsing

1. INTRODUCTION

It is worthwhile to use a person's direction of gaze as well as his or her position when he or she is navigating in the real world because navigation depends on the direction in which the person wants to go and on what kind of information the user needs most. On the basis of this idea, we are developing a gaze-based guidance system for navigating in the real world by extending the 3D photo-logging system [6] we have previously developed for recording the real world with photographs.

Our 3D photo-logging system uses viewpoint-based image retrieval [5]. The key idea is to use 3D viewpoint information [7] as metadata when taking pictures in the

real world to facilitate easy access to pictures of interest. The position and orientation of the camera are regarded as a user's point of view in a 3D space and are referred to as 3D viewpoint information, which is a key element for bridging 2D images and 3D spaces (e.g., 3D computer graphics models, 3D virtual worlds, or the real world).

3D Viewpoint Information is basically a 3D vector from the observer's eye to the point where the observer is looking at. When it is not easy to obtain the 3D coordinates of the point correctly, we can use a direction vector, which starts the observer's eye to the direction the observer is facing, as the 3D Viewpoint Information. The direction vector can be calculated from the extrinsic parameter, i.e., the position and orientation, of the camera which the observer is using.

The most important feature of the 3D viewpoint information is that it enables users to search pictures intuitively without the need to specify key words [7]. Keyword-based searches require users to formulate appropriate queries; however, combining several key words is often insufficient. Users may experience difficulty combining several keywords to formulate a query that can distinguish between images that are very similar. Moreover, keyword-based searches require users to have some knowledge about the image content, so they can make an appropriate query. This is too restrictive for practical use.

Another retrieval method called "content-based image retrieval" [4, 8], uses features, such as color, shape, and texture, that are automatically extracted from images. However, photographs taken in the real world, especially those taken outdoors, are usually affected by natural conditions, such as sunlight and weather, and artificial conditions, such as traffic and crowds. This causes diversity in the image features of the same object under various conditions over time and hence reduces the efficiency of image retrieval. Viewpoint-based searches, on the other hand, are robust against diversity in image features.

Moreover, 3D viewpoint information is useful for organizing and analyzing data associated with spatial data. It is has many applications, such as archaeological information systems [3]. The 3D photo-logging system was designed to take full advantage of 3D viewpoint information. A prototype of our gaze-based guidance system has been developed based on the 3D photo-logging system.

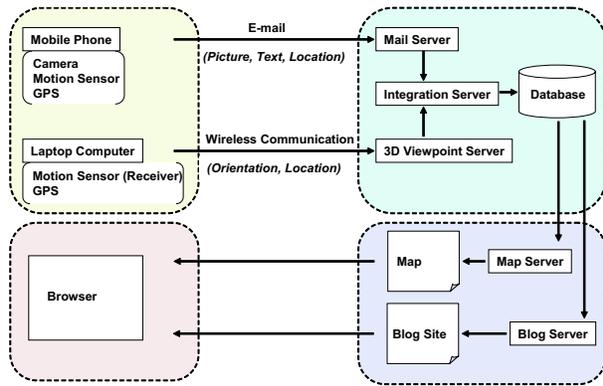


Figure 1: System configuration.

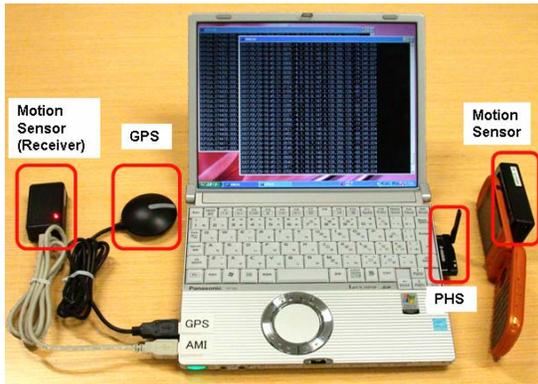


Figure 2: Hardware setup for user terminal.

2. 3D PHOTO LOGGING

2.1 Overview

We use the term “photo logging” to express the method of recording objects and events in the real world as photographs. 3D photo logging means photo logging with 3D viewpoint information. The 3D viewpoint information is composed of a person’s position and direction of view when he or she takes a photograph. His or her position is obtained from a global positioning satellite (GPS) while his or her direction of view is obtained from the motion sensor.

The goal of building the 3D photo logging system is to create a framework for recording, organizing, analyzing, searching, and presenting every object and event in the real world with “portable devices,” such as mobile phones, in an intuitive and casual manner. Objects include, for example, buildings, monuments, landscapes, and archaeological sites, and events include, for example, festivals. Anything that exists or happens at a particular location can be a target of the 3D photo-logging system.

2.2 System Configuration

Our prototype system consists of a server machine and user terminals. A database server, a mail server, and an



Figure 3: Experimental use of 3D photo-logging system.

integration server are run on the server machine. Figure 1 shows the configuration of the 3D Photo Logging system.

The user terminals are equipped with mobile phones to which motion sensors are attached, a GPS sensor, and a notebook PC for logging the position and orientation of the mobile phone, as shown in Figure 2. The motion sensor obtains the orientation of the mobile phone when the user takes a photograph. The motion sensor data is transmitted by wireless connection to a receiver, which is connected to the laptop PC via a universal serial bus (USB) port. The GPS sensor obtains the location of the mobile phone. Figure 3 shows our trial use of the 3D photo-logging system at a tourist site.

To obtain the posture of the user from his or her mobile phone, we used AMI601-CG [1] developed by the Aichi Micro Intelligent Corporation. AMI601-CG is an evaluation kit of AMI 601, a six-axis G2 motion sensor, which has a three-axis magnetic sensor and a three-axis accelerometer [2]. We can calculate the direction of the person’s gaze from data about his or her posture obtained with the mobile phone.

2.3 Processing Flow

When a person takes a photograph with a mobile phone, he or she sends the photo to the mail server via e-mail accompanied with a text explanation. Usually he or she writes the name or identifier in the “subject” field of the e-mail and writes some comments in the body of the e-mail so that these text explanations can be transformed into the title and description of a blog entry. This increases the readability of the blog.

The motion sensor attached to the mobile phone detects the orientation of the mobile phone every second and sends the information with a timestamp to the laptop PC by a wireless link. The location of the phone is detected by the GPS sensor connected to the PC through a USB port. Note that the position of the mobile phone can be obtained by the GPS sensor installed in the mobile phone itself; however, the reading will not be as accurate. The data on the position and orientation of the mobile phone is automatically uploaded to the 3D-viewpoint server. The integration server determines



Figure 4: Sample of person’s Blog site generated using photo-logging data.

the correct position and orientation data for the picture by comparing the timestamp in the e-mail containing the picture and text with that in data on the position and orientation. It then stores this data in the database.

Every time the mail server receives e-mail with a picture attached, the picture is processed and is stored in the database along with metadata. The metadata includes the 3D viewpoint information as well as picture ID, e-mail date, title (subject of the e-mail), and description (content in the body of the e-mail). The blog server then updates the blog site so that the most recent picture is shown as the latest entry.

2.4 User Interface

Visualizing photographs in several different ways is easy because they and their associated 3D viewpoint information are stored in a general-purpose database. The system enables users to browse photographs “blog style,” by treating each photograph and corresponding explanation as a blog entry, displaying them in reverse-chronological order. Figure 4 shows an example of a blog generated using the photo-logging data.

Alternatively, the collected photographs can be displayed on a 2D or 3D map. When people browse photographs on a 2D map, markers with line segments are displayed on the map to show the position of the camera, i.e., the location from which the pictures were taken and the direction in which the camera was facing. The orientation of the camera is decomposed into an azimuth and an elevation angle, and the line segment attached to the markers represents the azimuth angle. Clicking on a marker, a user can

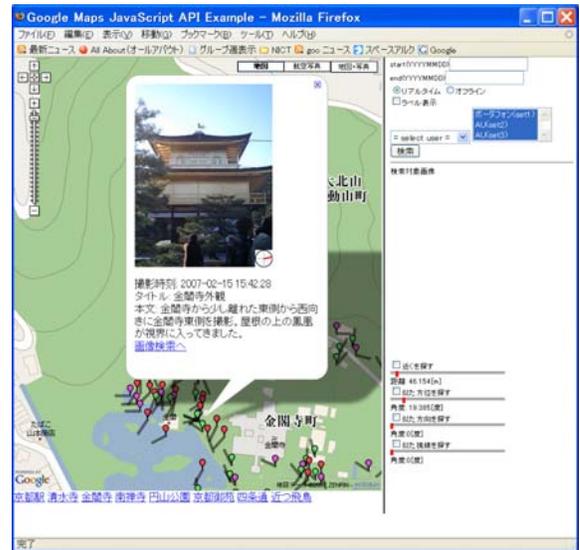


Figure 5: Map-based user interface. The markers denote the positions of the person when he or she took a picture, and the lines denote the direction in which the he or she was facing.

see a photograph with the corresponding elevation data represented graphically. Figure 5 shows a map-based interface.

2.5 Searching Images

The 3D photo-logging system uses multiple criteria for searching photographs, namely, the distance between two cameras, the degree of similarity in the azimuth angles, the degree of similarity in the elevation angles, the degree of similarity in the directions of gaze represented as 3D vectors, and the shooting date. After the user chooses a picture from those shown on the map, sets the values for the criteria, and clicks the search button, the system displays the results on the map graphically, i.e., with markers and line segments. Users can search photographs intuitively by interactively changing the value for the criteria. Figure 6 shows an example of an image search. One picture retrieved from the search results is displayed on the map.

3. GAZE-BASED GUIDANCE WITH MOBILE PHONES

We have implemented a proof-of-concept system by using the 3D photo-logging system. The system configuration is depicted in Figure 7. Suppose that we have a large collection of photos of tourist sites stored in the 3D photo-logging system. That is, each photo has 3D viewpoint information. Mobile phone users could then search for photographs of the tourist sites they are currently visiting by pointing and clicking with their mobile phones.

When a person is visiting a tourist site, he or she can receive a picture of his or her approximate location by accessing our experimental Web site. The system automatically locates the person’s position and the direction in which he



Figure 6: Image searches. A query image (on the right) and the corresponding search results (shown as markers on the map).

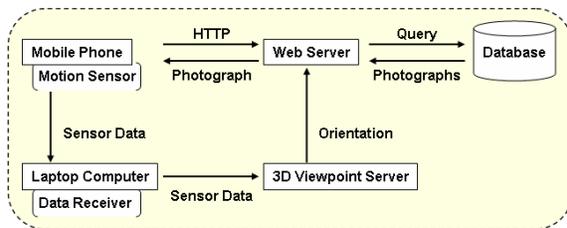


Figure 7: Configuration of searching function for mobile phones.

or she is moving, and uses this information to query the database. Note that the prototype system uses a motion sensor attached to the back of the mobile phone to detect the camera's orientation. In the current system, users receive pictures, as shown in Figure 8; however, they can either receive pictures with text or text only so that they can easily obtain information about what they are looking at.

4. CONCLUSION AND FUTURE WORK

We are developing a Gaze-based guidance system based on a real-world application of a 3D photo-logging system. The 3D photo-logging system records the picture together with its corresponding 3D viewpoint information and its explanation. Once the image database is created, we can use the data to send mobile phone users information on the basis of the direction of their gazes and their positions.

We believe that gaze-based guidance systems will increase the ease of access to real-world information in the real world. We intend to develop methods of finding the most appropriate photograph and/or information by using



Figure 8: Example of gaze-based guidance for mobile phones.

contextual data as well as the direction of a person's gaze and his or her position.

5. REFERENCES

- [1] Aichi Micro Intelligent Corporation. *Computer Graphics Controller (AMI601-CG) Instruction Manual AMI-MI-0132E*, 2007. http://www.aichi-mi.com/3_products/601-cgmanual_e.pdf (accessed 31 May 2007).
- [2] Aichi Steel Corporation. *AMI601 Delivery Specifications*, ver.1.3_061011 edition, 2006. http://www.aichi-mi.com/3_products/ami601ev1.3_061011.pdf (accessed 31 May 2007).
- [3] P. Drap et al. Photogrammetry and Archaeological Knowledge: Toward a 3D Information System Dedicated to Medieval Archaeology: A Case Study of Shawbak Castle in Jordan. In *Proceedings of the ISPRS International Workshop 3D-ARCH 2007*, 2007.
- [4] M. Flickner et al. Query by Image and Video Content: The QBIC System. In *IEEE Computer*, volume 28, pages 23–32, 1995.
- [5] R. Kadobayashi and R. Furukawa. Combined use of 2D images and 3D models for retrieving and browsing digital archive contents. In *Videometrics VIII, Proceedings of SPIE-IS&T Electronic Imaging 2005*, volume 5665 of *SPIE*, pages 134–143, 2005.
- [6] R. Kadobayashi, K. Kayama, T. Umezawa, and I. E. Yairi. Sensing Human Activities and Environment for Creating Knowledge Cycle. In *Proceedings of the First International Symposium on Universal Communication*. National Institute of Information and Communications Technology, 2007.
- [7] R. Kadobayashi and K. Tanaka. 3D Viewpoint-based Photo Search and Information Browsing. In *Proceedings of the 28th Annual International ACM SIGIR Conference (SIGIR 2005)*, pages 621–622, 2005.
- [8] J. R. Smith and S. F. Chang. VisualSEEK: a Fully Automated Content-Based Image Query System. In *Proceedings of ACM International Conference on Multimedia*, pages 87–93, 1996.

A Mock-up System for the Early Testing of Location Based Services

Florian Klomp maker
University of Paderborn
Fürstenallee 11
33102 Paderborn
florian.klomp maker@c-
lab.de

Claudius Stern
University of Paderborn
Fürstenallee 11
33102 Paderborn
claudius.stern@c-lab.de

Christian Reimann
Siemens IT Solutions and
Services
Fürstenallee 11
33102 Paderborn
christian.reimann@c-
lab.de

Holger Santelmann
Siemens IT Solutions and
Services
Fürstenallee 11
33102 Paderborn
holger.santelmann@c-
lab.de

ABSTRACT

In this paper we describe a mock-up system for Location Based Services for the use on mobile devices. It allows an early testing of applications, user interfaces, interaction techniques and computer vision algorithms. Since no special hardware and effort is needed to virtually place a user on an arbitrary location, it is a fast and low cost prototyping system. A specialised Wizard-of-Oz testbed is used to incorporate reality and to simulate application logic and localization. We use an User Centered Design approach by allowing the involvement of the user in the application's design process.

Categories and Subject Descriptors

D.2.11 [Software Engineering]: Software Architectures;
H.5.2 [User Interfaces]: Evaluation/methodology

1. INTRODUCTION

The processes of software engineering as well as the design of novel applications dealing with the real world is time-consuming and cost-intensive. Applying User Centered Design (UCD) can help developers to early evaluate application approaches, user interfaces and interaction techniques. UCD refers to the involvement of the user and has the goal to understand her needs and to analyse the requirements to perform a task and hence comprises an iterative design process and evaluation [1].

The design process contains constructional design at the one hand and behavioural design on the other hand [2]. Constructional design means the real, clearly structured software development process which is handled by programming experts. The behavioural design process considers the view of the user who interacts with the system using multimodal Human Computer Interaction (HCI) techniques. This process is accomplished by designers, HCI experts and evaluators. The complete UCD development process requires thus an accurate schedule and a sensitive cooperation between both

development groups.

Applications interacting with the real world mainly exist in the area of research for demonstration purposes. End users are rarely familiar with these applications and there interaction techniques. Hence, the challenge is the creation of highly intuitive, fast and reliable user interfaces. To test and evaluate these in the run-up, systems, tools and proceedings have to be designed that allow a fast investigation and evaluation of HCI techniques in the field of real world interaction. Our system is based on an existing mock-up system for the early testing of Augmented Reality (AR) applications and thus has the ability to easily provide AR functionality like adding virtual objects to the real world.

Applications on mobile devices that use Location Based Services have to consider the special characteristics of the devices as well as methods for the acquisition of context information. Graphical User Interfaces (GUIs) on mobile devices are often designed according to standard desktop GUIs. They rarely take the small display size and limited input modalities into account. A goal of our system is to develop suitable and highly intuitive GUIs and interaction techniques for mobile devices. Context information is needed for Location Based Services and has to be procured by the device. Therefore several device sensors like GPS, Bluetooth, infrared or other optical sensors can be used. Since it is quite difficult to test this data in a laboratory, we developed a mock-up system that emulates all needed context information in order to place the user on an arbitrary location where she uses Location Based Services.

The motivation for the creation of our system is founded in the development of services for the LOMS (Local Mobile Services) project [3]. Here local, location dependent services are to be developed which are highly context sensitive, use different end device sensors and provide intuitive user interfaces using AR. Further on tools for the fast creation of Location Based Services are to be created that support

non computer experts in the creation process. Therefore an early testing of applications and interaction techniques is essential.

This paper is structured as follows: After presenting some related work in section 2, section 3 introduces our mock-up system. We first address the reproduction of reality in our laboratory and afterwards explain the context of Location Based Services and the implemented line of sight detection of the system. At last we address the Wizard-of-Oz approach that we used. Section 4 introduces our ideas of future work and section 5 finally summarises this work.

2. RELATED WORK

Since Wizard-of-Oz simulation allows an early testing of user interfaces and interaction techniques it is a very easy, fast and reliable way to involve users into the development process. Some research is already done in this field and we here present a few examples.

A prototype system for mobile devices and applications is LiLiPUT (Lightweight Lab Equipment for Portable User Testing in Telecommunications) [6]. The system provides an early testing of mobile applications through several observer approaches. The user of a mobile device here wears a hat that is equipped with two cameras recording the user's surrounding and her facial expressions and with a microphone for recording her utterances. Further on a person (the observer) follows the user in order to make notes of her behaviour. The observer also wears a camera that records the user in her surrounding. Several applications providing information depending on the user's location are now tested and afterwards an evaluation is accomplished by analysing the camera and audio recordings as well as the notes of the observer.

The Wizard-of-Oz testbed presented in [11] describes a testing system for a medical robot. It uses information from both X-ray images and the surgeon to calculate a correct position and then physically guides a surgeon in the drilling through a patient's hip. Further on it provides a touchscreen with a GUI for displaying additional information and instructions. The sessions with several surgeons showed that Wizard-of-Oz prototyping allows users to be heavily involved in the external design of interactive products. The results from the sessions led to some decisions of the final user interface. For example the decision was made to not differ between double and single clicks on the touchscreen. Further on some inconsistencies in the instruction flow were detected and eliminated. Also some menu functions were completely removed because none of the test surgeons used it.

[7] introduces Topiary, a framework for Wizard-of-Oz testing of Location Enhanced Applications which provides maps, scenarios and storyboards for the easy creation of Wizard-of-Oz test situations. During a test, a wizard follows the user and permanently updates the environment of her, e.g. the positions of persons and things. The end user's user interface is automatically updated by Topiary if things change or the user interacts with it. The behaviour of the user interface is defined in the Topiary storyboard.

In [8] a system for Reality Reproduction is presented which

provides immersive video and surround sound. The aim is to let the user fully immerse into a virtual world and to test user interfaces and applications that deal with contextual information.

Mobile, CV based Line of Sight detection and information overlay is presented in [9]. Here a camera equipped PDA with GPS receiver is used to analyse images taken by the user. These are uploaded to a server machine that analyses which buildings are currently in the user's field of view. Additional information about a single building is then presented on the PDA.

3. MOCK-UP SYSTEM

This section describes the framework of our mock-up system, early testing procedures and the application currently running on the system.

3.1 REALITY REPRODUCTION

Since Location Based Services depend on the position the user currently remains it is difficult to test these kind of applications in a laboratory during the development process. Therefore our system is capable to simulate the reality of an arbitrary place. For raw position tracking GPS is used directly on the mobile device. Since there is no possibility to receive real GPS data in our laboratory we developed a GPS emulator. The emulator sends GPS NMEA strings via Bluetooth to the mobile device and thus acts like a GPS mouse. This enables us to virtually place the user on an arbitrary place by just changing the transmitted GPS data. We use a laptop pc as emulator which provides an application to easily change the GPS coordinates (see section 3.4). Since GPS only provides raw accuracy and does not address the direction the user is currently looking at, we further use Computer Vision (CV) based line of sight detection. We called this part of our system Reality Reproduction since GPS data and image data is created artificially for simulating an arbitrary place on earth.

Meanwhile there are only a few – and if, inaccurate – possibilities to get the actual heading or direction of a mobile device. Most of them use the incremental change of positions to compute an approximate heading but this method can only be used when the mobile device is moving around. In the case of a human carrying the device, the calculated heading is even less accurate. Particularly, these devices have only the information about past movements and the accuracy of e.g. GPS is not good enough to track rotation while the device is not moving. Assumed the mobile device is carried by someone who is stopping every now and then to look around, the device would not recognise the corresponding rotations. Of course, it is possible to track the user's rotations with an additional sensor, e.g. like in [10], where an external digital compass was used. In that approach, the data from the external orientation sensor was used to show the corresponding section of a panoramic image, which is located on a server and has been enriched with additional information before. The mobile device is connected via a wireless network to a server system, which holds the image database. The main disadvantage of this approach is its need for an external sensor to gather the orientational information. Furthermore, an additional sensor would be necessary to gather position information inside the museum;

since neither the digital compass nor a GPS system would provide the necessary information, thus another sensor has to be found. The intent of the approach described here, is to use sensors already available in the mobile device instead of attaching external ones. Currently, a trend is integrating as much functionality as possible in continuously shrinking hand-held devices. So today, you will not have a pure mobile phone, but rather a small PDA integrating Bluetooth, wireless LAN, UMTS and even a GPS receiver. Thus, all necessary sensors for an outside-the-house navigation are existent, but still it is hard to guess the actual heading of the mobile device while it is not moving. Almost all of these multi-functional hand-held devices integrates cameras with at least VGA resolution. So, it stands to reason using the integrated camera to get the actual heading.

Our approach uses prepared panoramic images for Reality Reproduction. Such an image is displayed on a huge screen using a projector. The panoramic image must accord to the GPS coordinates received from the GPS emulator. The user now records parts of the panoramic image by pointing the embedded camera of her device towards the screen. Using CV algorithms the direction of the user can be analysed. Figure 1 shows a model of the recording of an image part from a projected panoramic image.

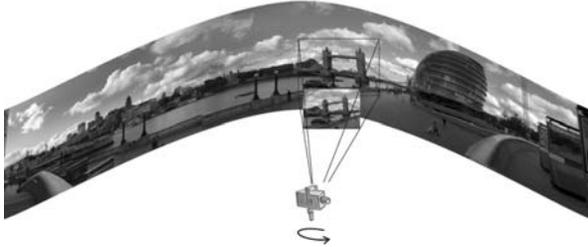


Figure 1: Panorama of London: Camera records an image part

3.2 LOCATION BASED SERVICES

The scenario described in this paper is to be implemented within the scope of the LOMS project [3]. Recorded image parts of a real surrounding are to be transmitted to a remote LOMS service and analysed there. The service afterwards responds with corresponding information about visible points of interest. The localization happens as previously described in a two-stage way via GPS and line of sight detection using CV. The GPS data is also transmitted to the LOMS service.

The main advantage of this system is, that besides complete applications also different interaction techniques, user interfaces and CV algorithms can be tested and evaluated. Further on AR functionalities for a rich superpositioning of the camera image can be applied. The use of remote services allows a faster analysis and enhancement of a recorded camera image since processors of mobile devices are rarely able to do such complex computations.

3.3 LINE OF SIGHT DETECTION

In most cases a heading information is needed to identify objects of relevance, e.g. interesting buildings, while looking

around. Hence it would be sufficient to get the information on user's request with a small delay. The user points her camera-and-GPS-equipped hand-held to a direction and initiates a request. Such a request consists of the image taken and the global position information obtained by the GPS receiver, both sent to a mobile service server holding the panorama picture database. The server then analyses the received image and extracts distinctive image features (e.g. SIFT points, see [5]) to match them afterwards against the database. The typically huge search space in the database is reduced thanks to the knowledge about the global position. As can be seen in [4], objects and their spatial orientation can be detected with this kind of image features. Using an image based line of sight detection also offers the possibility to enrich the recorded image later with additional information (e.g. using Augmented Reality which needs a high accuracy and addresses lens distortion). The extracted and the stored keypoints are matched against each other. A weighted sum of all positive matches is built up; and with its help the centre of the matched coordinates can be found. The found coordinate in the panoramic image can be interpreted as a heading or line of sight. With all these information together (global position, heading, up-to-date camera image), an augmented image can be sent back, where e.g. interesting points of view are marked. Along the way, additional information about the user's location or about the near by interesting locations can be sent to the user's hand-held.

3.4 WIZARD-OF-OZ SIMULATION

To allow an early testing of applications and interaction techniques it would be cumbersome and time-consuming to develop a complete application in each case. Hence we created a mock-up system which is based on Wizard-of-Oz simulation. Wizard-of-Oz simulation replaces the application logic by a human wizard. The wizard always provides a current GUI for the user and analyses her actions. If the user interacts with the GUI the wizard thus provides a new look of it containing changes caused by the action. Afterwards an evaluation of the application and interaction technique can be accomplished through user interviews.

Hence in our system the human wizard firstly cares about sending GPS coordinates that fit to the displayed panoramic image. The GPS emulator application provides a GUI that allows the easy change of GPS coordinates. It therefore provides a clickable map. If a position on the map is selected the corresponding GPS coordinates are calculated and henceforth continuously sent to the mobile device. Figure 2 shows the emulator application which here uses a map of a virtual theme park which can easily be replaced by e.g. a real city map. The user now applies the camera of her device to record image parts from the panoramic image. Once she made a decision of her line of sight she requests additional information about it by pressing a button on the GUI. The recorded image is now transmitted to a server machine which analyses the line of sight. If the user is looking towards a special point of interest, the wizard sends a new GUI to the user containing information about e.g. buildings or mountains she is currently looking at.



Figure 2: GPS emulator with virtual map

3.5 APPLICATION

Our application currently uses a panoramic image and GPS data of the city of London. The wizard initially provides the projection of the panoramic image and virtually positions the user into the city of London by using the GPS emulator. The GPS emulator from now on acts as a GPS mouse and continuously sends the GPS NMEA string via Bluetooth to the mobile device. The user now records an image part of the panoramic image with her camera equipped device and sends it to the server machine using WLAN. The server then computes the line of sight and displays the result to the wizard. The wizard decides which image is to be sent back and thus provides a superimposed image to the user that contains additional information about a building the user is currently looking at. After receiving the result image the mobile device displays it to the user. Figure 3 provides a scheme of the process containing concerned hardware and the data flow.

4. FUTURE WORK

Our system currently uses GPS data and a panoramic image of the city of London. We are going to create datasets for additional locations especially to test the line of sight detection with different CV algorithms on different panoramic images. We are also planning to use more special displays for the panoramic images e.g. an L-Shape or CAVE for a better surrounding of the user. The Reality Reproduction could be extended using further stimuli (e.g. noise) for a better immersion of the user. Besides CV algorithms we will also test different user interfaces and interaction technologies on mobile devices. Several scenarios will be created and user questionnaires will be accomplished in order to develop highly intuitive user interfaces for our application. This is then going to be implemented into the LOMS framework where different transmission technologies for service access may be used. Also user tests in the real world are imaginable to compare results with that from the laboratory tests.

5. CONCLUSION

In this paper we presented a mock-up system for Location Based Services. Inside our laboratory we are able to reproduce reality and to virtually place a user on an arbitrary place. Therefore a panoramic image is displayed on a huge screen and a GPS emulator is used to send GPS coordinates

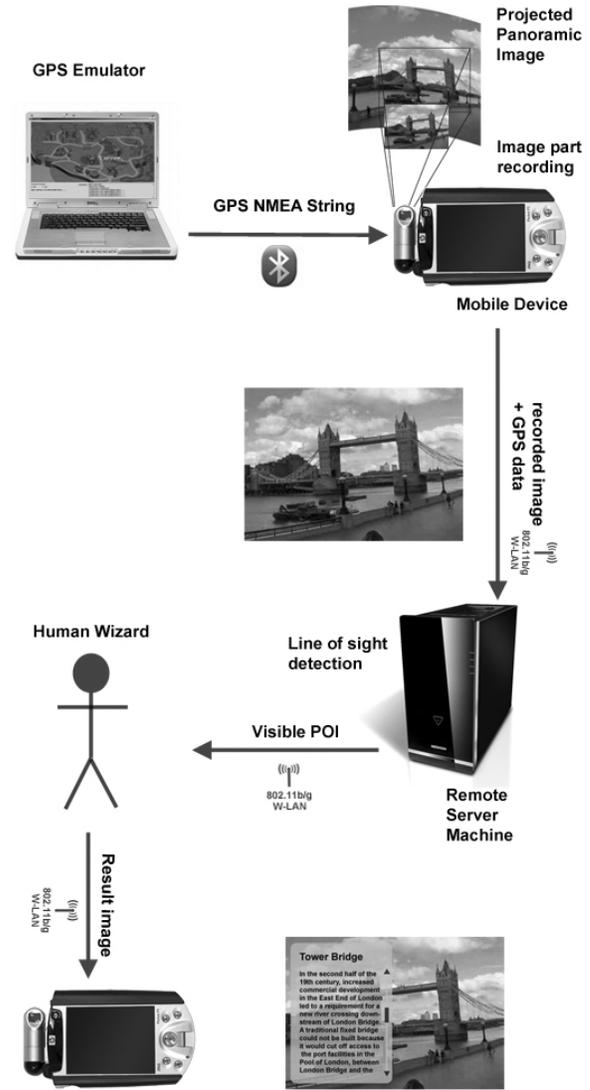


Figure 3: Application scheme

according to the panoramic image to the mobile device. Using CV algorithms and a client / server approach, image parts from the panoramic images recorded by the mobile device are sent to the server which analyses the line of sight of the user. Using Wizard-Of-Oz simulation a human wizard always knows, what the user is currently doing and where she is looking at. The wizard hence changes the GUI of the mobile device manually and can so provide additional information about buildings or landmarks the user is currently looking at. This Wizard-of-Oz simulation allows a fast and cost-efficient testing of the application, user interface, interaction technique and CV algorithm.

6. ACKNOWLEDGMENTS

We would like to thank the sponsors of the LOMS and D-LOMS project which are ITEA, BMBF and DLR. We also would like to thank Markus Becker for supporting us in soft-

ware creation processes.

7. REFERENCES

- [1] K. Vredenburg, J.Y. Mao, P.W. Smith and T. Carey. A survey of user-centered design practice. Conference on Human Factors in Computing Systems, 471–478, 2002
- [2] J.L. Gabbard, D. Hix, J.L. Swan II. User-centered design and evaluation of virtual environments. IEEE Computer Graphics and Applications, Volume 19, 51–59, 1999
- [3] LOMS: Local Mobile Services, <http://www.loms-itea.org/>
- [4] D.G. Lowe. Object recognition from local scale-invariant features. Proc. 7th Int. Conf. Computer Vision, 1150–1157, 1999
- [5] D. Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, Vol. 20, 91–110, 2003
- [6] P. Reichl, P. Froehlich, L. Baillie, R. Schatz and A. Dantcheva. The LiLiPUT prototype: a wearable lab environment for user tests of mobile telecommunication applications. Conference on Human Factors in Computing Systems, 1833–1838, 2007
- [7] Y. Li, J.I. Hong, J.A. Landay. Design Challenges and Principles for Wizard of Oz Testing of Location-Enhanced Applications. IEEE Pervasive Computing, Volume 6, 70–75, 2007
- [8] P. Singh, H.N. Ha, Z. Kuang, P. Oliver, C. Kray, C. Blythe, P. James. Immersive Video as a Rapid Prototyping and Evaluation Tool for Mobile and Ambient Applications. Proceedings of the 8th conference on Human-computer interaction with mobile devices and services, 264, 2006
- [9] G. Fritz, C. Seifert, P. Luley, L. Paletta, A. Almer. Mobile Vision for Ambient Learning in Urban Environments. Proceedings International Conference on Mobile Learning, 2004
- [10] L.W. Chan, Y.Y. Hsu Y.P. Hung and J.Y.J. Hsu. Orientation-Aware Handhelds for Panorama-Based Museum Guiding System. UbiComp 2005 Workshop: Smart Environments and Their Applications to Cultural Heritage, 2005
- [11] L. Molin. Wizard-of-Oz prototyping for co-operative interaction design of graphical user interfaces. Proceedings of the third Nordic conference on Human-computer interaction, 425–428, 2004

Demonstrating Coherent Interactions between Personal Mobile Devices and Situated Installations

Ben Bedwell

Boriana Koleva

Mixed Reality Laboratory
School of CSIT, University of Nottingham
Nottingham, NG8 1BB
+44 (0)115 9 514 227

{bzb, bnk}@cs.nott.ac.uk

ABSTRACT

Mobile devices and situated installations have differing inherent advantages and limitations; coupling – the physical and computational combination – of personal mobile devices with situated installations allows designers to overcome these disadvantages and create interactional synergy. This paper presents a demonstration of a small macro-environment: three conceptually-linked installations to which visitors can couple their personal mobile phone and interact. We discuss two significant design issues – approaches to support the exploration of the macro-environment and the use of portable content - raised during the development of the demonstration. Finally we conclude the paper by outlining our first and future steps in developing a comprehensive guiding framework in this emerging domain.

Categories and Subject Descriptors

H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—*Artificial, augmented, and virtual realities*; H.5.2 [Information Interfaces and Presentation]: User Interfaces—*Input devices and strategies*

General Terms

Design, Human Factors, Theory.

Keywords

Coupling, exploration, context-awareness, targeted advertising, personalisation, mobile devices, situated installations.

1. INTRODUCTION

Personal mobile computing devices – mobile phones, PMPs (personal media players) and PDAs – are *connectable* (both physically and computationally by hardware and software interfaces) and *identifiable* (among other devices by MAC, IP addresses and so on, and as digital representations of their owner). As such, users could navigate an environment (the ‘macro-environment’, such as a museum) populated by situated installations carrying their mobile device, travelling to and coupling their mobile to installations in order to personalise the services offered by the installation. In addition these mobile devices provide increasingly sophisticated computing capabilities, i.e. *dynamic data storage, input and output*. The personal mobile device also represents a familiar technology interface; many users will be intimately familiar with their mobile phone.

We can now envisage a scenario such as that illustrated in figure 1, based on the activities of a museum visitor. In this example, a visitor enters the museum (the “macro-environment”) and arrives

at a terminal (A) where they use the terminal’s keyboard to enter their name, age, etc. and also physically couple their mobile to the terminal screen to act as a webcam in order to add a photograph to their registered profile. The visitor finishes their registration by decoupling their mobile from the terminal, and then walks over to a large map of the world, blank except for the outlines of continents and a number of glyphs pinned at various positions (B). The visitor photographs glyphs using their mobile which in turn reveal clues about the marked locations – in some cases this is an interesting fact, while a select few glyphs hide a flag which can be grabbed and held on the visitor’s mobile. Once the visitor has explored a number of glyphs and grabbed one or more flags, they are notified that they can explore some of their findings in more detail at the next installation. They leave the map and walk to the final installation (C) where they find that they can create slideshows on a large projector screen based upon the locations from which they grabbed flags, and control the shows using their mobile as a remote control. While watching a slideshow the visitor can save particularly interesting slides to their mobile. The slideshow content is filtered based upon the registered profile of the visitor.

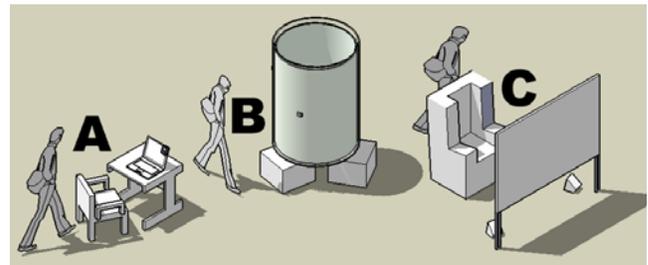


Figure 1. A coherent sequence of coupled interactions

Several key questions arise from a brief consideration of this example, including (but not limited to), “*how can users become aware of - and locate - situated installations? What interactions are afforded by coupling mobile device with installation and how does the user perceive these affordances?*” We began to explore theoretical approaches to map the design space (efforts which are discussed in the conclusion) while at the same time developed a small demonstration of situation similar to that in figure 1 in order to provoke and explore practical design issues.

2. INITIAL DEMONSTRATION

The prototype macro-environment and mobile client (installed on several mobile phones) have been designed in our lab to support two main activity states (inspired by Ullmer, Ishii and Jacob’s [11] TUI interaction types, exploration and performance):

- *Macro-exploration* – the mobile device is a tool used to explore the environment to discover an installation; and
- *Coupled action* – the mobile and discovered installation combine to afford the visitor rich situated interactions.

With reference to figure 1, the two states can be identified essentially as the intermediate steps (macro-exploration) and the labeled (A, B, C) steps (coupled action) respectively.

2.1 Macro-exploration

2.1.1 Entry and visitor profiling

Upon entry to the macro-environment and initialisation of a mobile client by a visitor, a profiling dialogue takes place between the visitor’s mobile client and a server application (the ‘visitor monitor’). During this process, a profile for the visitor (consisting of three parts) is created and stored centrally:

- one *technical* part (what software/hardware components does their mobile include?), built automatically upon entry to the environment;
- one *personal* part (what are the visitor’s demographics, interests, etc.?), empty until filled by the visitor at the first installation; and
- one part logging their *macro-exploration* (which installations have been visited and what content has been picked during those visits?), altered as the visitor’s experience progressed.

2.1.2 Navigation

We hypothesised that in a rich environment visitors would require clues to fully understand where useful installations are and how they can be used, especially on entry to the environment. Traditional signs do not consider the context of individual visitors, e.g. different galleries in a museum have the same exhibit signs for each visitor, regardless of how interesting the different visitors might find those galleries. We suggest that the personal mobile device can take on the role of an *advertising proxy* for the different exhibits. By directing visitors through their mobiles, we can discriminate between different visitors and target adverts based on context. In practical terms, a visitor would receive a number of adverts for contextually suitable installations, choose one from among those adverts, then be guided to the corresponding installation. By being able to design rules defining who receives adverts under different conditions, the designer can not only ensure that the visitors are directed to valuable installations but also gain control over the flow of visitors in the environment. The portability of the mobile device enables ‘the most suitable information to be delivered at any time and place’.

As such two questions needed to be addressed during the design of our demonstration: *when* should a visitor receive adverts and *what* adverts should they receive (i.e. what installation do they advertise)?

When should adverts be delivered?

In response to this question we implemented two approaches. Adverts may be ‘pushed’ to the visitor by the environment unrequested (i.e. the visitor monitor triggers delivery of the advert(s)), and the visitor may ‘pull’ or demand suitable adverts at will. These actions are comparable respectively to the passive (‘information sniffing’) and active (‘selection’ and ‘search’) forms of spatial exploration for content proposed by Fröhlich et al. [3].

However, we apply Fröhlich et al.’s terms (in addition to the term ‘dictation’) to the act of spatial exploration for installations rather than content. The approaches to advert delivery are summarised in table 1, where they are shown in relation to an axis representing the balance of control between user and system.

Table 1. Advert delivery and balance of control

<i>Push</i>	Dictation	System control
	Sniffing	↑
<i>Pull</i>	Search	↓
	Selection	Visitor control

We investigated the supposition that a combination of the two approaches is necessary in order to provide the visitor with the most satisfactory experience. Adverts were *dictated* to visitors when they entered the environment and every time they decoupled from installations. If such a push (or a pull, as described later) results in the visitor monitor returning just one advert to the visitor, photo-follow directions to the relevant installation are triggered immediately and one coupling ‘slot’ at the installation is reserved for the visitor (so that there is still capacity for the visitor when they finally reach the installation). If more than one advert is returned, the adverts are presented to the visitor as choices; directions are triggered once the visitor confirms a choice. If any suitable installations are determined by the visitor monitor to be fully occupied, they are advertised (rather than hiding the installation from the visitor), but these adverts cannot activate directions.

Whenever choices of adverts are presented in this way, the visitor has time to rest; visitors can also choose to interrupt any directions that they are currently following (and release their reservation at that installation) and return to this resting state. Once at rest, the visitor can then pull an advert by *selecting* any installation that they discover by themselves (by scanning a barcode attached to the installation) or by *searching* for installations (instructing their mobile to pull a fresh set of adverts, updating their choices). Trials by Izadi et al. [7] suggest that the discovery of hidden information (installations in this case) by user-controlled exploration causes the user to engage more deeply with their discoveries, implying that the element of initiative inherent to these approaches, in contrast to dictation, is rewarding. The visitor can also focus their attention elsewhere and leave their mobile to *sniff* for fresh adverts pushed to it by the visitor monitor at regular intervals.

From preliminary trials we are beginning to find that allowing the visitor to rest is essential. Visits to a learning environment are much less beneficial if the visitors are not given time to reflect upon their experiences and discuss them with other visitors: it must always be remembered that visitors to museums, galleries, and so on often visit as part of a group and thus the experience is a social, typically organic one where restrictions on behaviour detract from the enjoyment of the visit. In addition, we do not suppose that macro-environments will contain only installations; they will often contain non-interactive exhibits that visitors may wish to experience without any interference from their mobiles. While we have found that pure dictation does not necessarily provoke a negative response from visitors (visitor guides are often totally linear, yet valuable and enjoyable), providing only a limited amount of initiative causes resentment towards the system – one visitor to an early prototype of the demonstration macro-

environment, before the option to interrupt directions was included, commented that, “I am intelligent – why do I have to let the mobile keep telling me when to move?”

What adverts should be delivered?

There are a number of factors – common to all macro-environments – that we suggest can influence which adverts should be chosen by the visitor monitor to be delivered to a visitor following a push or pull.

Resource management is an important consideration for designers of macro-environments. Many of the installations the designers may create – in particular those that form tangible couples - will be inherently limited in their coupling capacity. The installation illustrated at point A in figure 1 for example has a physical holster in which the visitor places their mobile to begin the local experience, while the installation at point C allows a visitor to create and control slideshows displayed on a large projector screen. Both installations are thus limited to couple with only one mobile at a time hence we did not wish to frustrate visitors by advertising those installations when they were already occupied. It is also possible to frustrate visitors by advertising installations that are irrelevant. Again we may refer to the slideshow installation: if the installation creates and visualises slideshows based upon content collected at a previous installation (in this case at the map installation) there is little point advertising the installations to visitors that have not visited the map, or have visited it but collected no useful content. Particular installations may also be irrelevant to visitors if they use coupling interfaces that do not match those coupling interfaces found on their mobile devices. For example, the map installation in our macro-environment utilises a camera on the visitors’ mobiles to allow visitors to photograph the glyphs representing locations that they wish to reveal information about; if a visitor’s mobile device does not include a camera then a direction to this installation would be frustrating.

As such when we create an installation we define three attributes for it:

- *technical requirements* (which technical components does a visitor require to couple with this installation?); and
- *storyline requirements* (visits to which other installations/which collected content are prerequisites for this installation?), both defined when the installations are initialised; and finally
- *coupling profile* (who is currently coupled to the installation and what total capacity does the installation have?), updated as visitors couple with and decouple from the installation.

In practice, each decision made by the visitor monitor (following a choice to push, or receipt of a pull request from a visitor) required a series of cross-references, using the visitor’s profiles and the attributes of every installation to determine the visit suitability of each installation. The installations are classified as either suitable or unsuitable, and then the suitable installations are prioritised depending upon whether they have been visited before and whether they are a prerequisite to another installation. Having classified each installation, the visitor monitor passes adverts for any classified as suitable back to the visitor’s mobile, and they are handled by the visitor as described previously.

Our prototype contains only three installations, and preliminary trials have involved up to a maximum of three visitors at any one time, hence we have yet to test our proposals for macro-exploration with respect to scalability. Key issues to be tackled by further development of our prototype include the impact of higher (more realistic?) numbers of visitors upon flow around a macro-environment, and how the practical implementations of our four approaches will need to be modified to be most effective in dealing with this impact. Also we must investigate the impact of the social structure of the visitors, e.g. do socially connected visitors such as families and friends explore a macro-environment differently? Do they use their personal mobile devices collaboratively? Do adverts need to be chosen for visitors based not just upon personal context but also upon social context?

2.2 Coupled action

Due to the brief nature of this paper we shall discuss only one issue amongst those raised concerning the coupled actions at the installations in the demonstration: the use of *portable content*.

One of the most valuable roles identified for the personal mobile device in the macro-environment is as a *vessel* for content, blurring the boundaries between container, token and tool object types described by Holmquist et al. [6]. While installations (and their content) are situated within specific micro-environments, a mobile device is the perfect container to carry and spread content that might otherwise remain tethered to one point-of-interest. We refer to content that can be moved as *portable content* and our prototype macro-environment illustrates two basic uses for such content.

The flags which may be grabbed from the map installation can be considered *portable selections*, as they hold little interest for the visitor by themselves, but can be used as triggers or *seeds* to create personalised experiences when carried to other installations. In the case of the demonstration, as previously described, the visitor may carry the selections to the slideshow installation, where after coupling they can be used to initiate slideshows based upon content concerning the selected locations.

Our slideshow installation allowed the visitor not just to view slides seeded from their portable selections but also to rate and/or grab individual slides. Slides grabbed by the visitor formed part of their *personal collection* (which we conceptualised as a wallet) that the visitor could browse once they decoupled from the slideshow installation and even when they left the macro-environment through a separate wallet application. As the slides were too rich in information to be shown on a mobile display, each ‘card’ added to the wallet simply showed a thumbnail when browsed by the visitor. Our intention is that the user is enabled to temporarily re-create a slideshow installation outside of the macro-environment on-demand, using their PC, TV, or other large display to re-view their wallet.

We saw unexpected social behaviour provoked even by our small trials, e.g. when a group of visitors were close friends visitors often became the ‘audience’ at the slideshow installation, leaving their mobile device at rest while their friend (the ‘conductor’) coupled to the installation. We are beginning to use this prototype to explore how the experience that the coupled visitor has effects the audience, especially when the coupled experience is seeded – in this case the slideshows viewed by audience members will be different to those available when they couple as they will have

gathered different portable selections. In fact we deliberately encouraged this effect by also altering the slideshows created from the same seed with respect to the visitors' personal profiles, e.g. different slideshows would be created for 10 year-old and 25 year-old visitors even if they both used the same portable selection 'Japan'. We may hypothesise that such design can encourage the audience to return to previous installations to find seeds they have missed, and that among some visitors a sense of competition may be aroused. One alternative to discovering original content in situ is for visitors to grab already-discovered portable content from each other. In fact using the previous example of the seed 'Japan', it will be impossible for the 25 year-old to discover some of the slides seeded by the 10 year-old, hence this method would be necessary for him to collect the same slides to his wallet. We intend to explore the value of collaborative management of portable content at multi-couple installations, such as trading tables like Dynamo [8]: does trading of content reduce its role to currency and thus discourage reflection on the content itself?

3. CONCLUSION

We have briefly presented a demonstration of a small macro-environment fully implementing a range of approaches to spatial exploration for installations, and coupled interaction. The demonstration provoked a number of design problems, an explanation of all, especially the technical aspects, being well beyond the scope of this paper. Instead, two interesting challenges brought to our attention during development of the demonstration were discussed: section 2.1 illustrated four approaches to supporting visitor exploration of the environment, while showing how the visitors' and system's needs for control over this process must be reconciled; section 2.2 introduced some simple uses of portable content in the macro-environment, and touched upon the social behaviour that might be initiated through its use.

In parallel to the demonstration we have been developing a theoretical framework for the design of coherent sequences of coupled experiences. The two main pillars of this framework are a dual model of human-computer interaction and system interaction (from entry into the macro-environment to exit from the macro-environment), as well as a systematic mapping of different configurations (physical and computational) of mobile-installation couples.

For the former we draw strongly upon our experience with the demonstration presented here. For the latter, authors such as Norman [9], Gibson [4] and Dourish [2] have made important contributions to our understanding of the perceptual effects of physical configuration and appearance, while useful systemisations of devices in terms of physical and computational properties have also been proposed [5]. Of particular interest from the field of tangible computing is work on the use and implications of mechanical constraints, such as that by Shaer et al. [10] and Ullmer et al. [11], as there is scope here for the application of this research to describe the mechanical effects of the coupling to the installation upon the mobility of the mobile device.

In isolation these two theoretical components of the framework will provide for environment designers a useful descriptive vocabulary and boundaries to the design space. While these tools may allow designers to effectively implement their visions of the macro-environment, we further intend to raise a number of salient design questions based upon the theoretical components, in a

similar manner to Bellotti et al.'s [1] design questions for the more general domain of 'sensing-UIs', which can not just guide, but provoke the design of environments which utilize the full potential of this emerging design space.

4. REFERENCES

- [1] Bellotti V., Back M., Edwards W., Grinter R., Henderson A., Lopes C. 2002, 'Making sense of sensing systems: five questions for designers and researchers', *CHI*, Minneapolis, Minnesota, USA, pp.415-422.
- [2] Dourish P. 1999, *Embodied Interaction: Exploring the Foundations of a New Approach to HCI*, [Online]. Available from: <<http://www.dourish.com/embodied/embodied99.pdf>> [7 November 2005].
- [3] Fröhlich P., Simon R., Baillie L., Anegg H. 2006, 'Comparing Conceptual Designs for Mobile Access to Geo-Spatial Information', *MobileHCI*, Helsinki, Finland, September 12-15, pp. 109-112.
- [4] Gibson, J. 1977, *The Theory of Affordances*. In *Perceiving, Acting and Knowing*. R. E. Shaw & J. Bransford, Eds. Lawrence Erlbaum Ass. Hillsdale, CA.
- [5] Gross M., Do E. 2004, *Toward Design Principles for Invisible Interfaces*.
- [6] Holmquist L., Ljungstrand P., Redstrom J. 1999, 'Token-Based Access to Digital Information', *HUC*, Karlsruhe, Germany, pp.234-245.
- [7] Izadi S., Rodden T., Benford S., Fraser M., Flintham M., Schnadelbach H., Greenhalgh C. 2002, 'Citywide: Supporting Interactive Digital Experiences Across Physical Space', *Personal and Ubiquitous Computing*, vol.6, no.4, pp.290-298.
- [8] Izadi S., Rogers Y., Rodden T., Underwood M., Brignull H. 2003, 'Dynamo: A public interactive surface supporting the cooperative sharing and exchange of media', *UIST*, Vancouver, Canada, pp.159-168.
- [9] Norman D. 1999, 'Affordance, conventions, and design', *Interactions*, May-June, vol.6, no.3, pp.38-43.
- [10] Shaer O., Jacob R., Leland N., Calvillo-Gamez E. 2004, 'The TAC paradigm: specifying tangible user interfaces', *Personal and Ubiquitous Computing*, vol.8, no.5, pp.359-369.
- [11] Ullmer B., Ishii H. and Jacob R. 2005, 'Token+Constraint Systems for Tangible Interaction with Digital Information', *ACM Transactions on Computer-Human Interaction*, vol.12, no.1, pp.81-11

Contextual Bookmarks

¹ Niels Henze, ² Mingyu Lim, ³ Andreas Lorenz, ⁴ Michael Mueller, ⁵ Xavier Righetti,
⁴ Enrico Rukzio, ³ Andreas Zimmermann,
² Nadia Magnenat-Thalmann, ⁶ Susanne Boll, ⁵ Daniel Thalmann
¹ OFFIS - Institute for Information Technology, Oldenburg, Germany
² MIRALab - University of Geneva, Switzerland
³ Fraunhofer-Institute for Applied Information Technology FIT, Germany
⁴ Computing Department, Lancaster University, UK ⁵ VRIlab, EPFL, Switzerland
⁶ Media Informatics and Multimedia Systems, University of Oldenburg, Germany
henze@offis.de, {mingyu.lim, thalmann}@miralab.unige.ch, {xavier.righetti,
daniel.thalmann}@epfl.ch, {m.mueller, rukzio}@comp.lancs.ac.uk, {andreas.lorenz,
andreas.zimmermann}@fit.fraunhofer.de, susanne.boll@informatik.uni-oldenburg.de

ABSTRACT

The role of digital information in everyday life divides out activities in interacting with the physical and the digital world. There is no connection between these worlds that is easily accessible, even though physical objects, persons, and real world events often have digital counterparts. The physical reality is actually overlaid by an additional virtual or digital layer. As we are acting in the physical and the digital world it is desirable that we can use links that bridge the gap between both worlds. In this paper we describe our approach to narrow this gap. Starting from a scenario that shows the demand for such connections, we present an architecture that allows users to bookmark specific situations. Based on these contextual bookmarks the user can request additional digital information. Our first prototype enables the user to bookmark content shown on public displays by taking photos of the display using a mobile phone. Our system combines content analysis of the photo with context information such as position, creation time, etc. in order to form the basis to establish a link to the digital world. The presented architecture will serve as a flexible solution to find and integrate further connections between the physical and the digital world.

Categories and Subject Descriptors

H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities, H.5.2 [Interfaces and Presentation]: User Interfaces - Interaction styles; D.2.2 [Software Engineering]: Design Tools and Techniques - User interfaces

General Terms

Algorithms, Design, Human Factors

Keywords

Mobile interaction, user interface, contextual bookmark.

1. INTRODUCTION

With the advent of the era of ubiquitous computing environments, it has become common for people to have their own mobile devices (e.g. phones, PDAs, and multimedia players) and access the Internet not only at home and in the office but also from hot

spots and other wireless networks offering Internet connection outside buildings. All these facilities let us face a flood of digital multimedia services through which we interact in both the digital and the real world because physical objects, public events, and people in the real world could have digital counterparts. For example, various advertisement posters in train stations or bus stops can also be found on the Internet in a digital form. We can not only enjoy a film in a cinema in the real world but also by buying and downloading the corresponding file as a digital item. However, today there is no easy accessible link that connects physical objects to digital items.

If there is a public event, which we are interested in and we want to remember, we need to record the information by writing down a memo in a calendar or a personal scheduler in a PDA or a mobile phone, which generates high demands on the people. Therefore, this paper proposes a contextual bookmark approach that aims at bridging the gap between the real and the digital world.

In this paper we introduce the concept of *contextual bookmarks* in the subsequent section. Then, section 3 motivates our approach through an illustrative usage scenario described. Section 4 describes related work regarding the identification of real-world objects. Section 5 discusses the architecture of our approach and the current state of our development efforts. Finally, section 6 concludes the paper with further work.

2. CONTEXTUAL BOOKMARKS

We define a contextual bookmark as a combination of a snapshot of a physical object taken with a mobile device and meta-information about the content related to this physical object. According to and extending our scenario described below, a physical object can be a public display, an advertisement poster at a train station, or an exhibit of a museum. The meta-information about the physical object basically results from content analysis and also comprises context information acquired from the environment such as time, location or temperature. Furthermore, this meta-information may cover information about the user's preferences and intentions.

A contextual bookmark represents a physical object in a digital form. A user of our system uses his or her personalized wearable

device to record and define a contextual bookmark. Currently we use a Nokia N95 which is one of the most recent camera-equipped mobile phones that is equipped with additional sensors (e.g. GPS) for acquiring context information. In a later phase this device will be extended by additional input controls (e.g. wearable buttons), output capabilities (e.g. headphone, see-through glasses), wearable storage and various sensors (e.g. a light or temperature sensor).

The contextual bookmark provides a handle to the content and the respective situation, in which it has been recorded. Browsing the list of contextual bookmarks on the mobile device enables the user to exploit services that are related to the context, in which the bookmark has been defined. Once the user selects a service associated with a contextual bookmark, he or she can, for example, access the content of the contextual bookmark via the mobile device or via a nearby output device depending on the current context and preferences (e.g. public display, TV, radio, laptop, etc.). Furthermore, the user gains access to more detailed information about the content, which was not available in the situation in which the contextual bookmark has been taken.

Although our first prototype system is specific for a movie trailer bookmark as a useful example, it can also be adapted to contextual bookmarks of various other types of physical objects such as posters advertising upcoming events, songs played on the radio, and even people who were met at a conference or in a business meeting. As such, our approach, instead of providing a solution for a single scenario, aims to support a flexible solution that enables to integrate various sensor data, media types, means of data access, and matching algorithms for different scenarios.

3. SCENARIO

This section presents a scenario which shows the need for and usefulness of a mobile contextual bookmark application.

Jim is on a business trip in Berlin and sits in the metro. The train is equipped with small advertising screens; suddenly, the trailer of the latest Tarantino movie pops up on the screen. Tarantino is one of Jim's favourite movie writers and he does not want to miss this movie again. He points at the screen with his mobile phone and bookmarks the trailer as a movie he likes to see.

During the day, Jim spends endless time in business meetings discussing the next release of the company's software. It is already late afternoon, when the meetings finish and he is on the way back to the hotel. He updates his digital calendar because business meetings are finished for today. The system automatically checks the cinemas near Jim's hotel and asks him if he wants to go to the movies tonight at 8:30pm to watch the movie he bookmarked in the train. Jim remembers that he has an appointment with a former class mate tonight and refuses with a sigh.

The next day, Jim is back at home. He takes his mobile device and browses the bookmarks he had taken in Berlin: There are some sights like Brandenburg Gate, the Berlin Wall and Federal Chancellery, the presentations of the business partners, and some advertisements he bookmarked to remember. There is also the movie trailer he had bookmarked in the train. The bookmark provides Jim with a link to download the trailer. When he watches the trailer he realizes, that he still had not have time to watch the movie. He decides to invite some friends to watch the movie on his home cinema system. Jim sets up everything, gets some beers

and calls some friends. He selects the bookmark and advises his mobile phone to get a license for downloading the movie to his set-top box. At home, the device retrieves his personal movie pay-per-view account from his PC, requests a license from the database and asks Jim to acknowledge the payment. He confirms and the license is available.

Because some friends were late that night, they are already behind the schedule when they start watching the movie at 8:50. Jim selects the movie license; the system automatically determines the best fitting output screen and downloads the movie in highest quality to the set-top box of his home cinema system. The movie starts and everyone enjoys watching it. Unfortunately, Joe has another appointment and as they started late he cannot keep watching the movie at Jim's place. Jim acknowledges sharing the license of the movie with Joe. Joe pays a small fee for the last 20min to finish the movie on the train with his mobile phone. Joe's device requests the license from Jim's phone and downloads the last 20 min in lower resolution that fits with the capabilities of Joe's phone. Joe takes it with him while the rest of the group finishes the movie at Jim's place. They all had a great time.

4. RELATED WORK

There have been approaches to record and retrieve information based on markers or image analysis. However, they require external markers tagged on real objects or sophisticated content-based image analysis. These concepts are not sufficient for understanding the exact meaning of media items, which users really want in the current context. Some projects related to our contextual bookmark approach are presented in the following.

Hansen et al. introduce the term "mixed interaction space" to distinguish camera-based interaction from other types of sensor-based interaction on mobile devices [1]. Indeed, physical space plays an important role in camera-based digital interaction that is controlled by both movement and orientation of a mobile phone in the space. But there are many ways for mobile phones to interact with their environment, especially with large situated displays. Ballagas et al. classified more than 15 different interaction techniques using various communication technologies and interaction paradigms [2]. Concurrently, Ailisto et al. have established a detailed comparison table between four potential commercial technologies for physical selection [3], which comprise Visual Code, IrDA, RFID and Bluetooth.

However, Rukzio et al. note that there is very little support to build such kind of systems [4]. Hence, they present an architecture named "Physical Mobile Interaction Framework" which is based on existing standards such as the Java 2 Micro Edition (Java ME) and the Contactless Communication API. The authors plan to support all relevant interaction techniques between the device and the object by providing abstractions for the programmer, hiding technical details of the communication.

Attaching markers to real world entities makes it possible to find corresponding digital entities. Simple implementations of this approach are barcodes developed by Silver and Woodland back in 1948. Barcodes can only store a very limited amount of data but nonetheless are used to identify physical objects [6]. Thus, systems using more sophisticated markers have been developed. Examples are 2D barcodes [7] radio-frequency tags [13] and infrared tags [14]. However, all marker based approaches need a marker that is somehow attached to the object. If a user want to

select a marker to acquire information about the corresponding entity the user or the device must find this marker. This is often not desirable because visual markers must be small, with long range radio frequency based approaches the user's device can not decide which marker was selected, and infrared tags are outshined by the sun. In addition, all markers can be masked by other people or objects and for some real world entities, like human beings, it is not even feasible to attach markers.

Raj et al. present a mechanism that allows users to implicitly download content available on public displays [5]. The ContentCascade framework enables the user to download either summary information or movie clips via Bluetooth. They suggest measuring the interest of the content by analyzing the user's behaviour and the length of the interactions with the public display. Cheverst et al. developed and evaluated the Hermes photo display, a prototype that enables users with a suitable mobile phone to both send and receive pictures via Bluetooth [8]. The display uses Sun's Bluetooth API to connect to the mobile phone and utilizes OBEX push to transmit the files. The advantage of this approach is that it requires no additional applications to be installed on the mobile phone. The Hermes door display [9] is another prototype developed by Cheverst et al. and installed on the outside of ten offices' doors in Lancaster's Computing department. The Hermes door displays give offices' owners and visitors the ability to leave notes on the door using their mobile phone. During the 24 months of its use, more than 6000 notes were added. All three Bluetooth based systems need additional hardware infrastructure integrated in the public display and shares the limitation of marker based approaches described above.

Among other projects, we can also cite PhoneGuide [10], an enhanced museum guidance system that uses camera-equipped mobile phones and on-device object recognition. Föckler et al.'s main technical achievement is a light-weight object recognition algorithm that is realized with a single-layer perception neuronal network on the device itself. According to their results, over 90% of the photographed museum exhibits can be recognized without the addition of passive or active reference markers. The limitation of Föckler's promising approach is the small number of recognizable objects and that it does not work for all types of real world objects. We believe that both can significantly be improved by taking the user's context into account.

5. ARCHITECTURE AND IMPLEMENTATION

We are currently working on a first prototype which realizes a small subset of the functionality described above. The prototype will consist of multiple computers with attached large displays that show different movie trailers. Persons can use their mobile phones to define a contextual bookmark by taking a photo of the displays playing the trailer. The bookmark can be transmitted to further computers to play the connected movies.

The aim of the prototype is to provide a flexible architecture that enables the integration of further matching techniques and to evaluate the proposed interaction. The general architecture of the system integrating further extensions is outlined in Figure 1. On base of this architecture we can analyse context and content of a given snapshot of the user's situation. In the following subsections we describe the lifecycle of a contextual bookmark

from sensing the environment through matching snapshots of the sensed context to exchanging bookmarks.

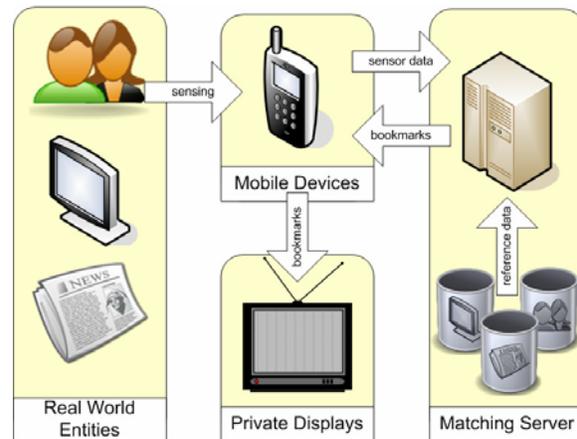


Figure 1. Generic architecture of the prototype currently under development

5.1 Sensing the Environment

When the user invokes the bookmark function on his or her mobile device a snapshot of the current context is created. The first prototype will be based on mobile phones (e.g. the Nokia N95) and Java ME, the most common platform for mobile phones. After starting the system the user can focus the mobile phone's camera at the public display. When the user activates the bookmark function a photo is created and enriched with the actual position using the integrated GPS receiver and a timestamp. This information represents a snapshot of the user's context at this point in time.

To ease the integration of additional sensor information and react to the respective devices individual capabilities, for instance not all mobile phones have an integrated GPS receiver, the snapshot contains self-describing sections each containing a distinct feature of the context. Thus, further extensions, e.g. recording audio, sensing the user's emotions, or determine his or her tasks using digital agendas can be integrated easily.

5.2 Matching

The snapshot is transferred to a server that retrieves links to related digital items and services. Analysing the snapshot of the context is a processing power and memory consuming task. Thus, the architecture envisages implementing the matching of context to digital items and services on a separate server. The server offers its services via a web service interface. The first step of the matching process is an analysis of the kind of context data included in the received snapshot. According to this, the data is delivered to matching processes that are able to process parts of the snapshot.

For the first prototype we will develop a single matching process that receives a photo taken by the mobile device to find the according video shown on a public display using the creation time and the user's position. The matching process needs all videos potentially shown on one of the public displays. Scale Invariant Feature Transform (SIFT) keypoints [11] are extracted from the

video frames and stored in a repository. When a photo is delivered all public displays nearby the provided position are asked for a list of videos played at a timeframe around the snapshot's time of creation. The resulting video frames are compared to the photos SIFT keypoints using the Best Bin First algorithm [12].

The same matching process can be used to match photos to other visual data, e.g. posters or presentation slides. Additional matching processes will be developed, for example, for face recognition or music matching. We estimate that with the use of additional context information and users preferences the matching precision could be enhanced, for example, by assuming that it is more likely that users interested in specific topics bookmark items belonging to this topic more often.

Each item that could be a result of a match is annotated with a list of links to digital information and services. These lists are returned to the mobile device where the lists are combined with the snapshot of the context to create a contextual bookmark.

5.3 Exchanging Bookmarks

When the contextual bookmark is created from the snapshot of the context and links to digital information and services the user can use the bookmarks to activate the links on his mobile device, transfer them to private devices, for instance to show the movie belonging to a bookmarked trailer on his TV, or send the bookmark to friends and colleagues.

To send contextual bookmarks to another user we will provide a server based solution that enables the users to virtually find each other. To ease the handover to nearby devices and friends we will also implement a Bluetooth based solution that scans the surrounding for other devices running our system and shows them to the user. The users can select nearby devices to easily exchange bookmarks. This solution can also be used to for location based services, for example, to equip advertisement posters with the ability to proactively provide the environment with additional digital information.

6. Conclusion

In this paper we presented our vision of a system that narrows the gap between the digital and the physical world. Users can create contextual bookmarks using everyday devices like their personal mobile phone. We presented a scenario that motivates and justifies our work. Based on this scenario and on the analysis of the related work in this area, we outlined an architecture and its implementation that simplifies the integration of further matching techniques.

The idea presented in this paper can be extended by further matching techniques to address more types of media and real world situations. The overall aim is that the user can create contextual bookmarks of almost any real world item without much effort. For this purpose it is also necessary to address the interaction with the user. Therefore, we will explore how the user expresses the intended action when creating a contextual bookmark.

7. ACKNOWLEDGMENTS

This paper is supported by the European Community within the InterMedia project (project No. 038419). We thank all partners for sharing their ideas with us.

8. REFERENCES

- [1] Hansen, T.R., Eriksson, E., and Lykke-Olesen, A., "Mixed Interaction Space – Designing for Camera Based Interaction with Mobile Devices", Conference on Human Factors in Computing Systems, pp. 1933-1936, 2005
- [2] Ballagas, R., Borchers, J., Rohs, M., and Sheridan, J.G., "The smart phone: a ubiquitous input device", IEEE Pervasive Computing Journal, vol. 5, pp. 70-77, 2006
- [3] Ailisto, H., Korhonen, I., Plomp, J., Pohjanheimo, L., and Strömmer, E., "Realising Physical Selection for Mobile Devices", Mobile HCI 2003 Conference, Physical Interaction (PI03) Workshop on Real World User Interfaces, Udine, Italy
- [4] Rukzio, E., Wetzstein, S., and Schmidt, A., "A Framework for Mobile Interactions with the Physical World", Proceedings of Wireless Personal Multimedia Communication (WPMC'05), 2005
- [5] Raj, H., Gossweiler, R., and Milojevic, D., "ContentCascade incremental content exchange between public displays and personal devices", International Conference on Mobile and Ubiquitous Systems: Networking and Services, pp. 374-381, 2004
- [6] Lange, B.M., Jones, M.A., and Meyers, J.L., "Insight lab: An immersive team environment linking paper, displays, and data", Conference on Human Factors in Computing Systems, pp 550-557, 1998
- [7] Rekimoto, J. and Nagao, K., "The world through the computer: computer augmented interaction with real world environments", ACM symposium on User interface and software technology, pp 29-36, 1995
- [8] Cheverst, K., Dix, A., Fitton, D., Kray, C., Rouncefield, M., Sas, C., Salsis-Lagoudakis, G., and Sheridan, J.G., "Exploring Bluetooth Based Mobile Phone Interaction with the Hermes Photo Display", Conference on Human computer interaction with mobile devices and services, pp. 47-54, 2005
- [9] Cheverst, K., Dix, A., Fitton, D., Kray, C., Rouncefield, M., Salsis-Lagoudakis, G., and Sheridan, J.G., "Exploring Mobile Phone Interaction with Situated Displays", PERMID Workshop, Pervasive, vol. 5, 2005
- [10] Föckler, P., Zeidler, T., Brombach, B., Bruns, E., and Bimber, O., "PhoneGuide: Museum Guidance Supported by On-Device Object Recognition on Mobile Phones", Conference on Mobile and ubiquitous multimedia, pp. 3-10, 2005
- [11] Lowe, D.G., "Object recognition from local scale-invariant features", Proceedings of the international Conference on Computer Vision, pp. 1150-1157, 1999.
- [12] Beis, J. S. and Lowe, D.G., "Shape Indexing Using Appropriate Nearest-Neighbour Search in High-Dimensional Spaces", Proceedings of the 1997 IEEE Conference on Computer Vision and Pattern Recognition, pp 1000-1006, 1997.
- [13] Want, R., Fishkin, K.P., Gujar, A., and Harrison, B.L., "Bridging physical and virtual worlds with electronic tags", Conference on Human Factors in Computing Systems, pp 370-377, 1999
- [14] Want, R., Hopper, A., Falcao, V., and Gibbons, J., "The active badge location system", ACM Transactions on Information Systems, pp 91-102, 1992

Bluetooth Familiarity: Methods of Calculation, Applications and Limitations.

Barry Lavelle, Daragh Byrne, Cathal Gurrin, Alan F. Smeaton, Gareth J.F. Jones
Centre for Digital Video Processing

Dublin City University

Dublin 9, Ireland

cathal.gurrin@computing.dcu.ie

ABSTRACT

We present an approach for utilising a mobile device's Bluetooth sensor to automatically identify social interactions and relationships between individuals in the real world. We show that a high degree of accuracy is achievable in the automatic identification of mobile devices of familiar individuals. This has implications for mobile device security, social networking and in context aware information access on a mobile device.

Categories and Subject Descriptors

H5.3 [Information Interfaces and Presentation]: Group and Organizational Interfaces

General Terms

Algorithms, Bluetooth, Familiarity.

1. INTRODUCTION AND BACKGROUND

Bluetooth is a short-range wireless protocol by which enabled devices can exchange content. It is increasingly routinely included in a wide variety of devices from home computers to portable laptops, PDAs, mobile phones, keyboards, mice and headphones. Bluetooth, although originally devised to support content exchange, can be used for much more than this, for example, Bluetooth familiarity. By monitoring the presence of nearby devices Nicolai [11,12] popularised the concept of familiarity within the Bluetooth space and the work demonstrated that social context could be drawn from general encounters with devices.

He suggests that there are the main types of devices that are encountered: '*familiar*' devices, '*familiar stranger*' devices and '*strangers*'. A *familiar* device is one belonging to a familiar individual who typically form the core social group within a person's life, friends, family, work colleagues, etc.. A *stranger* device typically belongs to an unknown, or a rarely encountered individual, who is outside the friends/family/work-colleague group. Finally *familiar strangers* (a category proposed originally by Milgram [10]) are devices owned by individuals that you encounter on a somewhat regular basis, for example, people who eat in the same restaurant at lunchtime. Typically you will never have interacted with these people.

In this paper, we discuss the collection of Bluetooth sensor data to determine co-present device familiarity. First, we extend Nicolai's work and outline a more robust mechanism for automatically calculating a measure of familiarity for encountered Bluetooth-enabled devices. We demonstrate this technique to be effective at rating familiarity for encountered devices through experimental

means. Bluetooth familiarity has a wide range of applications, relevant to real world mobile interactions including: context-aware information retrieval and content delivery; social networking, and privacy and security for Bluetooth interaction. We explore these applications and as well as some of the limitations of familiarity in the final sections.

2. FAMILIARITY DETERMINATION

In order to automatically assign each encountered device into one of the 3 established familiarity categories we calculate a cumulative score. This score represents a device's presence relative to the other encountered devices and is determined based on duration of presence during short periods each day.

Previous work on familiarity determination [11] only employed a basic metric to determine familiarity. In Nicolai's approach a device becomes familiar after only 5 encounters. While such an approach works well for short periods, it does not scale nor does it provide a realistic determination of familiarity in periods longer than a week. Employing this approach in the long-term allows 'strangers' to become erroneously classified as familiar after a small number of infrequent encounters. Our approach to determination seeks to overcome these scalability issues, be more robust and to provide a more accurate representation of device familiarity, useful in more operationally realistic timeframes.

In our technique, each day is divided into short intervals for which a presence score of each encountered device is calculated. Short intervals are used to provide a more representative and comparative measure and to allow for differences in recording span of days or any gaps in data where a logger might be disabled which could affect. The use of intervals will also allow for the calculation to account for varied device discovery frequencies. (e.g. polling every minute would use a larger interval size compared with 10 second polling). The resulting interval scores for a day are then summed and added to previous day's scores, providing a cumulative measure of a device's overall presence. An advantage of using this approach is that the scores do not have to be recalculated for all data in the set but only additions to the set.

2.1 Familiarity Calculation Technique

With the availability of a presence score a determination on familiarity may be made. For this, we use dynamic thresholds to set a point at which a device transition between familiarity categories. Initially a device will begin as a 'stranger' and then as a result of increases in their cumulative score can become 'familiar' or a 'familiar stranger'.

Using a static threshold, as in [11], presents problems as a stranger may become familiar after a number of short infrequent encounters over a long enough timeframe. To prevent this, we utilise a small dynamic cumulative threshold in combination with a base threshold to prevent strangers becoming familiar through longevity of encounters. With each additional encounter the cumulative threshold is incremented with a small value. As such to become considered ‘familiar’ a device must pass the sum of the static and cumulative thresholds. The static threshold marks the point at which a ‘stranger’ becomes a ‘familiar stranger’.

Our technique is summarised in the below formula:

$$CS_d = \sum_{i=0}^I \frac{F_d}{AVG(F)} \times \frac{T_i}{T_d}$$

$$Familiar = CS_d > \alpha + (\beta \times I_d)$$

$$FamiliarStranger = \alpha < CS_d < \alpha + (\beta \times I_d)$$

where CS_d = Cumulative Score of Device d ; I = a given interval (a specified unit of time at which device presence is determined e.g. 5 mins); F_d = Frequency of encounters of device d within given interval; $AVG(F)$ = Average of all encounter frequencies within given interval; T_i = time in seconds for i ; T_d = time in seconds for day d ; α = Static Baseline threshold ; β = Cumulative Dynamic Threshold; I_d = Total intervals where device d is present.

2.2 Experiment

In order to validate our approach to determining familiarity among Bluetooth enabled devices, we collected Bluetooth sensor data using a proprietary Java ME Bluetooth logging application which employs the JSR-82 API for device discovery. During operation, the application polls for other Bluetooth enabled devices every 10 seconds. For each encountered device the following data is gathered; the device’s hardware address (acting as a unique device identifier), the current ‘friendly-name’ and a timestamp of the encounter. The hardware address is used to map a particular device to a user. Periodically the application automatically uploads data via a GPRS connection to a database. By running this application and through the Bluetooth sensor on the mobile device, it was possible to gather the device interactions for an individual device over an extended period of time.

Once the Bluetooth sensor data was collected and using the hardware address as a unique identifier, the encounters for each device were extracted and segmented into appropriate intervals. The familiarity score for each device was then calculated using this data as explained above.

2.2.1 Experimental Procedure

In total, six participants were involved in the experiment, each of which ran the logging application on their phones at all times the phone was operational for a period of 24 days. Typically these people worked in the university environment (regular hours) and interacted with people at work and outside of work during the course of a day. No incentive to participation was made, however any costs for upload of data incurred by the participants were covered. Participants of the experiment were asked to start the logging software when they turned on their mobile devices and to

go about their daily routines keeping the phone with them as normal.

On completion of the data gathering process, the participants were presented with an exhaustive web-based list of all the (unique) devices they had encountered, ordered from most encountered to least frequent. Users were asked to select those devices they considered to be familiar. A list of all friendly-names for each device was provided to help in the judgment. From this, we formulated a ground-truth data set for each user, against which we could evaluate the results of our familiarity calculation technique.

2.2.2 Statistical Analysis

From the six participants 210,529 individual device encounters were recorded over a 24 day period. On average, each participant had 35,088 device encounters (min. 6,162, max. 62,803) recorded for 1204 unique devices (min: 243, max: 2233) and recorded for 15 days (min. 10, max. 24). Interestingly, we found the number of encounters for each unique device follows a power-law natural logarithmic distribution as shown in Figure 1.

Power-laws are used in mathematics when one wishes to relate one quantity to the power of another. A power-law implies that small occurrences are extremely common whereas large occurrences are extremely rare [9]. Our initial findings show this is coherent with the familiarity scores of users found for each participant’s social network. This finding is in line with what we expected, as typically a user will know only a relatively small group of people very well and the remainder of people encountered over a day will be largely unknown. This natural distribution of devices is also important as it verifies that suitable thresholds can be dynamically calculated as the Bluetooth data set expands. It is also possible that we could use this distribution to tune thresholds to individual differences. This dynamic elicitation of thresholds will become part of our future work.

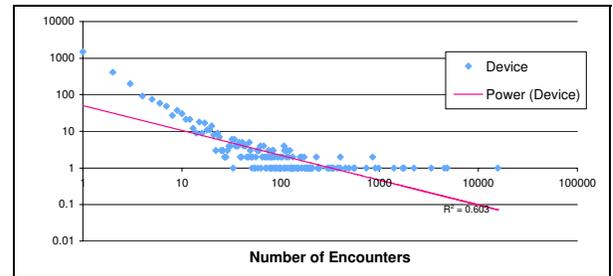


Figure 1. Natural distribution of devices. Only a small number of devices have a high number of encounters while a large number of devices have a small number of encounters.

2.3 Results & Evaluation

As stated, using the time, duration and number of encounters during each day, a familiarity score for each device was calculated. The allocated score was relative to the other devices encountered at given intervals over the course of the day. Precision and recall figures were then calculated against the user specified ground-truth and the effectiveness of our approach measured, though only precision figures are presented here.

To assess the impact that different granularities of time had on the results, familiarity scores were calculated for each device at a range of intervals (including 1 day, 6 hours, 2 hours, 1 hour, 30

minutes, 15 minutes and 5 minutes) to examine the affect of interval size on the accuracy of the resulting familiarity scores. The resulting scores were then sorted in descending order. Precision performance at 20 devices is shown below in Figure 2. Interestingly we found the interval size used had only a marginal effect on the performance of the familiarity scoring for the most familiar devices. Additionally we noted the 5 and 15 minutes offer marginally optimised performance.

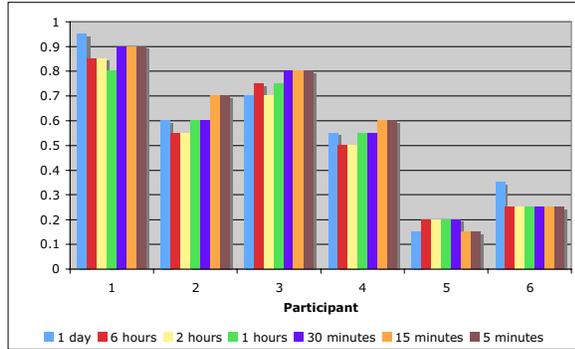


Figure 2: Precision@ top 20 devices

The precision figures for participants 5 and 6 were found to be considerably lower. This can be attributed to the fact that the volume of data recorded was considerably less than the other participants. After analysing their data, we found that they logged infrequently and as a result data was captured sporadically. Based on the analysis of our data we would estimate that between 15,000 - 20,000 device encounters are required before we can reasonably identify friendly devices. The results of this evaluation demonstrate that our methods offer a robust and accurate to determining familiarity of Bluetooth-enabled devices within more operationally realistic timeframes.

3. APPLICATIONS OF BLUETOOTH FAMILIARITY

3.1.1 Security & Privacy

Bluetooth has proven effective as a means of exchange of information, however, it is not without issues of security and privacy. We believe that Bluetooth familiarity can be employed to effectively mitigate against a number of these.

We have previously demonstrated that there is a large overlap in the Bluetooth 'friendly names' used to label devices, with as much as 25% of names overlapping [7]. Typically when exchanging information between Bluetooth enabled devices, the device discovery is socially mediated so confusion or problems in device identification resulting from name overlap can easily be overcome. One can, however, envisage times where it would be helpful to quickly identify how well the person is known to you in order to facilitate faster and more accurate identification of known devices. In Figure 3 we provide an example of how Bluetooth familiarity might be used to aid in this process.

Another issue where familiarity may be of use is 'bluejacking', a simple exploitation used to send unsolicited messages or files to Bluetooth-enabled devices [14]. This may provide a route by which mobile phone viruses can propagate or by which malicious users can gain full access to all content on the recipient's device [2]. However, this can easily be thwarted through familiarity.

Presenting the familiarity score of the sender when a new Bluetooth file transfer request or message is received would enable the device owner to make a better determination of the appropriate action (see Fig. 3). It could also enable the automatic rejection of messages from unfamiliar sources and thereby prevent Bluetooth spamming, which is increasingly being employed in advertising [5] or unwitting access to malicious content [2,14].

Familiarity could similarly be employed within social applications to restrict access to personal phone-based content. For example, BlueTuna [1], an application that enables music recommendations between collocated Bluetooth devices mobile phones, could allow users to restrict access of the music and recommendations to only the most familiar devices .



Figure 3. Examples of how familiarity can be employed in current Bluetooth device discovery and message sending.

3.1.2 Social Networking

Using mobile phone context data to infer social relationships is an established concept. Eagle et al. [4] have demonstrated that mobile device data such as application usage, cell tower IDs (i.e. current location), and proximal Bluetooth devices can be used to determine social patterns in daily activity. Bluetooth familiarity is particularly useful in determining social networks of users as our familiarity calculation offers the ability to determine a devices' relative importance based on their frequency and duration of encounters. This can then be used to map the relationships between devices, potentially across multiple users and devices. The resulting social network could be used in a variety of mobile applications, particularly Bluetooth socialising and dating software for example Nokia Sensor [13].

3.1.3 Context Awareness

Typically applications are unaware of the context in which they operate, however, the use of context data in information access promises increases in the effectiveness, ease and relevance of information retrieval. Applications can be made 'context-aware' "by exploiting the wide range context data available describing the environment, the searcher and the information itself" [6]. These applications seek to automatically and continuously mine sources of context data for information access, examples of these might include: the location of the user; sensors such as for ambient temperature; biometric or body media devices; and application use and activity. With the widespread availability of mobile devices, wireless access to the internet and the availability of contextual cues from mobile devices, context-aware applications are particularly relevant to the mobile space [3].

Realising context-aware retrieval is however not without its challenges, but we believe that Bluetooth combined with familiarity solves a piece of the puzzle: namely, people-awareness. By continually monitoring Bluetooth activity and

associating proximal devices with a familiarity score in real time, the relevant people (determined by familiarity) and the social situation of the owner could be inferred at any given time (e.g. home alone or out with friends). Using this information it will be possible to adapt, alter and adjust the type and amount of information presented to the user in real time. Not only can system output then be tailored but the presentation and interface can also be adapted to the context of use. Combining a real-time understanding of Bluetooth co-presence and familiarity with other sources of context information such as GSM or GPS location or biometric context, will further increase the power, scope and usefulness of Bluetooth-enabled context-aware applications and particularly so in domains such as human digital memory and ubiquitous computing.

4. LIMITATIONS OF BLUETOOTH FAMILIARITY

Bluetooth signals can easily penetrate walls and ceilings without degrading and they also operate in a 3-dimensional plane. This means Bluetooth-enabled devices can be detected in rooms above, below and adjacent to the owner and this presents a challenge for Bluetooth familiarity. From the human perspective, in order for a person to become familiar to us, they must spend time physically proximal to us. Someone sitting in an adjacent room while we are in another will not become familiar to us. Unfortunately due to the nature of Bluetooth, devices in nearby rooms can quickly gain familiarity and there is no easy means to negate against this. This subtle difference between the human and the machine, in what is understood to be proximal has a mild impact for familiarity. Consequently, it can be expected that a small number of devices not personally known by the device owner may be considered extremely familiar when using Bluetooth interaction and co-presence data. These devices typically belong to those working in the same building but not the same physical space as the owner. This difference in the concepts of personal and Bluetooth familiarity should be born in mind when applying our work.

We realise that the collection of Bluetooth activity and presence information raises concerns over privacy. Our work is academically motivated with consent provided by all participants and is designed only to demonstrate the potential of such information. Suitable consideration for privacy should be made when employing Bluetooth presence and familiarity, particularly, if information is not processed or stored on the mobile device, which carries out the logging.

5. CONCLUSIONS AND FUTURE WORK

In this paper we have presented an approach for utilising a mobile device's Bluetooth sensor to automatically identify social interactions and relationships between individuals in the real world. By means of a user experiment, we have shown that by simply running a Bluetooth logger on a mobile phone and analysing interactions with other Bluetooth enabled devices, that an accuracy rate of 90 percent can be achieved in detecting the devices of familiar individuals. This finding has implications for mobile device security, social networking and the application of context awareness.

The focus of our current evaluation has been on establishing the effectiveness of the familiarity scoring for encountered devices and we recognise that additional work is required to determine appropriate thresholds. In our future research, we plan to examine

the most appropriate means to establish and optimise the thresholds for familiarity determination. In addition we hope to explore if thresholds need to be tailored to an individual or if a single threshold will generally apply to all devices.

6. ACKNOWLEDGMENTS

We would like thank all members of the CDVP who participated in the experiments. This work was partially supported by the Irish Research Council for Science Engineering and Technology and by Science Foundation Ireland under grant 03/IN.3/I361.

7. REFERENCES

- [1] Baumann, S., Jung, B., Bassoli, A. and Wisniowski, M. BluetunA: let your neighbour know what music you like.
- [2] Bialoglowy, M. Bluetooth Security Review, Part 1. (Apr. 2005) Retrieved from: <http://www.securityfocus.com/infocus/1830>
- [3] Brown, P.J. & Jones, G.J.F., Exploiting contextual change in context-aware retrieval, Proceedings of the 2002 ACM symposium on Applied computing, Madrid, Spain, pp. 650 - 656, 2002
- [4] Eagle, N. and Pentland, A., Reality Mining: Sensing Complex Social Systems, Personal and Ubiquitous Computing, Volume 10, Issue 4, 2006
- [5] Graham-Rowe, D. Billboards beam adverts to passing cellphones. Aug 2005. Retrieved from: <http://www.newscientist.com/article.ns?id=dn7883>
- [6] Jones, G.J.F. Challenges and Opportunities of Context-Aware Information Access, International Workshop on Ubiquitous Data Management, pp. 53-62, 2005
- [7] Lavelle B., Byrne D., Jones G.J.F., Smeaton A.F. Bluetooth Friendly Names: Bringing Classic HCI Questions into the Mobile Space, British HCI 2007, Lancaster, UK, September 2-5 2007.
- [8] Marcus, A., Ferrante, J., Kinnunen, T., Kuutti, K. and Sparre, E. Baby faces: user-interface design for small displays. ACM SIGCHI, Los Angeles, CA, Apr 18-23, 1998.
- [9] Mitzenmacher, M. A Brief History of Generative Models for Power Law and Lognormal Distributions, Retrieved from: <ftp://ftp.deas.harvard.edu/techreports/tr-2001.html>, 2001
- [10] Milgram, S. The Familiar Stranger: An Aspect of Urban Anonymity. The individual in a social world, pp. 51-53. Reading, MA: Addison-Wesley, 1977
- [11] Nicolai, T., Behrens, N. and Yoneki, E. Wireless Rope: Experiment in Social Proximity Sensing with Bluetooth, PerCom 2006, Pisa, Italy
- [12] Nicolai, T., Yoneki, E., Behrens, N. & Kenn, H. Exploring Social Context with the Wireless Rope. In Proceedings of MONET'06, Montpellier, France.
- [13] Nokia Europe. Nokia Sensor. Retrieved from: <http://europe.nokia.com/A4144923>, 2005
- [14] Thom-Santelli, J., Ainslie, A. and Gay, G. Location, location, location: a study of bluejacking practices. In Extended Abstracts of CHI 2007. (San Jose, USA, April 2007). ACM Press, New York, NY, 2693 - 2

Semantic Integration and Language Access to Mobile Data

Raimondas Lencevicius
Nokia Research Center Cambridge
3 Cambridge Center
Cambridge, MA 02142

Raimondas.Lencevicius@nokia.com

Alexander Ran
Nokia Research Center Cambridge
3 Cambridge Center
Cambridge, MA 02142

Alexander.Ran@nokia.com

ABSTRACT

Real-world data can significantly enhance the functionality of mobile services. For this, real-world data needs to be collected, stored and integrated with other information available on mobile devices. A flexible and user-friendly interface to the data is also needed. This paper describes an experience in collecting real-world data and integrating it into a semantic data repository. We use an innovative Natural Query language and engine to automatically connect the resulting repository to the natural language user interface. The resulting system on S60 mobile platform successfully answers user questions about Personal Information Management (PIM) data extended with real-world data.

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval], H.5.2 [User Interfaces]: *Natural language*

Keywords

Data access, query language, natural language.

1. INTRODUCTION

Mobile devices make a perfect user interface to the real-world environment. They are constantly carried with the user [2] enabling gathering of user location information. Mobile devices are equipped with more and more sensors including GPS receivers, Bluetooth transmitters and receivers, RFID receivers and others. They also receive and store information about such real-world events as messages, phone calls, meetings, application usage and access to digital services. It is therefore natural to expect that this real-world data should be collected and made accessible on mobile devices. However, there are some open questions that need to be resolved in order to make this kind of data useful and accessible both to the programs and to the mobile device users. Collected real-world data must be structured and integrated with other information available on mobile devices such as, for example, the information found in the user's phone book or calendar. There also needs to be an intuitive interface that allows flexible access to collected information.

In this paper we present a framework that collects real-world data, structures it according to an extended PIM ontology, augments and integrates it with earlier collected data, and stores it in an RDF repository.

To access the data, an intuitive interface is needed. Natural language based information access is increasingly viewed as a promising alternative to graphical user interfaces (GUIs), especially in the domain of mobile devices. We have developed a Natural Query language and engine that can automatically map

meaning representation produced by language systems into formal database queries. This enables us to provide a natural language interface to the integrated real-world and on-device data.

The paper describes our data gathering framework (Section 2) and explains our solution to data storage (Section 3). Natural Language Interface to the stored data is simplified via using Natural Query system (Section 4). Then we describe our experience with the system (Section 5). We finish with the related work and conclusions.

2. DATA GATHERING

Data collection on mobile devices is an active field of research [3][8]. We have extended one of the frameworks available within Nokia to collect events that occur on a mobile device: phone calls, SMS messages, nearby Bluetooth devices, and GSM locations. All of these events are tagged with a timestamp when they occur. For phone calls the device records the phone number called (or the phone number that called the phone user) and the call length. For messages, the phone number and the message text is recorded. A GSM location change event is recorded when the cell tower associated with the phone changes. Finally, the phone periodically scans for Bluetooth devices in its vicinity and records their names and IDs.

Although the real world data gathered is interesting by itself, it becomes even more important when connected to the data already available inside the device. Mobile devices store a rich set of structured information. The address book or phone book application contains names, phones, addresses and affiliations of personal contacts. The calendar application contains entries for meetings with participants, meeting location and time. All these data are related. Retrieving these data based on their relation could be very useful for device owners. With such retrieval capabilities they could learn who called them when they were in California, or when is their next meeting with Ann from Accenture. Unfortunately, the relations between different data items are not explicit when the events occur or information is entered in some application. Therefore it is important to integrate the collected data by explicating its relation to data available on the device. To achieve this goal, we have developed an extended PIM ontology that covers all relevant types of information available on the mobile device: from observed events, information from external data stores, to on-device data from several mobile applications. Once the data was structured and augmented with relations, it is stored in RDF [10] repository.

3. STRUCTURED DATA STORAGE AND ONTOLOGY

We created the PIM ontology to cover all data available in the device. We considered using such standard ontologies as W3C

foaf [4] and vcard [15]. However, the information available on the mobile device was richer than the types supported by standard ontologies. Main classes in our ontology are *Person*, *Organization*, *CalendarEntry*, *EmailAddress*, *Location*, *Observation*, *Message*, *Call*, and *PhoneNumber*. Part of the ontology is shown in Figure 1.

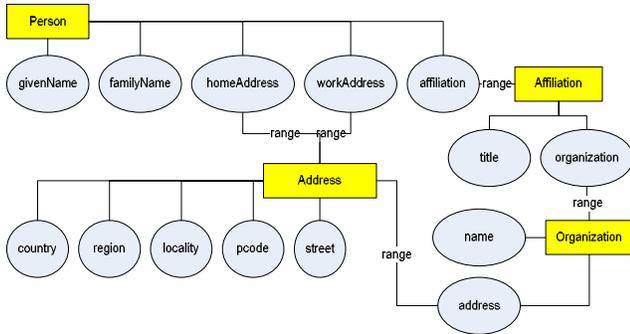


Figure 1. Part of Mobile PIM ontology

There are about ten more secondary classes and class attributes. The real-world data observations are stored in the objects of *Observation* subclasses: *BTDeviceObserved*, *CallObserved*, *MessageObserved*, and *LocationObserved*. The attributes of these objects connect with other objects of the repository. For example, the *phoneNumber* attribute of a *CallObserved* is of type *PhoneNumber*, which is also used in the attribute *phoneNumber* of a *Person* or *Organization* class. Therefore the gathered real-world data directly integrates with the on-device data.

For some other data, programs or users have to add information to facilitate integration. For example, a Bluetooth device ID and name attributes have to be added to the *Person* class and filled in with concrete values in order to associate the *BTDeviceObserved* observation to a specific person carrying a Bluetooth device.

Another area where observed data integrates with on-device data is the location information. A significant part of ontology deals with locations at various granularity levels: from meeting rooms, to office buildings, cities, and countries. We use the *partOf* relation between different objects to represent geographic or organizational inclusions. For example, a relation can indicate that Boston is a part of Massachusetts, which in turn is a part of the USA. This attribute is also used to describe the GSM location containment within a certain geographical object. Since GSM locations are somewhat imprecise, we have chosen to associate them with town or city level geographical entities. This provides sufficient information in most cases.

Overall, we found that our RDF repository is significantly more flexible than a relational database. For example, it naturally supports multiple classes of contacts, multiple affiliations per person, and supports a sophisticated typing system.

4. NATURAL LANGUAGE INTERFACE

Although the repository of integrated real-world and in-device data can be used in variety of ways, for example, via querying it using SPARQL [14], we were interested to provide an intuitive and flexible user interface to it. We decided that a

general natural language interface to a rich data set could be more effective than a GUI based application.

As a rule, information bases and language systems are developed independently of each other. Therefore information bases are not designed for interaction using natural language and their integration process is mostly ad hoc, manual process. Figure 2 is a sketch of a typical architecture that is used to provide a natural language interface to databases and other back-end or native services.

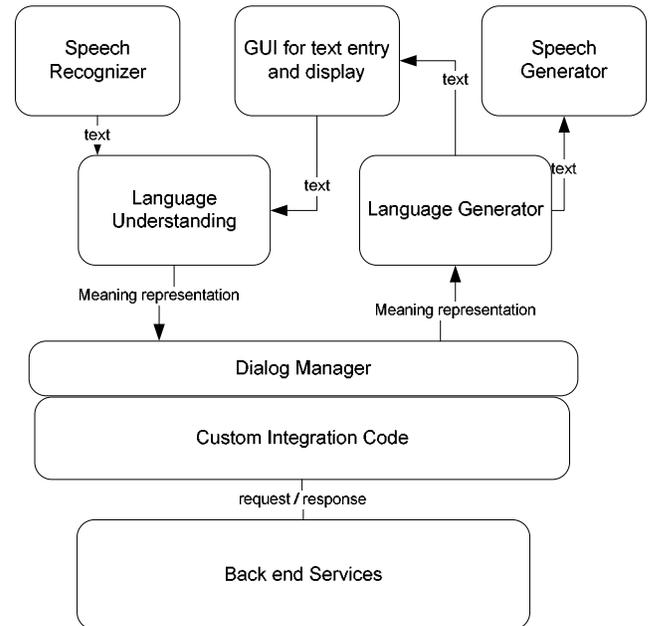


Figure 2. Architecture sketch of Natural Language Interface to Services

The speech recognition and generation components translate between text and speech modalities. The language understanding component converts the text into a formal representation of meaning sometimes called *semantic frame* [12]. The language generation component converts the formal meaning representation to a natural language text [1]. The dialog manager uses the context of conversation to complete frames received from the language understanding module or created by the custom integration code from responses of backend services. The custom integration code also translates meaning representation frames it receives from the dialog manager into a standard database query or backend specific API requests.

We have designed and implemented the Natural Query (NQ) language and engine [9] that removes the need for custom integration code. NQ can automatically map meaning representation produced by language systems into precise database queries. NQ employs two mechanisms: language tags and data graph search to return requested data using only the information in the meaning representation of the user request.

Language tags are words, expressions, and linguistic tokens attached to database elements such as classes and properties. Multiple tags can be attached to a single element and a single tag can be attached to multiple elements. Language tags are the names of the corresponding categories used by the language system(s).

Figure 3 illustrates language tags associated with a part of our PIM ontology. A generalization like “Contact” can be attached to specific classes like “Person” and “Organization”. Tags like “in” can be attached to all location elements. A general reference like “Name” can be attached to multiple elements like “givenName”, “familyName”, and so on. In our RDF repository of real-world and in-device data, we added language tags to the RDF objects using a subproperty of RDFS label field.

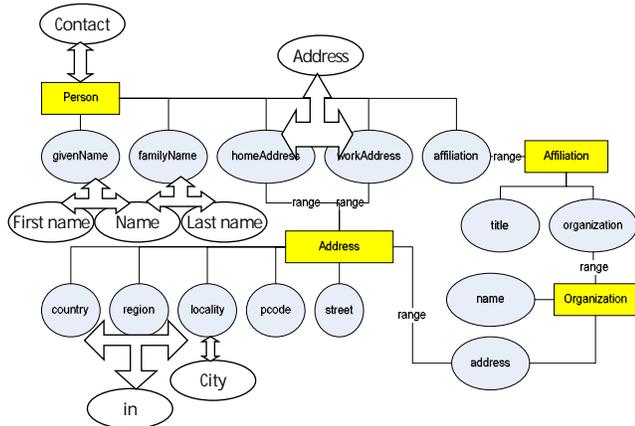


Figure 3. Language tags for database elements

While ad hoc integration needs to have the information about the organization of the database, NQ avoids the need for such information by using a graph search to achieve the same objective. Given a question “Who are my contacts at IBM in Ulm?”, NQ finds paths connecting the nodes known from the meaning representation, such as “Person”, “name”, “Organization”, “City”, “Ulm”, and “IBM”. One of such paths is highlighted in Figure 4.

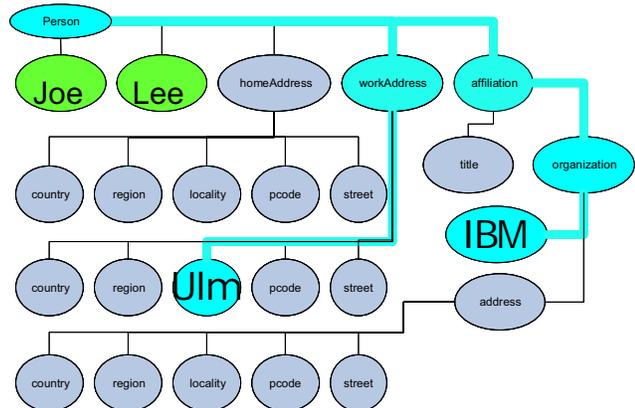


Figure 4. Answering query via graph search

We have created a proof of concept implementation of NQ in Python [7] that runs on S60 [11] mobile phones. Full description of the Natural Query system is presented in [9].

5. EXPERIENCE WITH THE SYSTEM

We tested our system on a PIM test data set containing 550 contacts with about 150 meetings and 250 phone calls, which is normal for executives with a lot of contacts and meetings. The repository contained over 11000 RDF triples. We asked over 50 natural queries corresponding to over 600 parameterized

questions. We did not count various parameters, such as different cities or names, since these numbers can be made arbitrarily large and the resulting number of questions does not really reflect the capabilities of the system.

The system can answer questions ranging from “What is the email of John?” to “Where does Ann work?” to “My meetings next week with John” and “Who called me yesterday”. Some of these questions would convert to quite complex relational or SPARQL queries. For example for the query “Who called me yesterday”, we need to find all telephone numbers of calls that occurred yesterday and then find all people who have these telephone numbers. NQ query for this is very simple: “fromClass = 'Person', select = ['givenName', 'familyName'], where = [(["ReceivedPhoneCall", 'start'], TimeInterval('yesterday')])”.

If we classified questions according to domains, one domain would contain questions about the personal information data from an address book application, for example “Who works as a real estate broker?”. Another set of questions is about meetings, for example, “When are my meetings next month at MIT?”. Yet another set is about calls and messages, for example, “Who called me last Friday?”. Finally there are questions spanning multiple domains, for example, “What are emails of people who participated in a meeting on Monday?”, “Who called me when I was in Finland?”, and so on.

All these types of queries (Figure 5) were successfully created and executed on the extended PIM data store.



Figure 5. Example questions

We found out that we could easily ask questions both about the in-device data and the collected real-world data. Integration of the two enhanced our question answering capability significantly, allowing such questions as “Who called me when I was in Helsinki?”, “Which messages did I receive during the meeting with Juha?”. Although the detection of someone’s Bluetooth device is a weak indication the phone user met other person with a Bluetooth device, in our experiments we assumed such implication. This allowed us to ask questions such as “Who did I meet last week?”.

Test NQ queries mostly returned expected answers (96% recall, 92% precision) (Figure 6) including the approximate answers where the exact answers were not available. For example, the question “When was my meetings with Sam last month?” had no exact answers, so the system returned approximate answers of

meetings with Sam that did not occur last month as well as the meetings that occurred last month, but did not include Sam.

The performance of the system was acceptable with answers taking from less than a second to several seconds. The system implementation is a prototype in Python that was not optimized for memory or speed. The detailed evaluation of system performance is outside the scope of this paper. We are planning to optimize the system performance in the near future.

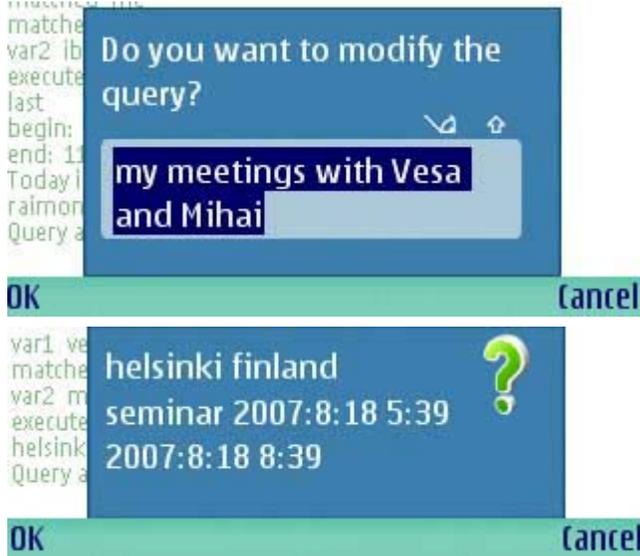


Figure 6. Example question and answer

6. RELATED AND FUTURE WORK

Real world data has been gathered on mobile devices by a number of projects including Context [8] and Reality Mining [3]. In our work, we have extended one of the data gathering frameworks available at Nokia.

Mobile data storage in RDF repositories is investigated by ConnectingMe [5] project at Nokia Research Center. We are collaborating with ConnectingMe in the ontology and repository development.

We have not discovered any research directly corresponding to the Natural Query approach. The Precise system by Popescu et al. [6] attaches language tokens to database elements in a way very similar to language tags of NQ. Also the query derivation approach of Precise is based on database graph search. NQ uses a more flexible data model, supports incomplete answers, and collects data for explanations.

In the future, we plan to connect our system to such natural language and speech systems as TINA [12] and Galaxy [13]. We plan to perform user trials to evaluate our system and its user interface to real world data. We will collect additional data such as email messages, songs listened, and pictures viewed and taken. We will also optimize the current prototype implementation.

7. CONCLUSIONS

Mobile devices are now able to continuously collect real world data and present it to the users. In addition to real world data, mobile devices host structured and semi-structured

information bases. We have demonstrated integration of such data with the collected real world data using a flexible and powerful RDF repository and a common ontology. We have designed and implemented a query language and engine NQ that can automatically map meaning representation produced by language systems into formal database queries. We have used NQ to access extended PIM (Personal Information Management) data on mobile phone. Our experience indicates that real-world data gathering and integration with in-device data, together with a natural language interface is a valuable addition to capabilities of mobile devices.

8. REFERENCES

- [1] Baptist L. and S. Seneff, "Genesis-II: A Versatile System for Language Generation in Conversational System Applications," Proc. ICSLP '00, Vol. III, pp. 271-274, Beijing, China, Oct. 2000.
- [2] Chipchase, J., "Why do People Carry Mobile Phones?," http://www.janchipchase.com/blog/archives/2005/11/mobile_essentia.html, 2005.
- [3] N. Eagle, "Machine Perception and Learning of Complex Social Systems", Ph.D. Thesis, Program in Media Arts and Sciences, Massachusetts Institute of Technology, June 2005.
- [4] FOAF Vocabulary Specification 0.9, <http://xmlns.com/foaf/0.1/>, 2007.
- [5] Lassila, O. et al, "ConnectingMe", <http://research.nokia.com/research/projects/connectingme/index.html>, 2007.
- [6] Popescu, A., Etzioni, O., and Kautz, H. 2003. Towards a theory of natural language interfaces to databases. In Proceedings of the 8th international Conference on intelligent User interfaces (Miami, Florida, USA, January 12 - 15, 2003). IUI '03. ACM Press, New York, NY, 149-157.
- [7] Python for S60, <http://sourceforge.net/projects/pys60>, 2007
- [8] Mika Raento, "Context software - A prototype platform for contextual mobile applications". In Proceedings of the International Proactive Computing Workshop. University of Helsinki, 2004.
- [9] Ran, A., and Lencevicius, R., "Automating Access to Structured Data from Natural Language", Submitted for publication, 2007.
- [10] Resource Description Framework, <http://www.w3.org/RDF/>, 2007.
- [11] S60 platform, <http://www.s60.com>, 2007
- [12] S. Seneff, "TINA: A natural language system for spoken language applications," *Computational Linguistics*, vol. 18, no. 1., pp. 61-86, March 1992.
- [13] S. Seneff, E. Hurley, R. Lau, C. Pao, P. Schmid, and V. Zue, "GALAXY-II: A Reference Architecture for Conversational System Development," *Proc. ICSLP 98*, Sydney, Australia, November 1998.
- [14] SPARQL Query Language for RDF, <http://www.w3.org/TR/rdf-sparql-query/>, 2007.
- [15] Vcard, <http://www.w3.org/TR/vcard-rdf>, 2007.

Koubachi: A Mobile Phone Widget to Enable Affective Communication with Indoor Plants

Philipp Bolliger
bolligph@inf.ethz.ch

Benedikt Ostermaier
ostermaier@inf.ethz.ch

Institute for Pervasive Computing
ETH Zurich
Switzerland

ABSTRACT

We introduce *Koubachi*, a system designed to enable interaction between plants and humans. A plant is equipped with several sensors, which are connected to a wireless sensor node. By performing various measurements, the sensor node can determine the current health status of the plant. The user can access this data at all times through a small mobile phone application, a so called mobile widget. In case the node detects a change in the health status, the mobile widget is brought to the foreground in order to get the user's attention. We implement the principles of affective communication [8] to notify the user about the plant's health status since we believe that this special form of information propagation has various benefits compared to non-emotional communication.

1. INTRODUCTION

Taking proper care of indoor plants can be quite demanding. This is due to the fact that different plants have very different requirements regarding the amount of water, light and fertilizer they need. Furthermore, these demands may even change over the seasons. Thus it can necessitate a lot of time and knowledge to take care of an indoor plant. Our own experiences have shown that it is particularly cumbersome to learn about the plant's actual condition, i.e., if it has enough or maybe too much water, or if it needs more light, especially as we can only learn the plant's condition with a considerable time-lag. If the leaves turn brown, it is often too late. Even more so, a common practice to determine the current soil humidity is to actually stick a finger into the soil, certainly not very accurate and not always a pleasant experience. However, the care itself is a very satisfying activity for most people as it gives the pleasurable feeling of accomplishment. Hence the idea emerged to give indoor plants new abilities to measure their health status and to communicate their needs to humans. In doing so, we want to make use of the fact that most people establish a relationship with their plants while taking care of them. Ideally, a plant should be able to appeal to human emotions in order to call attention. We believe that the simplest yet most effective method to achieve this goal is to use smiles, which are able to express emotions in a very human manner.

We believe that mobile phones are well suited as user interface device for this kind of application. Mainly because they are the most common electronic devices that people voluntarily and constantly carry around. Furthermore, people are used to utilize their phones for communication with other

people, using both synchronous voice calls and asynchronous communication techniques like short messages. Our contribution is that we augment this communication paradigm to enable the communication with things, like plants, via the mobile phone.

In this paper we present the monitoring system and the mobile phone widget that is used to display the health status of a plant. We begin by surveying related work in Section 2. In Section 3, we outline the hardware used to build our prototype sensor board and describe the logic to determine the plant's health status. Further, in Section 4 we present the design concept and the interaction model applied to the user interface. In this section, we also describe the widget's implementation and its functionality. We present our conclusions and make an outlook onto future work in Section 5.

2. RELATED WORK

By today, it is almost a matter of common knowledge that personal mobile devices are most prominent components in many ubicomp systems [6, 7]. Moreover, Roduner et al. [10] have just recently shown the strengths of using mobile phones to interact with everyday objects. Since mobile phones are ubiquitous and people are used to check their phone often, for example to see if they received a new text message, we decided to use this platform in order to communicate the plant's condition.

A project with very similar goals is *Botanicalls* [5], which is carried out at New York University. Much like Koubachi, Botanicalls aims at building relationships between plants and humans. However, this project utilizes conventional telephone calls in order to inform humans about the needs of the plants. Although the telephone system is truly ubiquitous and as such a universal communication medium, we believe that answering or making phone calls in order to get information about a plant's current health status can become rather cumbersome and thus impedes building up a positive relationship. Our goal however was to enable a human-plant interaction that "*deliberately influences emotion*"¹, thus creating an affective way of communication with everyday objects.

The question of the right level of computational interpretation has also been raised in the *Vineyard Computing* project [3]. Using ethnographic research methods, the authors found

¹From the definition of *Affective Computing* by R. Picard[8].

that the users, i.e., the vineyard managers, want to be provided with interpreted information rather than raw sensor data. However, they did not want the system to take action itself. The system should rather present data that suggests a tangible next step, so called *actionable data*. With Koubachi, we pursue a very similar goal. Instead of autonomously watering a plant upon a low soil moisture sensor reading, we provide the user with a simple and self-explanatory image that suggest to water the plant.

3. PLANT MONITORING SYSTEM

The main component of Koubachi is the monitoring system, which allows for the gathering, storage and processing of data that is vital to the plant’s condition. In order to acquire a detailed picture of the plant’s health status, we currently measure four different variables: the moisture of the plant’s soil with a *Decagon EC-5* sensor, the light level using a *Taos TSL252R* sensor as well as the air temperature and humidity with a *Sensirion SHT71*. Figure 1(b) shows the different sensors attached to an actual indoor plant. All of these sensors are connected to a BTnode [4], a wireless sensor node, which allows us to store and process the data and which is able to connect to other devices using Bluetooth. The computation of the plant’s health status is conducted autonomously by the node itself, in order to enable spontaneous notifications to different devices and users.

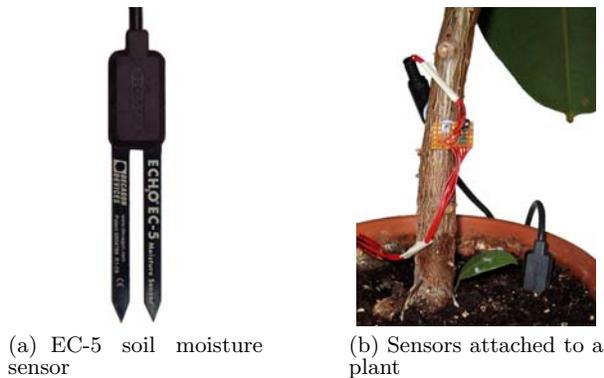


Figure 1: Plant monitoring system components

We use an approach where the plant’s health status is only updated when a critical boundary of one or more sensor values has been exceeded for a specific amount of time. This prevents rapid changes of the visualization and should avoid alarming the user when a suboptimal condition occurs for a short time only, e.g. a temperature drop due to a sudden ventilation of the room. These different parameters can be set individually for each type of plant. For some cases, since we record the data, we are able to indicate a critical situation before it actually happens. For example, we can warn the user that he needs to water the plant in the upcoming hours whenever the recorded soil moisture level shows a characteristic degression. We are also able to detect certain actions conducted by the human, like a watering of the plant. Figure 2 shows the progression of the soil moisture over the course of one week, in which three watering events can be clearly identified.

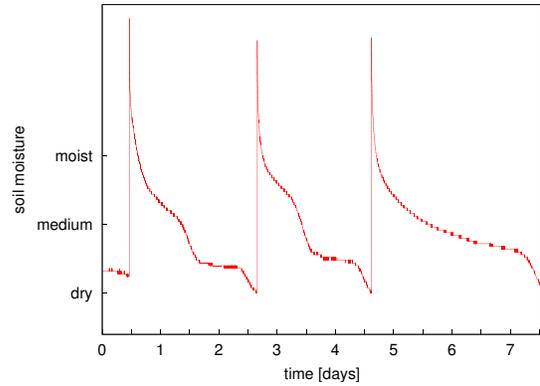


Figure 2: Exemplary progression of the soil moisture

Based on the input values from the sensors, the parameterized demands given by the type of plant that is monitored and the actual time and date, our health function will notify or alarm the user whenever the plant needs to be looked after.

To communicate the health status to the user’s mobile phone, we use the BTnode’s on-board Bluetooth radio. For example, if the user comes home and wants to check the status of his plants, he can start the Koubachi MIDlet² on his mobile phone. He is then presented a list of all his plants. By choosing a plant, he issues a query to the according BTnode, which will subsequently reply with the current status of the plant. Moreover, as the plant monitoring node should be able to autonomously alarm the user whenever a measurement shows a critical level, we store the Bluetooth address of the user’s mobile phone on the BTnode. By doing so, we can always initiate a Bluetooth communication from the BTnode to the mobile phone whenever the phone is in range.

4. USER INTERFACE

We utilize affective communication by giving the plants the possibility to express their needs in the manner of *subtle expressivity* [1]. Thus, we aim to support the process of relationship building between plants and humans. We chose to apply this paradigm because of three main benefits: firstly, affective communication is universal for humans and is therefore not constrained by linguistic or cultural boundaries [8]. Secondly, we believe that emotional communication is a very effective means to motivate humans to do something [1] and thirdly, by using emotions, one is able to build up a relationship with his or her communication partner [2].

Although voice calls seem to be a practical technique to communicate with things and may transport emotions, we do not think that they provide a usable solution – because the interface is synchronous, cumbersome and the repetition of messages may start to annoy the user. On the other hand, message-based communication techniques do not seem to provide a feasible user interface. Hence, we decided to use custom visualizations in order to enable plants to express emotions by utilizing the mobile phone of the user.

²A MIDlet is a Java program for embedded devices.

4.1 Design Principles

We use the well-known “smiley” in order to visualize the plant’s current health status. Such smileys are often used to transport emotions in on-line communication and are therefore well-known amongst users. As they imitate a human face, recognition of the intended message is done by interpreting the facial expression, which is intuitive and not hindered by cultural barriers [9].

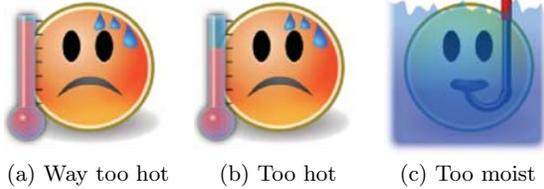


Figure 3: Smileys showing the plant’s health status

In our system, smileys can have three states: (1) A happy face, indicating that all parameters are within optimal values, (2) a neutral face, indicating that some parameters might pass tolerable values in the future, and (3) a sad face, indicating that one or more parameters of the plant show critical values. Furthermore, to advise the user of the concrete problem of the plant, we give additional visual clues. For example, if the plant has run out of water, the smiley will loll its tongue, whereas if too much water was given to the plant, the smiley will be depicted immersed in water, using a snorkel as shown in Figure 3(c). To further denote the severeness of a specific need, we amplify the smileys’ meaning by displaying a three level indicator. Figure 3(a) for example indicates that the plant is in an environment where the temperature is too high. By looking at the thermometer one can tell that the problem is quite severe. The smiley shown in Figure 3(b) on the other hand indicates that the environment is only a little bit too hot. The humorous illustration of the plant’s state is intended as we believe it will help to create a relationship between the human and the plant, much like it happened with the famous Tamagotchi³ toy a few years ago. To make it easier for the user to link the smiley to the actual plant, we do not display the smiley free-standing but draw it on a picture of the plant as illustrated in Figure 4(b). To further increase the level of affection, the user can assign nicknames to his plants.

In order to make looking after the health status of the plant as convenient as possible, we decided to utilize so-called *widgets*, small applications that are continuously running or can be started quickly.

4.2 Interaction

Conventionally, “interaction” between the human and the plant has been the human caretaker providing water and a good environment and the plant responding by exhibiting healthy leaves and flowers. With Koubachi however, we introduce a new approach to human-plant interaction, distinguishing between *plant-initiated interaction* and *human-initiated interaction*. Plant-initiated interaction is always

³Tamagotchi is a hand-held digital pet that needs to be fed and taken care of or else it will die.

based on the Koubachi widget and starts with the activation of a plant’s personal page, i.e., depicting the plant individually in full-screen. This only happens when the status of a plant (happy, neutral or sad) changes. Hence, the user is notified not only when the health status of the plant degrades, but also when it improves, thus avoiding to make a plant-initiated interaction something solely alarming.

By contrast, in human-initiated interaction, the user himself is activating the widget, i.e., he can check the current health status of his plants at any time. Based on information depicted in the widget or on own experience, the human is able to perform the appropriate care of the plants. It is expected that the user learns the needs of his plants with time, thus gradually reducing the number of situations of distress of his plants and therefore the amount of negative feedback from Koubachi.

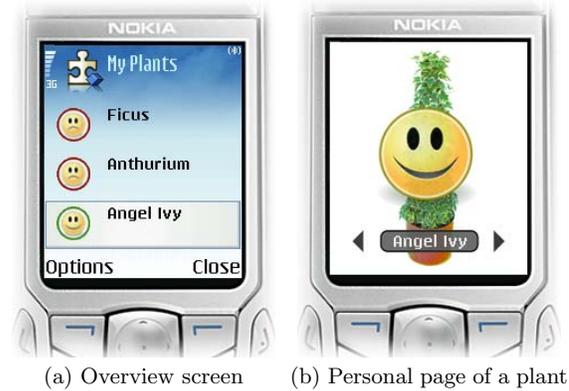


Figure 4: Screenshots of the mobile phone widget

4.3 Widget Implementation

As illustrated in Figures 4(a), 4(b) and 5(b), our widget implementation features three different views: (1) An overview of all plants attached to Koubachi, including their current health status, (2) a *personal page* of a plant, displaying it augmented with a smiley and additional information in full-screen and (3) a *detailed view* of the current sensor readings of a plant. We made the usage of the widget an enjoyable experience by utilizing a simple user interface, including appealing graphics and providing smooth transitions between the views. The widget is expected to run continuously in the background, both to enable plant-initiated interaction and to minimize the time it takes to access the widget.

We chose the concept of personal pages in order to create an emotional form of interaction and to improve usability. By showing each plant by itself, we are able to utilize the entire screen and provide the user a focused view on the subject, thus maximizing the impact of the depicted scene. In addition to that, we display the nickname of the plant below its picture. Personal pages of plants can be cycled through quickly by using the left/right arrow keys. Each personal page has also a “rear side”, containing detailed data for the corresponding plant, which can be displayed using either the middle or the down key.

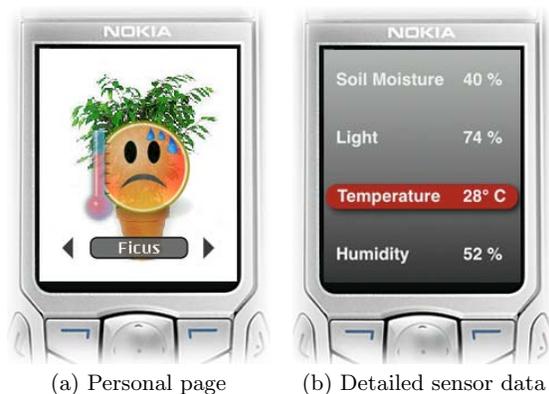


Figure 5: Front and rear side of a plant’s personal page, indicating that the environment is too hot

For user-initiated interaction, the human is presented the overview screen. Here, all of his plants are listed with their nicknames and small smilies indicating their current health status. This way, the user can see the status of all of his plants at a glance. However, to learn about the specific needs of a plant, he can display its personal page since the status of a plant is only indicated by a happy, neutral or sad smiley, without any additional hints. He can do so by selecting the appropriate plant from the list.

For plant-initiated interaction, the widget is automatically activated and shows the personal page of the plant demanding attention the user (Figure 5(a)). If the user is interested in detailed data of the plant, he can instantly switch to the rear side of the personal page, containing the detailed view of this plant (see Figure 5(b)).

5. CONCLUSION AND FUTURE WORK

Building upon the principle of subtle expressivity [1] and using state of the art sensor technology, we developed and built a system that supports taking proper care of indoor plants. By using smilies to communicate the plant’s current status, we support a pleasurable plant-human interaction that encourages the user to properly address the needs of his plants. A small hallway test with six people showed that the usage of the mobile phone widget is not only helpful but real fun. All six test persons were able to operate the widget without a word of introduction or explanation. Moreover, all test persons immediately understood the meaning of the smilies and intuitively knew which action to take.

Since the project is still in its early stages, we are currently improving the visualization by gathering feedback from users. Moreover, we plan to build more sensor boards in order to deploy the system to various different plants. Subsequently, we would like to conduct a larger user study to show the benefits of our approach. We are also planning to connect the sensor node to the internet, in order to allow for a true remote monitoring of plants.

6. ACKNOWLEDGMENTS

We would like to thank Christoph Bani for building the first prototype and creating the artwork. We would also like to thank Marc Langheinrich and Robert Adelman for their thoughtful reviews and advice.

7. REFERENCES

- [1] T. Bickmore and R. Picard. Subtle expressivity by relational agents. In *Proceedings of the CHI 2003 Workshop on Subtle Expressivity for Characters and Robots*, 2003.
- [2] C. Breazeal and B. Scassellati. Infant-like social interactions between a robot and a human caretaker. In *Special issue of Adaptive Behavior on Simulation Models of Social Agents*, 2000.
- [3] J. Burrell, T. Brooke, and R. Beckwith. Vineyard computing: Sensor networks in agricultural production. *IEEE Pervasive Computing*, 3(1):38–45, January 2004.
- [4] J. B. et al. Btnodes - a distributed environment for prototyping ad hoc networks. <http://www.btnode.ethz.ch>.
- [5] K. Hartman, K. London, R. Bray, and R. Faludi. Botanicalls: The plants have your number. <http://www.botanicalls.com/>, December 2006.
- [6] S. S. Intille. Designing a home of the future. *Pervasive Computing, IEEE*, 1(2):76–82, 2002.
- [7] T. Kindberg, J. Barton, J. Morgan, G. Becker, D. Caswell, P. Debaty, G. Gopal, M. Frid, V. Krishnan, H. Morris, J. Schettino, B. Serra, and M. Spasojevic. People, places, things: web presence for the real world. *Mob. Netw. Appl.*, 7(5):365–376, October 2002.
- [8] R. W. Picard. *Affective Computing*. MIT Press, 1997.
- [9] K. Rivera, N. J. Cooke, and J. A. Bauhs. The effects of emotional icons on remote communication. In *CHI '96: Conference companion on Human factors in computing systems*, pages 99–100, New York, NY, USA, 1996. ACM Press.
- [10] C. Roduner, M. Langheinrich, C. Floerkemeier, and B. Schwarzentraub. Operating appliances with mobile phones – strengths and limits of a universal interaction device. In *Proceedings of Pervasive 2007, Toronto, Canada, May 13-16, 2007*, LNCS, Berlin Heidelberg New York, May 2007. Springer.

Swiss Army Knife meets Camera Phone: Tool Selection and Interaction using Visual Markers

Christian Kray
Informatics Research Institute
Newcastle University, UK
c.kray@ncl.ac.uk

Michael Rohs
Deutsche Telekom Laboratories
TU Berlin, Germany
michael.rohs@telekom.de

ABSTRACT

A key issue in ubiquitous computing in general and public display research in particular is how to enable interaction. Oftentimes, it is not clear how users can interact with a system and what functionality it provides. In the case of public displays, several methods have been suggested such as touch-enabled surfaces, gesture recognition, voice input, or text messaging. However, all these methods have some inherent flaws such as being unreliable, limiting the number of concurrent users or requiring complex configuration. In this paper, we introduce a novel approach based on visual markers that users can photograph using their mobile phones. By displaying the marker snapshots on the screen of their phone and moving them over a public display, an external camera can track the position of the device and identify the marker being displayed. We introduce a prototype making use of this idea, and highlight key benefits of our approach such as no need to install custom software on the phone, the elimination of network configuration and exposure of system functionality.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces—*input devices and strategies, interaction styles*

General Terms

Design, Human Factors

Keywords

situated interaction, mobile phones, marker-based tracking, public displays, large displays, visual communication

1. INTRODUCTION

In recent years, the technical infrastructure surrounding us and the number of devices/services at our disposal have been growing at a steady pace. This trend towards ubiquitous computing raises a number of research questions, which have not yet been fully addressed. For example:

- how can we discover what services are available to us in a particular ubiquitous environment?
- how can we interact with a ubiquitous infrastructure?
- how can we make this interaction intuitive and accessible to untrained users?

In this paper, we discuss issues related to these questions in the context of a specific type of ubiquitous technology, namely public displays, which are rapidly proliferating in areas such as airports, transport hubs, and shopping malls. We introduce a particular interaction technique based on *purely visual* communication using mobile phones that are equipped with a camera. Figure 1 outlines the basic idea underlying our approach. The remainder of the paper is structured as follows. In the following section, we briefly discuss related work, and then describe our approach in detail. We will demonstrate its feasibility in Section 4 using an example application, and will discuss benefits and drawbacks of our approach in more detail in Section 5. A brief summary of the main contributions will conclude the paper.

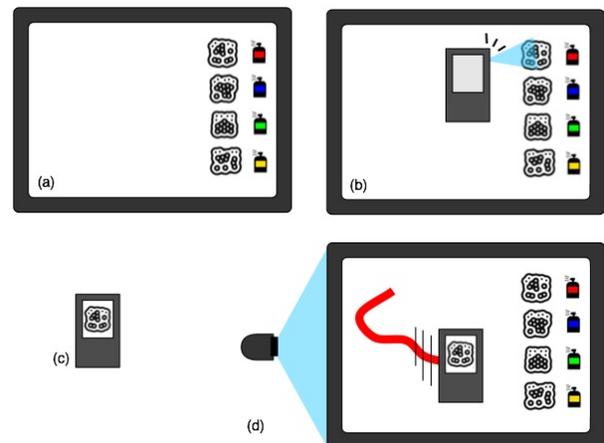


Figure 1: Basic idea: (a) toolbar displayed on public screen alongside corresponding visual markers (b) user takes photograph of marker next to tool to be used (c) user displays photograph of marker on phone screen (d) tool activates when phone is brought in front of public screen and into the field of view of the associated camera.

2. RELATED WORK

Users can interact with a (public) display in different ways. The traditional approach is to use a keyboard and a mouse (trackpad, trackball). More recent approaches include voice or gesture recognition, or a combination of these [7]; touch-enabled surfaces are another option [2]. If the public screen

is a tabletop display, custom-made tokens [4] or custom-made devices are another option [3]. Frequently, personal devices such as mobile phones are also used to interact with a public display (see [1] for a survey). One common way to realize this is through a network connection between the phone and the public display, e. g. via Bluetooth, wireless LAN (802.11) or infrared. If interaction is asynchronous, mechanisms such as text messaging (SMS) can be used. Some interaction mechanisms make use of an external or built-in camera. The built-in camera of a mobile device can help to sense the optical flow resulting from moving the device in space while the camera is recording video footage. Oftentimes, visual markers are used either in conjunction with this approach or on their own. Optical markers being shown on a large public display can be recognized using the built-in camera of a mobile phone, and help to select an item or to track the (relative) location of the mobile device with respect to the marker/large display [1]. The system most closely related to the approach discussed in this paper is C-Blink [6]. It enables mobile phone-public display interaction through a visual marker (i. e. a particular sequence of colors), which is shown on a the screen of a mobile device, and an external camera that tracks these markers. There are however several key differences compared to our system, e. g. C-Blink relies on custom software being installed on the mobile device and does not provide its users with a clear idea of the available functionality or the current state of the system. One advantage of C-Blink over our approach is that it only requires a display on the mobile device whereas our approach also depends on a camera to be present.

Without embarking on a fine-grained analysis of the alternative methods to enable interaction with a public display, it is safe to say that they all have a number of drawbacks. These range from requiring a lot of processing power (e. g. speech recognition) and the need for custom-made hardware (e. g. active tokens) to privacy problems (e. g. tracking facial expressions). Some approaches require extensive configuration, for example, to set up a network connection or to download/install a custom piece of software in order to be able to interact. Furthermore, some systems do not support multi-user interaction (e. g. most touchscreens), while others are fully opaque, i. e. users do not know what services are available to them and how to use them (e. g. speech recognition). Our approach addresses several of these problems: it does not require any custom-made hardware; the only requirements for a mobile device are that it features a display and a camera. We do not require a network connection or specific software on the mobile device either. Hence, a standard camera phone can serve as an interaction device straight away. Furthermore, our approach is user independent and inherently supports the simultaneous interaction of multiple parties. Another key advantage is its immediacy: users can easily tell what functions are available and which one they are currently using. Finally, even though we rely on an external camera, it is possible to anonymously interact with the system.

3. VISUAL INTERACTION

The basic concept behind the approach proposed in this paper is to use a *purely visual channel* for communication between a mobile device and a ubiquitous infrastructure. While we will focus on the interaction between a mobile phone and a public screen, the principle can be applied to

other device combinations as well.

One challenge in ubiquitous computing is how to make users aware of available services and how they can be used. In the case of a public display, an obvious solution to this problem is to simply display a list of available functions on the screen. Figure 1a shows an example in the context of a painting application (cf. Section 4), where spray cans for different colors symbolize the different tools available in the application. Note that next to the actual tool, visual markers are displayed (in our proof of concept prototype, we are using the reactTVision toolkit and the corresponding markers at the moment [5]).

Using the visual display of available functions on the screen, users can immediately select the tool they want to use by taking a photograph of the associated marker (Figure 1b). They can then use the built-in photo browsing application of their mobile phone to show this marker on the display of their phone (Figure 1c). If they have taken pictures of several markers they can select the tool they want to use by flicking through the corresponding photographs that are stored on their mobile phone. In the example case shown in Figure 1, the markers and tools being displayed on the public display can serve as a visual aid to identify which function is currently selected on the phone.

Once the display on the mobile phone shows a marker associated with an application-specific function, it will activate the function while it is in the field of view of a camera pointed at the public screen. For example, in the graffiti application we describe in Section 4, the phone will become a spray can (see Figure 1d). To de-activate the tool, a user can either move the phone out of the camera view, change the content of the screen on the phone, cover the screen of the phone or tilt the phone (so that the camera can no longer track the marker). To switch between different functions on the fly, users can either take a photograph of another tool or select a marker corresponding to the desired function from a list of previously photographed markers.

4. EXAMPLE APPLICATION

In order to explore and demonstrate the idea of a purely visual means of interaction between a mobile phone and a public display, we built a simple example application (see Figure 2). It enables users to ‘spray’ paint onto a virtual canvas using their mobile phones. In the figure, the canvas is shown in the lower left hand area of the plasma screen. On the right hand side of the screen, spray cans for different colors are shown alongside visual markers [5]. To select a particular color, users can photograph the marker shown next to it using their mobile phones (e. g. the topmost one to select red paint). In order to spray paint on the canvas, a user displays the marker corresponding to the desired color on the screen of their mobile phone. Once the marker is visible on the phone’s display, moving the phone inside the canvas area will activate the spray can and the user can paint by moving their phone in front of the canvas (see Figure 2). To stop spraying users can either cover the phone screen, remove the marker on the phone display, move the phone out of the canvas area or move it in a way so that the phone is not parallel to the canvas anymore.

The application was built by modifying software created by the reactable project, in particular the reactTVision toolkit [5]. The hardware setup consisted of a 50” plasma screen, a Nokia N95 mobile phone (display size: 40 × 55mm), an

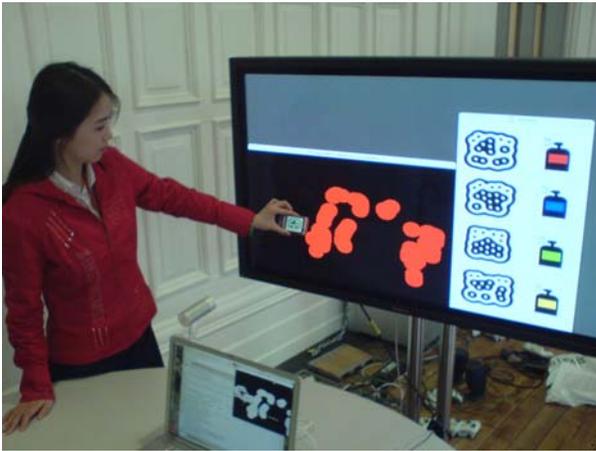


Figure 2: Virtual Graffiti prototype: users can spray paint on a virtual canvas using their mobile phone; photographing a marker next to spray can and then displaying it on the screen selects the color.

Apple iSight webcam (resolution: 640×480 pixel, 30 fps) and a G4 PowerBook (1.67Ghz G4, 1GB RAM). The camera was positioned in front of the display (ca. 40-50cm away) and was able to cover about a third of the display area of the plasma screen (i. e. the canvas area in Figure 2). Both the plasma display and the camera were connected to the PowerBook. No custom software was installed on the phone; we relied solely on the built-in capture and photo browsing applications.

Using this setup (even though it was far from ideal) we were able to paint on the screen when the phone displayed the photographed markers on its screen. While our prototype illustrates that the basic idea of purely visual, marker-based interaction is feasible, there are several aspects, which can be significantly improved in future versions. Most importantly, the reactIVision toolkit is not optimized for this setup (we mainly used it because it is easy to quickly build applications). It was developed to handle a lot of different markers (89 are provided with the software), which is more than would be needed for mobile phone-public display interaction. Reducing the number of different markers would make the recognition more reliable and can also help to make them smaller (thereby affording a larger webcam-to-screen distance). In our experience, the reactIVision toolkit is also not very forgiving with respect to tilting: when phones are not held parallel to the canvas/camera plane, the recognition rate drops sharply (it was built for use with tabletop applications, where it is safe to assume that objects are aligned with the tabletop due to gravity).

In addition, if markers were iconic representations of the function they are associated with, interaction would be more intuitive as users would not have to take photographs of abstract markers but of the actual tool itself (see Figure 3 for an example design). In order to cover a larger area and to improve recognition, a higher-resolution camera (or multiple cameras) would be beneficial.

5. DISCUSSION

While the prototype application presented in the previ-

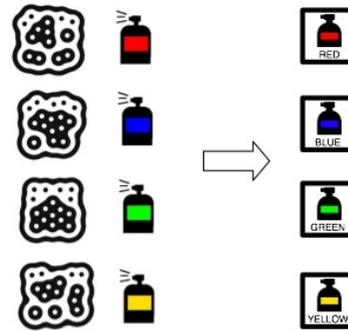


Figure 3: Toolbar exposing available functions: (left) photographing a marker next to a spray can and then displaying it on the phone's screen activates the corresponding color, (right) design for merging markers and tools.

ous section is fairly simple and could be improved in many ways, it nevertheless illustrate the potential of the basic concept. There are several key benefits that result from using a purely visual communication mechanism based on markers to enable mobile phone-public screen interaction:

- **configuration-free operation**

As the system does not require any network connection between the public display and the mobile device, it is not necessary to configure any network settings (unlike Bluetooth or WLAN based approaches). In addition, as there is no need for any special hardware (such as accelerometer or ultrasound sensors often required by custom-made interaction devices) besides the built-in camera, no configuration of such add-ons is required.

- **no custom client software**

The proposed approach can be implemented using solely the software already available on a camera-equipped device, i. e. the capture and photo-browsing application. Users thus do not need to familiarize themselves with any new software (as they frequently have to using, for example, Bluetooth based systems).

- **inherent support for multi-party interaction**

The camera(s) can easily track multiple markers at the same time - thus enabling multiple people to interact simultaneously (in contrast to most touchscreen solutions). The reactIVision toolkit we used to build the prototypical application was designed to track multiple markers simultaneously. However, camera placement with respect to the screen, mobile phones and users needs further investigation, i. e. regarding occlusion issues resulting from more than one user.

- **transparency of interaction**

Through the markers/icons being displayed (see Figure 3, right) it is immediately apparent to the users what functions are available to them (which is frequently not the case with voice-based interaction). We would argue that there is a fairly straightforward correspondence between 'picking up a tool' and taking a photograph of the icon representing the tool (which

can be a problem with gesture-based systems). The same can be said for activating a tool by displaying it on the screen of the mobile device.¹ A beneficial side effect of this is the fact that users can always tell which tool their mobile device currently incorporates by looking at its screen. Moreover, the large display could highlight the currently selected tool.

- **anonymous interaction**

Since there is no network connection between the mobile device and the public screen, interaction is by default anonymous, i. e. the system cannot determine *who* is interacting – it can only tell which tools are active at any given point in time. Many alternative approaches such as those using text messages or Bluetooth disclose at least the identity of the device to the display software.

- **ownership/control/empowerment**

By relying on unmodified mobile phones, our approach enables users to employ their personal devices in a way that does not require them to give up control: they do not need to install any custom software and they do not have to connect to a third party server. In addition, they can transfer their knowledge about their personal devices (e.g. how to take photographs with them) to control an application running on the public screen. Other systems oftentimes require users to learn new mechanisms such as a set of gestures or voice commands.

There are also a few open questions that need further investigation. In addition to the technical shortcomings discussed in Section 4, there is a need to research the particular properties of tracking visual markers being displayed on the screen of a mobile phone (e.g. the impact of reflections, robustness against tilting, optimization for being photographed). The same applies for evaluating the approach with a larger number of users (e.g. to determine a suitable set of markers, to identify appropriate metaphors, to provide interaction-related feedback). Furthermore, different hardware setups need to be explored, in particular different display technologies (e.g. back/front projection) and their properties as well as camera configurations (e.g. multiple cameras, different camera positions).

6. SUMMARY

In this paper, we have introduced a purely visual mechanism to enable interaction between mobile devices and ubiquitous infrastructure in general and public displays in particular. It is based on visual markers that users photograph using their mobile devices and then display on the screen of these devices to activate a particular tool. We briefly described a prototypical drawing application using this approach, and outlined key benefits of this form of interaction. The advantages include ease of use, configuration-free operation, multi-party interaction as well as the preservation of anonymity. Based on this initial research, we attribute considerable future potential to the basic idea of purely visual communication. We will hence further investigate both technical aspects (e.g. how to optimize the markers and the

¹We have not yet conducted proper user studies to test these conjectures but intend to do so in the near future.

display-camera setup) and issues related to interaction (e.g. suitable metaphors, user studies with a set of applications).

7. ACKNOWLEDGMENTS

We would like to acknowledge a hardware grant by Nokia Research Labs, Finland that provided the N95 handsets used. We would also like to thank the organizers and participants of the FTIR workshop in Münster, Germany, for inspiring discussions. Furthermore, we want to thank the authors of the reacTIVision toolkit for making their software publicly available.

8. REFERENCES

- [1] R. Ballagas, J. Borchers, M. Rohs, and J. G. Sheridan. The smart phone: A ubiquitous input device. *IEEE Pervasive Computing*, 05(1):70–77, 2006.
- [2] J. Y. Han. Low-cost multi-touch sensing through frustrated total internal reflection. In P. Baudisch, M. Czerwinski, and D. Olsen, editors, *Proceedings of UIST 2005*, pages 115–118, New York, 2005. ACM Press.
- [3] M. Hazas, C. Kray, H. Gellersen, H. Agbota, G. Kortuem, and A. Krohn. A relative positioning system for spatial awareness of co-located mobile devices and users. In K. G. Shin, D. Kotz, and B. D. Noble, editors, *Proceedings of the third international conference on mobile systems, applications, and services (MobiSys 2005)*, pages 177–190, New York, 2005. ACM Press.
- [4] S. Jordà, M. Kaltenbrunner, G. Geiger, and R. Bencina. The reactable*. In *Proceedings of the International Computer Music Conference (ICMC 2005)*, Barcelona, Spain, 2005.
- [5] M. Kaltenbrunner and R. Bencina. reactivation: a computer-vision framework for table-based tangible interaction. In *TEI '07: Proceedings of the 1st international conference on Tangible and embedded interaction*, pages 69–74, New York, NY, USA, 2007. ACM Press.
- [6] K. Miyaoku, S. Higashino, and Y. Tonomura. C-blink: a hue-difference-based light signal marker for large screen interaction via any mobile terminal. In *UIST '04: Proceedings of the 17th annual ACM symposium on User interface software and technology*, pages 147–156, New York, NY, USA, 2004. ACM Press.
- [7] W. Wahlster. Smartkom: Fusion and fission of speech, gestures, and facial expressions. In *Proceedings of the 1st International Workshop on Man-Machine Symbiotic Systems*, pages 213–225, Kyoto, Japan, 2002.

Mediating Multi-Party Negotiation Through Marker-Based Tracking of Mobile Phones

Michael Rohs
Deutsche Telekom Laboratories
TU Berlin, Germany
michael.rohs@telekom.de

Christian Kray
Informatics Research Institute
Newcastle University, UK
c.kray@ncl.ac.uk

ABSTRACT

Negotiating a date and a time for a meeting involving a number of people can be a difficult and time-consuming process – even when all participants are collocated and supported by technology. Oftentimes, it involves an auction-like procedure, where suggestions and conflicts are announced to the group and then checked individually until a feasible time can be found. We propose to use spatial proximity regions around handheld devices to significantly reduce the effort of exploring proposed meeting times in the context of a party of collocated people. In order to determine the location of devices on a table, we have developed a new tracking mechanism that relies on dynamic visual markers shown on the screen of the devices used. A preliminary evaluation of the underlying idea and the tracking mechanisms highlights advantages and drawbacks of our approach.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces—*input devices and strategies, interaction styles*;
H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces—*computer-supported cooperative work, synchronous interaction*

General Terms

Design, Human Factors

Keywords

mobile phones, marker-based tracking, meeting support, situated interaction, tangible interaction, spatially-aware interfaces, spatial relations

1. INTRODUCTION

Face-to-face meetings with business partners or with colleagues within a company constitute an important part of everyday activities in the professional lives of many people. Consequently, people frequently need to agree on a date and a time when they all will be available to meet. This task can become quite complex if there are a number of people involved or if the participants are very busy, and it may require several rounds of negotiation before a date can be agreed on. A range of devices is available to support users when engaging in this process, e.g. laptop computers, PDAs, and mobile phones. Automatic solutions to the meeting negotiation problem have only been partially successful so far [5], since users often enter low-priority events or block times

in their calendars that can be used for meetings if needed. Moreover, many users prefer to stay in control of their time planning rather than delegating it to an automatic tool.

In the context of this paper, we assume that all participants maintain a personal calendar using a smartphone or a PDA, and that the negotiation takes place in a location with one or more cameras for tracking phones on a table (e.g. a reasonably equipped meeting room). The task that we aim to support is the negotiation part of finding a date and a time for a meeting, in particular suggesting a date and time to the group, jumping to a suggested date and time in one's own calendar, and accepting/rejecting a suggested date and time. To achieve this we use simple spatial gestures and proximity regions around handheld devices. The spatial relationships of the involved devices reflect the negotiation state for a proposed date. The location of a device on the table is determined via a new tracking mechanism that relies on dynamic visual markers shown on the device screen.

The benefits of fine-grained spatial interaction for handheld devices have been recognized some time ago [2], but the idea has been difficult to implement. *TRIP* [1] is a vision-based location system using circular markers. The markers are tracked by cameras installed in the environment. It would be possible to show TRIP tags on phone displays, but their circular form factor would lead to a lot of wasted display space and prevent simultaneous display of application data. *Relate* [3, 4] is a system for the detection of spatial relationships between collocated mobile devices. It does not require an infrastructure, but determines relative position and orientation via ultrasound sensing implemented on USB dongles. This approach requires additional hardware on each device, whereas the marker tracking approach presented here is a software-only solution on the users' side. Our approach requires an external infrastructure, but has a higher accuracy and update rate. *Doodle* (www.doodle.ch) is an online tool for meeting negotiation. The initiator proposes a number of dates and sends a unique URL generated by the system to all participants, who can then individually accept or decline the proposed dates. This corresponds to one round of negotiation in our system, albeit with multiple proposed dates.

2. APPROACH

In this section we describe our approach to achieve the goals outlined above. We first present and motivate the basic idea before giving a detailed description of how we realized it technically. We also conducted a preliminary evaluation of both aspects, which we discuss in section 3.

2.1 Basic Concept

The basic idea underlying our approach is to use spatial proximity regions around mobile phones to trigger particular actions in response to other devices entering them, leaving them, or staying within them for a certain amount of time. For the scenario at hand, we define three regions around a mobile phone, and assign them to synchronization-related actions in the context of the meeting negotiation process. Figure 1 depicts a diagram of the different regions around a mobile device.

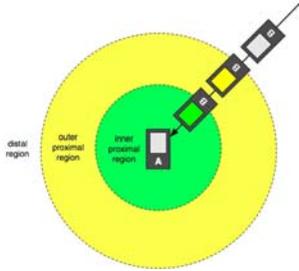


Figure 1: Proximity regions around a stationary mobile phone; see text for explanation.

We call the region furthest from a device its *distal region (DR)*. As long as devices are located in this region relative to a device, they are considered to be too far away for interaction and do operate independently. In this state, it is hence possible for the respective owners of these devices to individually explore their personal calendar.

The interaction is initiated once a device B enters the second region around a device A, which we call its *outer proximal region (OPR)*. In this case, we assume that the owner of the mobile phone A has selected a particular date and time and is proposing it to the group as a potential meeting time. The other members of the group can then check whether they are available at that time by moving their own devices close to the device of A. Once a device B enters the OPR of device A, the calendar application on device B automatically jumps to the date and time that is selected on device A. The owner of device B can immediately see whether the meeting time suggested by A is still available in their own calendar. By moving device B out of the OPR of A and into its DR, the calendar on device B reverts back to the state it was in before entering the OPR of A.

In order to confirm a suggested meeting time, a device needs to enter into and remain within the region nearest to the device proposing the date and time; we call this region its *inner proximal region (IPR)*. As shown in Figure 1, once device B enters the IPR of device A, it confirms the suggested meeting time. User B can abort the confirmation by moving device B out of the IPR of A. If there are more than two parties involved, *all* devices need to be in the IPR of A in order to confirm the suggested date and time. As long as there are still participating devices outside the IPR of A, those inside the IPR will merely indicate that they are ready to commit to the date and time but will not accept it. Figure 2 illustrates this behavior: in case (a), device C is still outside the IPR of A. Therefore, the suggested date is not yet accepted. In case (b), both C and B are inside the IPR of A; hence, the suggested meeting date is accepted by all participants.

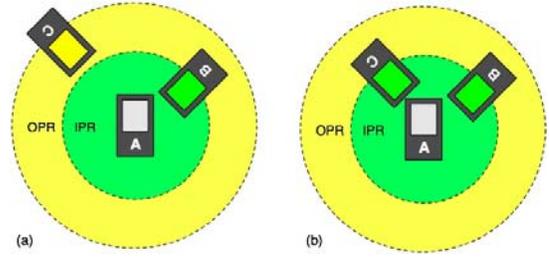


Figure 2: Multi-party negotiations: (a) no commitment as not all devices are inside IPR of A, (b) commitment performed as all devices are inside IPR.

The owner of device A can abort the interaction at any stage by moving device A in any direction. (Covering the visual marker on the display has the same effect as it effectively removes the device from the system.) This also invalidates the suggested time/date and all other devices within the proximal regions of device A revert back to the state they were in prior to entering into different stages of commitment. Table 1 summarizes the different actions and the corresponding negotiation stages.

Table 1: Spatial actions involving devices A and B and the negotiation stages associated with them.

Spatial action	Negotiation stage
B in DR of A	no interaction, i. e. independent operation
B in OPR of A	B displays date/time selected on A (exploration)
B in IPR of A	B accepts date/time selected on A (commitment) – also see text
moving B into DR of A	interrupts interaction
moving A while device are in its OPR/IPR	interrupts interaction

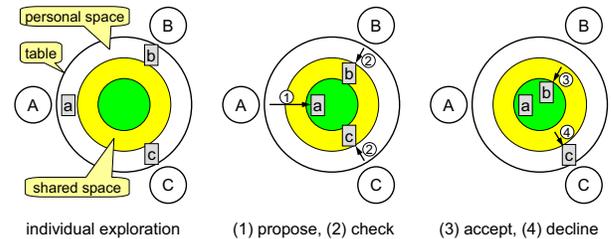


Figure 3: Users A, B, C at a table: (left) individual exploration of personal calendars; (middle) A proposes a date, B and C check the date; (right) B accepts and C declines the proposed date.

An alternative approach to the one described above is to define regions with respect to the physical position of the persons involved. This helps to increase scalability to larger tables and groups, since users do not have to move their devices over a long distance towards to the proposing device.

As shown in Figure 3 the area close to a user is defined as a personal space, in which one’s own calendar can be explored. By moving a phone into the central region, a user can propose a meeting time. Other participants can explore the suggested time by moving their devices into the shared space, accept it by moving them into the central area, or reject it by moving their device back into the personal space.

2.2 Technical Realization

The position and orientation of the phones is tracked by a dynamic visual marker that is displayed on the phone screen (see Figure 4). The marker is designed in such a way that it occupies only a small amount of screen space. In the prototype implementation on a Nokia N95 with a screen size of 240×320 pixels the marker just occupies 15% of the screen area. The marker is placed at the top of the screen in order to make it unlikely that users inadvertently cover the marker with their hands. The phone marker stores 42 bits of data that are protected by a (48,42,3) linear code with a Hamming distance of 3. For each detected marker the recognition algorithm provides the encoded value, center position, rotation, and distance.

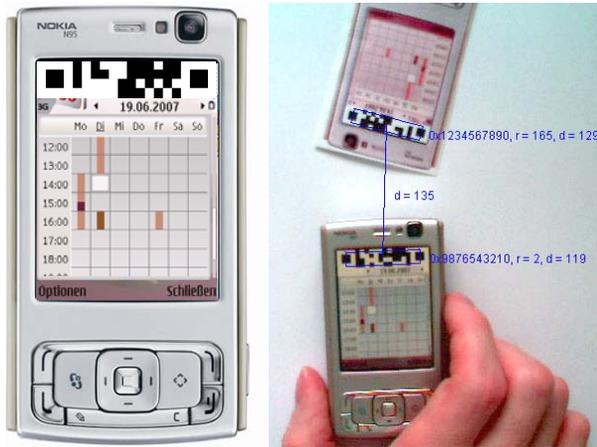


Figure 4: Visual marker at the top of a phone calendar application (left). Tracking of two phones (right). Marker values and distance overlaid.

Figure 5 shows the layout of a marker. It consists of two square corner stones surrounded by whitespace and a central data area of 12×4 elements. The total horizontal width of the marker is 20 elements. The recognition algorithm proceeds as follows: The grayscale image is thresholded to a black-and-white image. To find the corner stones connected regions are computed. Regions with an axis ratio greater than 0.7 (minor to major axis length) and a pixel count within a certain range are classified as corner stone candidates. Then, matching pairs of corner stones are identified. Ideally, matching corner stones have the same size and a distance of 16 elements. Since markers can appear tilted in the camera image, we allow a ±25% deviation from the ideal values of size and distance. In order to sample the data points of a potentially tilted marker, a homography is computed based on four points near the corner stones (indicated by diamond shapes in Figure 5). These points are found by moving ±1.5 elements in a direction perpendicular

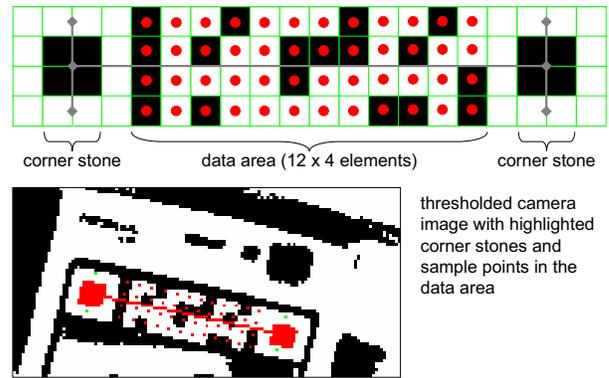


Figure 5: Elements of the visual marker for display on phone screens (top) and thresholded camera image with highlighted structures (bottom).

lar to the connection vector of the two corner stones. The sampled 48 points in the data area then undergo a parity check. The resulting 42 data bits are reported together with the position of the marker (in the coordinate system of the external camera), its rotation, and its distance (inverse of size multiplied by a constant).

Markers are tracked by a downward-facing camera mounted above the table. We used a typical webcam (Logitech QuickCam Fusion) with a wide angle lens (diagonal angle of view about 70°). The wide viewing angle means that the camera cannot be ceiling-mounted but at distances between 30-50 cm only. For ceiling-mounted operation a telephoto lens would be needed. The tracking software is implemented in Java. Initially we connected the video stream via the Java Media Framework (JMF). While being platform independent, the drawbacks of this solution were the long delay (about 500 ms) and the limited resolution (maximum 640×480 pixels). We therefore switched to Windows DirectShow and Java Native Interface (JNI). The delay is shorter and higher resolutions are possible (up to 1280×960 pixels). Marker sightings are reported via Bluetooth to the phones connected to the PC. In the prototype implementation, each phone computes its distance to the closest marker, determines what proximity region it is in and updates its display accordingly.

3. PRELIMINARY EVALUATION

In order to evaluate our approach, we performed a qualitative analysis of the effort involved in negotiating a meeting comparing our approach to a base case. In addition, we compiled an initial characterization of the marker-based visual tracking mechanism. At this stage, both are preliminary but nevertheless provide initial support for our approach. We are also planning a usability study at a later stage.

3.1 Qualitative Analysis of Efficiency

As described in section 1, our aim is to improve the negotiation of a meeting date among multiple people. We assume that they use a mobile device (such as a mobile phone) to maintain their personal calendar. We used the standard calendar application of a high-end mobile phone (Nokia N95)

as a base case.¹ Figure 6 depicts a simplified state diagram of the calendar application. In our analysis, we disregarded any shortcuts that were not accessible/visible through the user interface displayed on the screen. In the figure, arrows correspond to transitions between states, and the numbers correspond to the number of key presses that are needed to get from one state to the other. For example, in order to bring up the form for creating a new meeting on the day selected in month view, four consecutive key presses are needed. There are two steps in particular, for which we want to reduce the effort: exploring a date suggested by another person in one’s own calendar, as well as confirming a suggested meeting date and entering it into one’s own calendar.

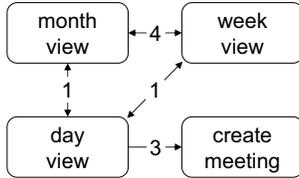


Figure 6: Simplified state diagram of the calendar application on a mobile phone (Nokia N95); arrows are labeled with the number of key presses needed to get from one state to the other.

In order to estimate the time needed to complete these tasks in the base case, we first need to determine the number of key presses a user has to perform during this process. This number depends on several factors:

- the view (month, week, day) that is currently selected
- the temporal distance between the currently selected day and the suggested date
- the number of meetings that are already scheduled for the suggested date.

Both month and week view are organized as 2D grids, which the user can navigate using the five-way navigation button. The day view is a linear list, which also uses the navigation button. We assume that only the day view enables the user to fully assess whether a time slot is available for a meeting or not. (The week view offers only a coarse view of time allocation but does not provide details about scheduled meetings such as their exact starting and end times.) Based on these considerations we can estimate the number of key presses needed to navigate to a suggested date (see Table 2).

It should be noted that Table 1 only represents an estimate, and that there are additional key presses needed if the day suggested already has a number of meetings scheduled. The minimum number of key presses can only be realized if $t_c = t_s$; the further apart the dates are, the more key presses are needed. Assuming optimal usage, there is an upper bound: by entering the pop-up menu and selecting the ‘go to date’ option, users can enter a target date numerically. The cost incurred by this operation is at least 9 key presses (11 if a different month is suggested and 15 if the suggested date is in another year).

¹We briefly explored the calendar tool that comes pre-installed with a mid-range phone (Sony Ericsson K750i) and found it to be very similar to the one on the N95.

Table 2: Key presses (KP) needed to navigate from current date t_c to a suggested date t_s .

view	key presses needed	min
month	1 KP per every week t_c and t_s are apart 1 KP per every day in the week t_c and t_s are apart 1 KP to enter day view	1 KP
week	7 KP per every full week t_c and t_s are apart 1 KP per every day in the week t_c and t_s are apart 1 KP to enter day view	1 KP
day	1 KP per every day t_c and t_s are apart	0 KP

In order to illustrate the key presses needed to navigate to a suggested date, let us consider the following example. A user is in month view and has selected Friday, 22 June. Another person suggests Thursday, 28 June as a meeting date. In this case, three key presses are required. In week view seven key presses would be required, and in day view six. If the suggested date would have been 12 July instead, five key presses would be required in month view, 21 in week view and 20 in day view. In each view, additional key presses would be required in case the suggested day already has more meetings scheduled than can shown at once.

This clearly demonstrates that under all but optimal conditions (i. e. the suggested date is the one already selected) it is reasonable to assume that a user will have to perform between 3 and 9 key presses, oftentimes more. In addition, the original state will be lost. Returning to the date that was originally selected incurs the same number of key presses again. Consequently, the costs will be doubled for every suggested date that is explored but then rejected.

If a date is accepted, further key presses are required. Depending on the view the user is in at that time, three or four key presses are needed to enter the form to create a meeting. In this form, a user has to fill in a number of fields (subject, location, start time, end time, etc.). Minimally, a start and end time have to be provided, which results in at least four key presses (if the automatically suggested times exactly match the proposed meeting times), respectively 14 individual key presses (if start and end time have to be entered manually).

While it is difficult to compare key presses to gestures (i. e. moving a device on a table), Table 3 illustrates that the proposed approach is at least less complex for the user: a single action is required to navigate to a suggested date, to reject the suggestion and return to the previously selected date, and to accept a suggested date. In the base case scenario, a user will also have to navigate through a number of forms or menus, which adds to the complexity, whereas this is not the case in our approach. Furthermore, if the user wants to include additional information about the meeting (e. g. its subject and location), a significant number of key presses will be required. Using our approach, if the proposer included this information in the suggestion, this would not incur any additional costs for other participants.

Table 3: Complexity of interaction to perform a single step of negotiation.

task	base case	our approach
exploration	ca. 3 to 9 key presses	1 gesture
rejection	ca. 3 to 9 key presses	1 gesture
acceptance	ca. 4 to 14 key presses	1 gesture

3.2 Characterization of Tracking Mechanism

Our experiments show that the minimum size of a marker element is about 3×3 pixels in the camera image. This means that the minimum required marker width is $20 \times 3 = 60$ pixels. Since the display width of our prototype device is 40 mm, the table surface must be sampled at a minimum resolution of $60/40 = 1.5$ pixels/mm. At a camera resolution of 1280×960 pixels a table area of 85×64 cm can thus be covered by a single camera. For robustness a slightly smaller physical area may be desirable. On a notebook with a 1.73 GHz Intel Pentium M and 1 GB of main memory the average processing time of the algorithm for a 1280×960 image is 115 ms, 37% of which is due to thresholding. However, JNI and other components introduce additional delays.

Since the marker's corner stones are symmetrical their orientation has to be checked by other means. One possibility would be to reserve two data elements as orientation indicators. Another way, which is currently implemented, is to use the error check as an implicit orientation indication. If the error check fails in one direction, the sampled bits are reversed and checked again. Given that the error protecting code is not symmetric only one direction will succeed.

4. DISCUSSION

After presenting an initial evaluation of the proposed approach, we want to discuss a number of questions regarding the performance, feasibility and other characteristics of using visual markers in this way.

Does the approach improve performance? As highlighted in the previous section and Table 3 further experiments are needed to more precisely characterize the performance of our approach. It is however fairly clear that using our approach significantly simplifies the whole process by reducing the complexity of the interaction needed. In particular, the cost of exploring suggested dates is reduced.

Is the approach suitable for meeting negotiation? Assuming that the date and time of a meeting is oftentimes agreed at the end of another meeting, we would argue that it is reasonably realistic to assume a setting where all participants are collocated and are using personal devices to access their individual diaries. A typical meeting room usually provides a technical infrastructure, which might well include a camera to record a meeting (e. g. the slides presented during the meeting) or equipment for video-conferencing. Even if this not the case, the cost of adding a simple webcam are very low. Based on the argument, outfitting more casual places (such as a café in a business district of a big city) would not be prohibitively expensive.

How to deal with privacy issues? Obviously, camera-based tracking can pose a threat to the privacy of participants, i. e. the camera can record the whole screen content, potentially exposing details from individual diaries. There

are three different responses to this:

1. We assume a benevolent environment (e. g. a meeting room owned by a trusted party).
2. We improve the overall security of the system. Examples for this approach include only showing markers on the screen and limiting the infrastructure to tracking the location of the markers.
3. We eliminate the need for an infrastructure. An idea we plan to investigate would be to use the camera on the mobile phone of a participant.

5. SUMMARY

In this paper, we presented a novel approach to negotiate meeting times among a group of collocated participants. In order to realize it, we used proximity regions and a tracking mechanism based on a new type of visual markers that are being displayed on the screen of handheld devices. We provided a preliminary evaluation, both with respect to its interaction properties and the characteristics of the tracking scheme. Based on these results, we discussed the benefits in terms of its performance and feasibility for meeting negotiation. We also reviewed privacy aspects of our approach.

Due to the overall promising results, we intend to further develop the approach and to validate it in a usability study. In addition to the actual device interaction, it would be very interesting to study the embedding of the approach in the social negotiation protocol and to investigate attention shifts between device and face-to-face communication. A potential way of making the system independent of any external infrastructure would be to use the camera of a participant's device instead of one provided by the environment. Other aspects we want to investigate are the use of dynamic markers as a visual communication channel and the use of the system for other applications involving proximity regions.

6. ACKNOWLEDGMENTS

We would like to acknowledge a hardware grant by Nokia Research Labs, Finland that provided the N95 handsets used. We would also like to thank the organizers and participants of the FTIR workshop in Münster, Germany, for inspiring discussions.

7. REFERENCES

- [1] D. L. de Ipiña, P. R. S. Mendonça, and A. Hopper. TRIP: A low-cost vision-based location system for ubiquitous computing. *Personal Ubiquitous Comput.*, 6(3):206–219, 2002.
- [2] G. W. Fitzmaurice. Situated information spaces and spatially aware palmtop computers. *Commun. ACM*, 36(7):39–49, 1993.
- [3] M. Hazas, C. Kray, H. Gellersen, H. Agbota, G. Kortuem, and A. Krohn. A relative positioning system for spatial awareness of co-located mobile devices and users. In *Proceedings of MobiSys 2005*, pages 177–190, 2005.
- [4] G. Kortuem, C. Kray, and H. Gellersen. Sensing and visualizing spatial relations of mobile devices. In *Proceedings of UIST 2005*, pages 93–102, 2005.
- [5] L. Palen. Social, individual and technological issues for groupware calendar systems. In *Proceedings CHI'99*, pages 17–24, 1999.