ISWC 2004

8th IEEE International Symposium on Wearable Computers

October 31–November 3, 2004

Proceedings of the Student Colloquium

ISWC 2004
8th IEEE International Symposium on Wearable Computers
October 31–November 3, 2004
Proceedings of the Student Colloquium
ISBN 3-902376-04-X

Wearable Computing Lab.
ETH Zürich

xybernaut

UMIT
private universität für gesundheitswissenschaften, medizinische informatik und technik
university for health sciences, medical informatics and technology

The papers of this publication comprise the proceedings of the student colloquium held in conjunction with the 8th IEEE International Symposium on Wearable Computers (ISWC) in Arlington, USA, 2004. All papers are included without change.

ISBN 3-902376-04-X

Additional copies may be ordered electronically from:

University for Health Sciences,
Medical Informatics and Technology - UMIT
Institute for Computer Systems and Networks
Eduard Wallnoefer-Zentrum 1
A-6060 Hall i. Tirol, Austria

Swiss Federal Institute of Technology Zurich (ETH Zurich)
Electronics Laboratory
Gloriastrasse 35
8092 Zurich, Switzerland

Cover designed by Stephanie Fahrni

Program Committee

Chris Baber, University of Birmingham
Michael Beigl, TECO, University of Karlsruhe, Germany
Holger Junker, ETH Zurich, Switzerland (Co-Chair)
Paul Lukowicz, UMIT Innsbruck, Austria / ETH Zurich, Switzerland (Chair)
Walterio Mayol, University of Oxford, UK
Jani Mäntyjärvi, VTT Electronics, Oulu, Finnland
Bernt Schiele, University of Darmstadt, Germany
Albrecht Schmidt, LMU Munich, Geremany
Kristof Van Laerhoven, Lancaster University, UK

Overview of Papers

Interaction Authoring for Adaptive User Interfaces in Ubiquitous Mixed Realities

Irma Lindt, *Fraunhofer Institute for Applied Information Technology*

Usability of Electronic Maintenance Manuals on Wearables and Desktops

David Liu, *School of Information Technology & Electrical Engineering, University of Queensland, LinkOne Product Team, Solutions Development, Mincom Ltd*

Implementing an AR-Search Database

Seng Heng Foo Angelina Novacheva, *Singapore Polytechnic Singapore, Singapore and Stuttgart University of Applied Sciences, Germany*

Context from multiple sensors for activity recognition

Georg Ogris, Institute for Computer Systems and Networks, UMIT, Innsbruck, Austria

Towards an ultra low power user activity recognition system

Niroshan Perera, Julian Randall and Gerhard Tröster, Wearable Computing Laboratory, Department of Electronics, Swiss Federal Institute of Technology, ETH Zürich, Switzerland.

Interactive Mixed Reality Rendering in a Distributed Ray Tracing Framework

Andreas Pomi and Philipp Slusallek, Computer Graphics Group, Saarland University, Germany

Location by solar cells: an experiment plan

J. Randall, O. Amft, G. Tröster, Wearable Computing Lab., ETHZ, Zurich, Switzerland

Real-Time Photorealistic Rendering for Augmented Reality in Unprepared Environments

Stephen DiVerdi, Department of Computer Science, University of California, Santa Barbara, USA

Fluid Beam - A Steerable Projector and Camera Unit

Lübomira Spassova, Department of Computer Science, Saarland University, Germany

Wearable Speaker Separation and Tracking System

U. Anliker, J. Randall and G. Tröster, Swiss Federal Institute of Technology, Wearable Computing Lab, ETH Zurich, Switzerland

Input, Control, and Application Design for a Useable Wearable Computer

Brian Mila, *Virtual Reality Application Center, Iowa State University*

Papers

An On-Line Sensor Data Processing Toolbox for Wearable Computers

David Bannach

Institute for Computer Systems and Networks
UMIT Innsbruck, Austria
david.bannach@umit.at

***Abstract*— This paper presents the design and initial implementation of a toolbox for recognition systems running on wearable computers. The toolbox is intended to facilitate the development of systems for on-line processing of sensor data. The special requirements of wearable devices are taken into account leading to a design that supports distributed computing. An initial implementation proves the feasibility and practical utility of the toolbox.**

I. Introduction

Developing systems for on-line context recognition always implies building a streaming network for the sensor data. The sensor readings have to be continuously acquired from one or more sensors under real-time conditions and then passed into diverse filters for conditioning, pre-processing and feature extraction. The filtered data stream is then fed into classifiers for the contextual recognition. The classifiers itself are often a nested structure or network of well known classification algorithms, e.g. Hidden Markov Model (HMM) or Bayesian classifier. Most of the pre-processing and feature extraction tasks are common in every recognition system and there is little sense in re-implementing them for each system.

Some components are too complex or computational expensive to run on a constrained autonomous sensor or even on a light-weight multiple purpose wearable computer. Also, to keep a low power dissipation for the wearable system, off-loading computation to a remote processing system is considered valuable [1].

The toolbox presented here aims at a general and flexible solution for the streaming network behind on-line contextual recognition systems. The goal is to create an infrastructure that facilitates the configuration of such systems using standard pre-built and customized components with clearly defined interfaces and providing means for remote processing.

A. Outline

Section II lists the specific requirements for an on-line sensor data processing toolbox for wearable computers. Section III describes the design decisions made for the toolbox and section IV shows how the toolbox can be used to configure a streaming network for context recognition. Finally section V contains a summary and conclusion.

B. Related Work

IU Sense [2], developed by students of the International University (IU) in Germany, is an extensible real-time application with graphical data displays in various views and user configurable layouts. It also provides offline data visualization from files and conversion into several different file formats.

The Context Toolkit [3] is an architecture for building context-aware applications. As such it helps building applications that use context information. The toolkit introduced in the present paper aims at the level below, i.e. it helps building systems that provide context information.

II. Requirements

The requirements of a general toolbox for wearable computing as envisioned above can be summarized as follows:

1. The toolbox should (as much as possible) facilitate the development process of systems for contextual recognition. Tasks that are common in such systems, like acquisition of the sensor data, conditioning, filtering, feature extraction, and some well-known classifiers, should be components of the toolbox and should be easy to configure. This way, the developer can focus on the actual problem.
2. The limited computing power of wearable devices mainly coming from limited battery capacity and the power dissipation of fast processors complicates the implementation of experimental, yet computing-intensive algorithms. It might also be desired to use existing mathematical libraries for some experiments. Most such libraries do not run on wearable devices or the effort for porting them exceeds the benefit at that time. Therefore, the toolbox should provide a means to include external, less restricted machines into the processing of the sensor data. This could be a desktop PC connected to the wearable computer via wireless LAN [4], [6].
3. Even if most algorithms were provided by the toolbox, it might still be desirable to use existing tools (e.g. Matlab) for certain tasks. The toolbox should provide a suitable interface allowing external tools to be integrated, independent of their programming language.
4. In collaborating teams where multiple persons are equipped with wearable computers it is desirable to exchange sensor data or higher level information with the team mates. Data from team mates may introduce new information (e.g. if

only one mate has a GPS receiver) or it may improve the accuracy of the information (e.g. all mates have GPS and distance between mates is known).

5. The limited computing power of wearable devices should be utilized in an efficient way. For instance, the XScale processor used in the QBIC [5] has no FPU. Operations on floating-point variables therefore take proportionally many CPU cycles and should be avoided. It is preferable to use processor-specific libraries like Intel provides for the XScale processor.
6. The system requirements of a streaming network highly depend on its configuration. For verifying the flawless operation of a toolbox, monitoring of the streaming network should be possible and data losses due to system overload must be reported.
7. Even if it is helpful to use fast machines in the background for experiments, in many realistic cases of wearable computing such machines are not available (e.g. when mountain hiking). The toolbox should support the seamless migration from fast machines to the wearable device.

III. System Concept

In order to meet the requirements listed above the following decisions on the design of the toolbox have been made:

A. Streaming Network

The toolbox allows to build a *streaming network* for the sensor data. The network consists of *tasks* and *connections*. Tasks are the processing entities that operate on data streams. They provide *in-ports* and *out-ports* for incoming and outgoing data streams respectively. Directed connections between tasks can be established from out-ports to in-ports of matching type, specifying the direction for the data streams.

Figure 1 shows an example configuration of a streaming network. The named boxes represent the tasks with in-ports on the top side and out-ports on the bottom side. In this configuration the data from the sensor is streamed into two different filters (mean and variance), merged to one (extended) stream and then fed into the classifier. The output of the classifier is logged to a file.

B. Task Classes

The different tasks in a streaming network can be categorized as follows:

Readers are tasks that have no in-port but one or more out-ports. They read the sensor data from external sources like sensor devices, files, or TCP sockets. The data streams are provided on the out-ports. Number and type of out-ports is specified by each task.

Filters have one in-port and one out-port. They process the data arriving on the in-port and stream the results to the out-port. Downsampling and oversampling filters are possible. Example filters are: mean, variance, RMS, FFT.

Splitters/Joiners are stream controlling tasks that split or merge the incoming streams respectively. They are useful for reorganizing and synchronizing data streams. The number of in-ports and out-ports depends on the actual task.

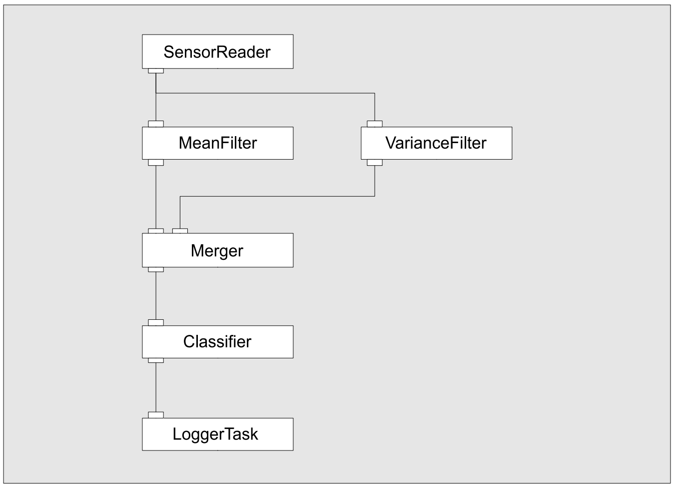

Fig. 1. Example configuration of a streaming network

Classifiers are similar to filter tasks but are mostly based on statistical data structures in order to classify the incoming data stream. Moreover, classifiers have a training mode where a labeled input stream[1] is used to train the classification model. Examples: Naive Bayes, Decision Tree, HMM.

Writers are tasks that have one in-port but no out-ports. They write the incoming data stream to a file, a TCP socket, or to other external entities.

C. Data Streams

The directed connections between out-ports and in-ports of tasks are the internal links of the streaming network. Each link is treated as a separate stream. The actual streamed data is specified by the originating task.

Writer tasks can be used to stream data to separate processes running on the same or remote machines, creating external links. Similarly, reader tasks can receive external data streams from other processes. This way the communication to other tools and remote machines is made possible.

Internal and external links are explicitly distinguished from each other due to different properties in terms of reliability and timing, thus enabling the implementation of internal connections to be optimized as much as possible and preventing errors introduced by false assumptions of connection properties.

IV. Implementation Details

The implementation of the toolbox is completely object oriented and is done in C++. An abstract `StreamTask` class based on POSIX threads is used as superclass for all tasks, therefore each task is a separate thread. Tasks have vectors of `InPorts` and `OutPorts` for receiving and sending `DataPackets` respectively. The `DataPacket` basically contains a timestamp and a data vector. It is instantiated by the reader task for every new data sample acquired from a sensor (or other sources) and then passed through the streaming network from task to task along the internal links[2]. Tasks

[1] Labels are either embedded in the data stream or as a separate stream on an additional in-port.

[2] Actually only a pointer is passed and the object's data stays in place for performance reasons.

processing the data packet do only change the data vector. When multiple receivers are connected to the same out-port of a task the packet is cloned.

The toolbox utilizes the XParam Object Serialization Framework [7] for runtime creation of objects. This enables the toolbox to be configured at runtime.

A. Building a simple streaming network

The task of creating a streaming network using the toolbox is simplified to the execution of two steps: 1) Write a configuration file. 2) Start the toolbox.

The configuration file specifies the `StreamTask` objects to be run in the toolbox, as well as a list of connections defining the links between the tasks (*source, out-port, destination, in-port*). The syntax is similar to C++ but is more relaxed (e.g. strings don't need to be quoted if recognizable as such). The exact syntax is defined in the XParam project. The configuration file for the example network shown in figure 1 is listed below (assuming sensor device '/dev/ttyUB1', window size 50, classifier configuration 'config1', and logfile 'logfile'):

```
tasks=[
    SensorReader( reader, /dev/ttyUB1 ),
    MeanFilter( mean, 50 ),
    VarianceFilter( var, 50 ),
    Merger( merger, [0, 1] ),
    Classifier( clf, config1 ),
    LoggerTask( log, logfile )
]

connections=[
    Connection( reader, 0, mean, 0 ),
    Connection( reader, 0, var, 0 ),
    Connection( mean, 0, merger, 0 )
    Connection( var, 0, merger, 1 )
    Connection( merger, 0, clf, 0 )
    Connection( clf, 0, log, 0 )
]
```

The configuration file is passed to the toolbox when started. The toolbox automatically creates the corresponding `StreamTask` objects (de-serialization), establishes the specified connections, and starts the tasks.

B. Outsourcing of tasks

Assuming the classifier used in the streaming network created above is too complex to run properly on the wearable device it must be outsourced to another machine, i.e. a second toolbox (toolbox2) is started on the additional machine. Figure 2 shows this scenario. In place of the classifier a `TCPWriter` task is inserted in toolbox1 writing the data stream to a TCP socket. In toolbox2 a `TCPReader` task receives the TCP stream from toolbox1 for further processing.

V. Conclusion

This paper introduces a general toolbox for on-line contextual recognition systems. The toolbox facilitates the configuration of recognition systems by providing highly customizable and extendable components for processing data streams and clearly defined interfaces for interconnecting these components building a streaming network for the on-line processing of sensor data. The toolbox allows components to be outsourced to remote hosts during development of new recognition systems and to sequentially migrate tasks back to the wearable computer as deployment advances and algorithms stabilise.

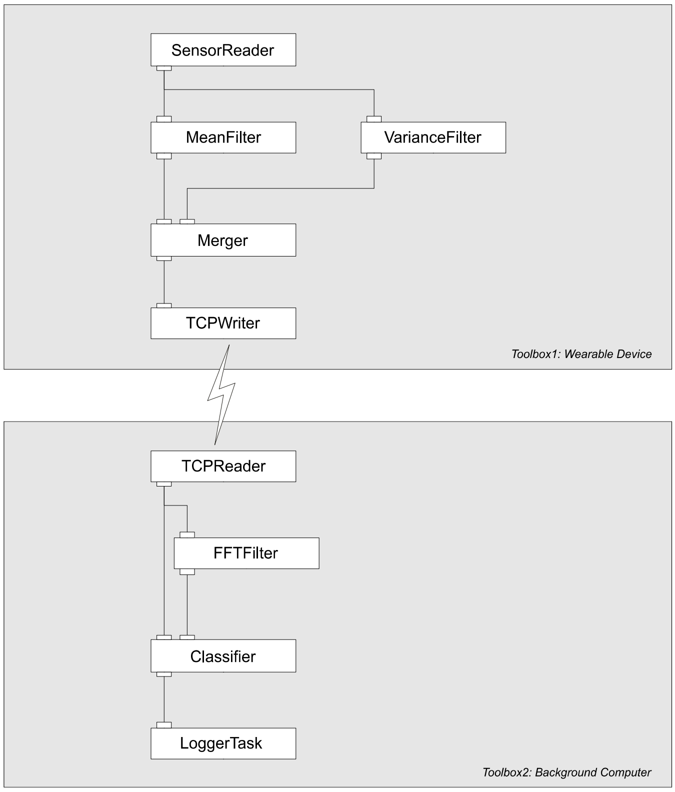

Fig. 2. Example configuration for outsourcing of tasks

The initial implementation of the toolbox shows its feasibility and practical utility. Nevertheless, there is still work to do to make the toolbox as flexible and usable as envisioned. This includes:

- Finding a proper language for specifying the type of data streamed by the tasks.
- Implementing more standard filters and classifiers.
- Creating a service for remote configuration, controlling, and monitoring of the toolbox.
- Creating a graphical user interface for configuring the toolbox.

References

[1] U. Anliker, J. Beutel, M. Dyer, R. Enzler, P. Lukowicz, L. Thiele, and G. Troester. A systematic approach to the design of distributed wearable systems. *IEEE Transactions on Computers*, 53(8):1017–1033, Aug. 2004.

[2] A. Caracas, E. Heinz, P. Robbel, A. Singh, F. Walter, and P. Lukowicz. Real-time sensor processing with graphical data display in java. In *IEEE Symposium on Signal Processing and Information Technology*. IEEE Press, December 2003.

[3] A. Dey, D. Salber, and G. Abowd. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-Computer Interaction (HCI) Journal*, 16(2-4):97–166, 2001.

[4] M. Kourogi, T. Kurata, and K. Sakaue. A panorama-based method of personal positioning and orientation and its real-time applications for wearable computers. In *Fifth International Symposium on Wearable Computers (ISWC)*, pages 107–114, October 2001.

[5] QBIC - Belt Integrated Computer, Wearable Computing Lab, ETH Zürich. http://www.wearable.ethz.ch/qbic.0.html.

[6] T. Truman, T. Pering, R. Doering, and R. Brodersen. The infopad multimedia terminal: A portable device for wireless information access. *IEEE Transactions on Computers*, 47(10):1073–1087, October 1998.

[7] XParam Object Serialization Framework. http://xparam.sourceforge.net/.

Wearable Computer for Building Surveying

Jan Doppleb, Nikolaus Steinke, Norbert Meyer, Alexander Kulik
Bauhaus University Weimar, Germany (tragbar@buan.de)
attended by:
Frank Petzold, Torsten Thurow *(Bauhaus University Weimar; Germany)*,
Christian Bürgy *(Wearable Consult, Dossenheim, Germany)*

Abstract

More than half of all building investments in Germany are in the renovation sector and this proportion will continue to rise. The aim of this project is to design a practice-relevant hardware and software concept consisting of a wearable, digitally supported equipment and the underlying software environment for the digital architectural surveying of buildings. The work is mainly based in the motivation of an interdisciplinary team to get a suitable building surveying system working, that later will allow the evaluation of both – the hardware and software aspects of such wearable computer systems.
This paper will discuss the concept, the design and the first prototypes of the hardware part. We illustrate the functionality, the ergonomics and the underlying research path.
Keywords: wearable computing, building surveying, mobile IT support, software and hardware architectures

1. Introduction

Since the beginning of the 1990's the focus of planning activities has shifted away from the construction of new buildings to renovation and construction within existing contexts. More than 60% of planning activity concerns existing buildings.[1]
Reliable and informative documentation is an essential pre-requisite for planning in general but especially for the planning task with existing buildings, where existing plans and building documentation are hardly available, very basic or often not up-to-date.

Figure 1: usability problems with current technology

In most cases the term "building survey" is understood to mean a geometric survey of the physical dimensions of a construction translated into architectural plans, sections and elevations. The geometric survey is typically undertaken using geodetic or photogrammetric measuring techniques and equipment that have been adapted for use in building surveying as shown in Figure 1.
The consequence: current building surveying and planning working practice is characterized by:
- a high level of mechanization;
- the need for specialists in geodesy / photogrammetry;
- a lack of simple tools for architects & engineers;
- the reduction to solely geometric data;
- the inclusion of a lot of redundant information.

2. Approach

We envision a fully computer-supported surveying process. This supporting IT system will help to avoid unnecessary data transmission between different analog and digital media, and thus avoid unnecessary media breaks (data transmissions). The envisioned system will fully support the surveying process in every single step.
To minimize the time and money spent for surveying of a building-site, we split the whole working process into different steps. These steps are ordered in a way as to get more detailed information with every new step. At the same time, we ensured that the information of a given step is saved for reuse in later steps, according to the surveying targets in the appropriate surveying phases (see table).
Building survey is not simply a geometric description of a building. In addition to the structured capture of the building geometry, other formal, informal and relational data is also captured and stored within the information container. The geometry structure is linked with formal-descriptive data.
The captured data should serve as a basis for further design and planning activities [2,3,4].
Before we could approach a system with intuitive capture and data entry interfaces, we first had to design and develop the core wearable computer that could be worn unobtrusively while performing sureveying tasks.

3. The Hardware Concept

There are different work tasks, which depend on the surveying-targets. One possible approach is to collect all available information in one step, but it will be more efficient to do only the necessary steps at a time. These different tasks require different tools, which the surveying person will carry on-site. Compareable to the carpenter's toolbelt as an example of an entire „wearable workshop", we want to develop a wearable system, that provides all the functionality to get all necesary specifications of the surveyed building and allows the connotation of collected data with further description in photographs, text and sketches. Similar to the carpenter, who equips himself for the next task on site by fixing all necessary tools at the straps that provide the typical toolbelt, the surveyor does so by connecting his tools to the USB-plug and fixing them to straps and pockets, that all provide our specific wearable system. Voice recognition and sound via microphone and loudspeaker located in the straps on the users shoulders refer to the basic equipment of our wearable system as like Bluetoth and WLAN functionality. If possible devices shouldn't be connected with disturbing cables.
The wearable computer unit always provides the surveyor with previously collected data and a specially designed surveying software that allows to interact with this data container. Every preferred display may serve for visualisation. We propose the use of a Virtual Retinal Display or an interactive pen display[7,8]. The following table shows examples of further connectable devices in reference to specific surveying tasks.

First site visit (inventory, condition)	- Digital Camera - Measuring devices
Sketch based spatial information	- Digital Sketchbook[1] - Distance-meter[2]
Detailed ground plan	- Distance-meter
Exact 3D geometry	- Tacheometer - Digital camera (photogrammetry)
Acceptance of work, Facility management	- Digital camera

[1] in best case with 6DOF-tracked pen and Seethrough-HMD to sketch in the air (figure 3)
[2] in best case with 6DOF-tracked distance-meter - just point for mesuring (figure 2)

Our approach is to keep all building site information together in one digital data container. Information should be linked immediately to its context. Thus, we use one basic computation module that will be powerful enough for CAD and voice-recognition. We design all attached periphery to be flexible, compatible and exchangeable, to reflect the carpenter's tool belt analogy.

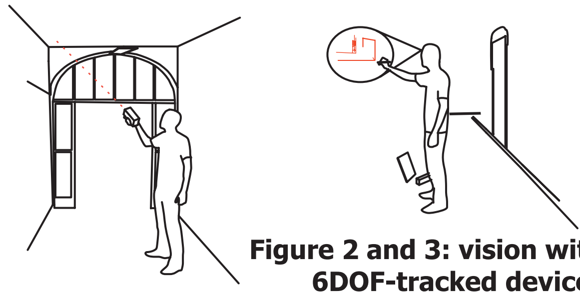

Figure 2 and 3: vision with 6DOF-tracked devices

3.1. Design Evaluation

In the beginning, we thought about clustering the whole computer into little modules, that we could distribute all over the body of the user. This way we wanted to achieve a wearable computer system to be worn very flat on the body, resulting in an unobtrusive device. Realizing this means to design a flexible motherboard where only small electronic devices would remain as solid objects mounted on the surface like little beads on a dancing dress. Flexible circuit boards, that provide more than two conducting layers would probably be a prerequisite and are not yet commercially available. Due to the fact that only rigid mainboards were commercially available, we had to postpone the use of a flexible board to future design iterations. Thus we decided for the ETX-family by Kontron which allows to interchange motherboards with the right scaled computation power for the specific task.
Thus we had to search for areas where rigid parts may be worn close to the human body. Due to the assumed need for relatively vast flat areas and relatively low weight to place there, we concentrated our research on the human torso. To verify the design approach of Gemperle et al.[5] for the surveying application, we plastered our body (Figure 4 and 5) and moved afterwards simulating typical movements of building surveying tasks, to recognize how the surface is breaking into pieces. Thus, we could identify the largest possible size and the ideal location for the computer modules that would not disturb surveying tasks while being attached to the body of the surveyor. Although not mentioned by Gemperle et al. [5] as areas to unobtrusevly wear objects, the plaster experiment showed that there are areas

Figure 4 and 5: plaster experiment

on the human body that qualify for wearing rigid accessories and feel comfortable.
According to the research at CMU, human collarbones should be left out as an area to place wearable parts. The high motion of them explains that thesis. It's generally right, but we recognized, that for lightweight flat volumes like circuit boards mostly are, the bladebones offer a relatively vast area to support them – preconditioned, that they are well cushioned to the users body and provide enough flexibility to move along with the underlying bones. (Figure 5)
Owing to the reachability that our carpenter's toolbelt approach requires, the further investigation went on to find an appropriate place for interface connectors (USB, Video, Sound) The plaster experiment resulted in an area above the users collarbones. (Figure 4) The proximity to shoulder and neck joints features these areas to be ideal for the placement of interface connector plugs. This way a comfortable working range is guaranteed even for peripheries that will not work wirelessly. The cable adjusts to the arm or head orientation (Figure 6 and 7).

Figure 6: cable route model from connector to HMD [8]

We developed different possibilities of wearing those parts according to our results of ergonomic studies and the requirements resulting from the different surveying steps. The idea of wearing a simple textile shirt, which connects the used parts with conductible textile, showed that textile is too flexible to keep the rigid parts in its position. Weight pulls them downwards – a disturbing slinging movement at ones body.
Another idea to wear the solid parts at a more rigid harness-like piece made of semi-flexible plastics, turned out to work well to some degree, but this solution lacks the possibility of changing it between persons with different physiognomies. So we decided for a system based on straps known from backpacks. Those would be adaptable to different body ergonomics and provide more stability than shirt-like textiles. In these studies, the positioning of the connectors turned out to work well, but the bladebone cases, that should contain motherboard, batteries and harddisk resulted to be uncomfortable if the weight and size of a heat sink, necessary for high-clocked processors, is added. So, we decided for a concept that, similar to modern backpacks with ventilation system, will keep the CPU away from the human body and secure permanent and sufficient aeration. The backpack straps still allow to use the place above the collarbones as connector plug for peripheries.

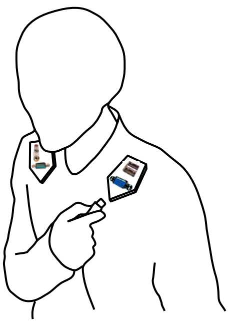

Figure 7: collarbone-placed interfaces

We propose two different styling concepts, that focus either on minimized size (Figure 8) or maximized ventilation and flexibility (Figure 9). The shape of the chassis will go through more design iterations, while we are finalizing the technical architecture inside. Finally, we will evaluate the working unit with a specially designed client / server software by T. Thurow [6].

4. Conclusion and Outlook

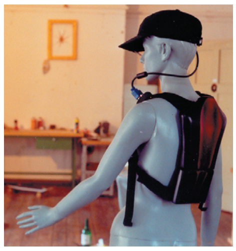

Figure 8: model of backpack-wearable computer

The IT support of building surveying and especially the improvement of the underlying processes is a complex task. In approaching this task with an interdisciplinary research team, we developed the concept of a wearable surveying system that runs a modular software architecture. The cooparation of modular software and hardware architecture enables the system to be designed after a tool belt analogy offering the task-specific tools needed at a certain point in the surveying process.
We will continue this interdisciplinary reasearch investigating different aspects, such as work process analysis, usability as well as hardware and software concepts. Simultaneously, we will evaluate the current system in field-tests at actual surveying sites.

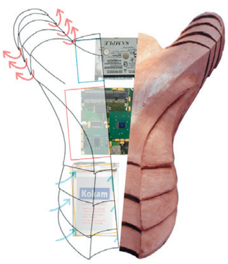

Figure 9: model of wearable chassis made of sheds

References:

[1]BMVBW (2001). Initiative Architektur und Baukultur – Bundesministerum für Verkehr, Bau- und Wohnungswesen http://www.bmvbw.de/architektur-baukultur/ (Stand 25.05.2002)
[2]Donath, C. Tonn (2004). Plausibility in architectural design.– software support for the architect-oriented design of colour schemes for interiors and buildings. ICCBE 2004
[3]Donath, D. (2003), Die Auseinandersetzung mit dem Bauwerk -Notwendigkeiten im Planen und Bauen,IKM 2003, Weimar, published on CD-ROM
[4]Tonn, C. Wolkowicz, T. Thurow, J. Ruth, D. Donath (2004). Plausibility in architectural design.– software support for the formal shaping and architect-oriented design of shell structures. ICCBE 2004
[5]Gemperle et. al, Design for Wearability, Proceedings of the Second ISWC. IEEE Computer Society Press, 1998
[6]Thurow, T. Donath, D. (2004) A vision of an adaptive geometry model for computer-assited building surveying, ICCBE 2004
[7]according to WACOM, http://www.wacom.com/lcdtablets/index.cfm, 2004
[8]according to Microvision, http://www.mvis.com/nomad/index.htm, 2004

Adaptive Vision Based Scene Registration for Outdoor Augmented Reality

Jason Catchpole
Department of Computer Science, University of Waikato
jjc6@cs.waikato.ac.nz

Abstract

Computer vision has proven to be useful in providing accurate registration in Augmented Reality systems. Systems in the outdoors may be required to work on a large range of possible scenes, under varying conditions. Because of the range of scenes, the use of one vision algorithm in all circumstances seems unlikely to produce the best results. This paper looks at a non-real-time system which relaxes the constraints of a real-time implementation, thus allowing for more scope in algorithm choice. The system will combine multiple vision and image processing algorithms and will chose the algorithm to use based on the type of scene and the conditions of the environment. This will enable the system to cope with a larger range of scenes and conditions, thus be more robust.

1. Introduction

Obtaining accurate registration outdoors is a difficult problem because of the range of environments a system needs to work in and because we cannot control the environment as much as with indoor systems, see Azuma in [1]. Several technologies exist to aid in registration [2] with varying degrees of accuracy, but many authors [1, 3] accept that in the near term only a hybrid system, combining different sensors, will provide the robustness necessary to form a complete solution outdoors.

Computer vision has been identified as a technique capable of providing accurate registration. Using computer vision by itself to locate the user anywhere in the world is impossible in most locations because of the ambiguity in differentiating similar scenes, and the amount of time taken to perform the many calculations, at least with any computer available in the near future. For this reason I propose a system in which vision is combined with GPS and inertial sensors which will provide a rough estimate of location and orientation. Inertial and vision sensors were identified by Welch and You et. al. [3, 4] as complementary technologies and have been used in many projects [3, 5].

Other projects have utilized computer vision by tracking fiducial markers [6], integrating region tracking with optical flow methods [5], tracking point features [7], matching to an existing CAD model [8], using Fourier-spectrum based techniques [9] to match to reference frames [10], tracking the horizon silhouette [11], tracking curves and lines [12], and tracking planar structures located in the scene [13]. These systems each work well in specific circumstances, but can be unreliable in scenes outside each of their ideal working conditions. Based on the immense range of possible outdoor scenes, from snow covered mountains and forests, to suburban streets and city centers, through to open fields and sunny beaches, the use of one algorithm for every type of scene seems unlikely to always produce the best results. Environmental changes in the scene itself such as fog, rain and snow, as well as lighting changes from dawn to dusk, and dynamic changes due to cars or people coming and going, objects being moved, occlusions and varying viewpoints, all compound to make a difficult problem worse, and could render an already brittle vision algorithm useless.

I conjecture that a better solution is to choose the algorithm based on the type of environment in which the system currently operates. This requires the system to detect the current type of scene, and determine the best algorithm to use. This choice will also be influenced by environmental conditions like those mentioned above.

2. The System

Orientation data for this project is provided by the Intersense Inertia Cube2 (IC2). This sensor uses a solid-state inertial measurement unit using micro-electro-mechanical systems technology, combined with accelerometers and magnetometers to compensate for gyroscopic drift. The use of a Kalman filter discards the portion of accelerometer measurements [14] due to motion instead of gravity, and reduces the “slosh” effect that was seen in early low cost trackers. Although the magnetometers considerably reduce drift, in the presence of magnetic field distortions the accumulation of drift in the yaw axis still occurs. In addition the accuracy of the

IC2 is not enough to be used solely as the source of orientation data as shown by You et. al. [3].

For basic positional information either standard or differential GPS will be used. GPS can be unreliable in cities, where buildings block GPS signals in many locations. As the focus of the project is primarily on the vision component of the system, the loss of GPS signal will not be considered. It will be assumed that GPS signals will be available, if only momentarily, after a significant positional change of the system.

Position and orientation sensor information only provide approximate registration. Computer vision is then used to refine the registration. As noted earlier no one algorithm is robust enough to operate in all conditions. It is my hypothesis that robustness can be improved by the use of multiple registration algorithms. This introduces the problem of selecting the most appropriate algorithm to use in any given situation. Image classification techniques could be used to make this selection; however, this is in itself a difficult problem. Jain et. al. [15] observe that even people cannot interpret an arbitrary scene "without some specific domain and contextual knowledge". Instead I propose to develop confidence measures for each algorithm by evaluating the features or properties required, and their presence and strength in the current scene. The technique with the highest confidence measure will be selected and used until confidence diminishes to a level which affects registration accuracy. At this point the process is restarted and an algorithm more suited to the new conditions, is selected. Alternatively the environment could be prepared beforehand so that for a given area the scene would be manually classified. This method provides an adequate fallback position should some locations be too difficult to classify through automated methods. This approach would require additional preparation; therefore ideally the vision system will determine the best algorithm to use based on what it is seeing.

If the algorithm best suited to the scene is ambiguous, it may be desirable to use multiple vision algorithms concurrently and to choose the best results. It may be possible to use image segmentation techniques [16] to segment the current image, and apply an algorithm to only a portion of an image.

The project will not concentrate on obtaining a real-time implementation as doing so would place too greater constraints on the choice of algorithms and techniques. It is envisaged that with increases in computational resources, combined with the mostly untapped power of programmable hardware such as graphics cards, which have been used for general purpose algorithms in recent projects [16], the approach will eventually be feasible in real-time.

With increases in the accuracy of GPS, inertial sensors, or other potential technologies, comes the benefit of closer initial estimates of the system's position and orientation. This means smaller search spaces [3], which reduces the computational requirements of the vision system. As the focus is not on a real-time system, dynamic registration errors due to system delays [17] will not be considered, however predictive methods by Azuma and others could be integrated, if a real-time version of the system was implemented.

An initial system has been implemented in C++ providing functionality to stream and capture live video using DirectShow. A modified version of [18] described in [9] is used to match live frames to pre-captured reference frames, the system then uses the transformation between the two frames to augment the current frame using simple objects rendered using OpenGL. The OpenCV library is used in the implementation, it provides many pre-built functions for image manipulation, image processing, feature detection and motion analysis.

3. Evaluation

In order to evaluate the system two approaches will be taken. The first will involve running each algorithm offline, on a range of pre-captured video and sensor information covering different types of scenes. Although the tests will be run offline, it is important to note that this will not be a simulation, the video footage will be combined with actual inertial and GPS measurements. These measurements could be hard coded, as the evaluation is offline. There presence however, provides an initial guess and reduces the number of false positives in the vision system. By running each algorithm on different scenes the conditions in which each algorithm works best can be determined. Evaluating how the system chooses which algorithm to use based on the current scene will involve running it on the pre-captured footage and analyzing which algorithm is chosen. This portion of the system will determine how the overall system will work as it dictates which algorithm is selected. The registration error will be determined based on the difference between the calculated transformation from the vision algorithm and the known transformation.

Changes due to varying illumination, movement of objects in the environment or differing sensor readings possibly because of changes in the Earth's magnetic field (which, as shown by Azuma can vary with time even at the same location [19]), might occur between successive runs of an algorithm when using live footage, and would invalidate any results, therefore the use of pre-recorded footage allows for repeatable results.

The second evaluation approach will involve implementing a system to be used on the University of Waikato campus. The system will be non-real-time and will produce augmented snapshots of a scene, for example taking a snapshot of a campus building and overlaying a wire frame map directly on the building. The results will be viewed through either a laptop, tablet PC or PDA which have ergonomic benefits over many current, cumbersome, low resolution, high cost head mounted displays. The augmentations will be designed to show the registration accuracy of the system simply by visual inspection of its results.

Although the system will not be real-time, the computational requirements of the system will be considered. By analyzing these requirements a timeframe to when such a system would be possible in real-time will be determined.

4. Research Goals

The goals of this project are to:

- Extend the range of environments in which computer vision systems can work.
- Improve robustness by enabling the vision system to cope with greater variation in environmental conditions, and determine the degree of change the system can cope with.
- Determine the computational requirements of the system and a timeframe for feasibility in real-time.

5. Summary

This project will seek to provide a better understanding of the requirements of a vision system which provides accurate registration outdoors. The range of conditions and environments the vision system can operate in will be extended through the adaptive use of image processing and computer vision methods which will be chosen based on the scene in which the system is working.

6. References

[1] R. Azuma, B. Hoff, H. N. III, and R. Sarfaty, "A Motion-Stabilized Outdoor Augmented Reality System," presented at IEEE VR'99, Houston, Texas, 1999.

[2] G. Welch and E. Foxlin, "Motion Tracking: No Silver Bullet, but a Respectable Arsenal," IEEE Computer Graphics and Applications, vol. 22, pp. 24-38, 2002.

[3] S. You, U. Neumann, and R. Azuma, "Hybrid Inertial and Vision Tracking for Augmented Reality Registration," presented at IEEE VR `99, Houston, Texas, 1999.

[4] G. F. Welch, "Hybrid Self-Tracker: An Inertial/Optical Hybrid Three-Dimensional Tracking System," UNC Chapel Hill, Dept. of Computer Science Technical Report TR95-048, 1995.

[5] U. Neumann and S. You, "Integration of Region Tracking and Optical Flow for Image Motion Estimation," presented at IEEE International Conference on Image Processing, 1998.

[6] H. Kato and M. Billinghurst, "Marker Tracking and HMD Calibration for a Video-Based Augmented Reality Conferencing System," presented at 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR), San Francisco, California, 1999.

[7] K. W. Chia, A. D. Cheok, and S. J. D. Prince, "Online 6 DOF Augmented Reality Registration from Natural Features," presented at International Symposium on Mixed and Augmented Reality (ISMAR'02), Darmstadt, Germany, 2002.

[8] R. Behringer, J. Park, and V. Sundareswaran, "Model-Based Visual Tracking for Outdoor Augmented Reality Applications," presented at International Symposium on Mixed and Augmented Reality (ISMAR'02), Darmstadt, Germany, 2002.

[9] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 5, pp. 1266-1271, 1996.

[10] V. Vlahakis, N. Ioannidis, J. Karigiannis, M. Tsotros, M. Gounaris, D. Stricker, T. Gleue, P. Daehne, and L. Almeida, "Archeoguide: An Augmented Reality Guide for Archaeological Sites," IEEE Computer Graphics and Applications, vol. 22, pp. 52-60, 2002.

[11] R. Behringer, "Registration for Outdoor Augmented Reality Applications Using Computer Vision Techniques and Hybrid Sensors," presented at IEEE Virtual Reality, Houston, Texas, 1999.

[12] G. Simon and M.-O. Berger, "A Two-stage Robust Statistical Method for Temporal Registration from Features of Various Type," presented at 6th International Conference on Computer Vision, Bombay, India, 1998.

[13] G. Simon and M.-O. Berger, "Real time registration of known or recovered multi-planar structures: application to AR," presented at 13th British Machine Vision Conference, Cardiff, United Kingdom, 2002.

[14] E. Foxlin, "Inertial Head-Tracker Sensor Fusion by a Complementary Separate-Bias Kalman Filter," presented at IEEE Virtual Reality Annual International Symposium, 1996.

[15] A. K. Jain and C. Dorai, "Practicing vision: Integration, evaluation and applications," Pattern Recognition, vol. 30, pp. 183-196, 1997.

[16] R. Yang and G. Welch, "Fast Image Segmentation and Smoothing Using Commodity Graphics Hardware," Journal of Graphics Tools, vol. 7, pp. 91-100, 2002.

[17] R. Azuma, "Predictive Tracking for Augmented Reality," University of North Carolina, Chapel Hill, Technical report TR95-007, 1995.

[18] A. V. Cideciyan, S. G. Jacobson, C. M. Kemp, R. W. Knighton, and J. H. Nagel, "Registration of high-resolution images of the retina," SPIE: Medical Imaging VI, vol. 1652, pp. 310-322, 1992.

[19] R. Azuma, "The Challenge of Making Augmented Reality Work Outdoors," presented at First International Symposium on Mixed Reality, Yokohama, Japan, 1999.

Video Stereoscopic Avatars for the CAVE™ Virtual Environments

Lance Chong, Hank Kaczmarski, Camille Goudeseune
Integrated Systems Lab, Beckman Institute, University of Illinois at Urbana-Champaign
E-mail: lchong@uiuc.edu, hank@isl.uiuc.edu, camilleg@isl.uiuc.edu

Abstract

This proposed study is a design and development project based on the concept that by using stereoscopic video cameras together with a retro-reflective backdrop behind the user, it is possible to extract a CAVE™ or I-Desk™ user's stereoscopic image and integrate it into the 3D projection. This paper discusses various possible technical solutions that might make this concept work and their potential limitations. If proven workable, this concept could be an ideal first stage of bringing augmented reality into existing immersive virtual environments.

1. Introduction

Since the CAVE™ was first introduced by the University of Illinois at Chicago in 1992 [1] it has so far mostly been used to display computer synthesized 3D graphics. It could be potentially a huge usability expansion if stereoscopic videos can be integrated seamlessly with computer generated graphics in the CAVE™. A possible first step could be to use live stereoscopic video imagery of two remote CAVE™ or other VE users "to be their own avatars". Remote users can interact with each other as if they are in person, instead of using 3D avatar models. If this concept is proven workable, we will use the experiences gained in this test to conduct further experimentation in this direction. The CAVE™ can potentially be turned into an augmentation work-booth for human visual perception ideal for computer-aided remote collaboration.

There are past VR developments also considered intergrading 3D video into virtual environments such as the blue-c system developed by ETH in Zurich uses active visible lighting and "adaptive silhouettes" method to algorithmically subtract user stereo video image from dark background of a CAVE™-like environment [2]. Other previous work has successfully used motion histogram approach to track human motion through differentiating the silhouette of human figure using IR back projection onto a screen behind the user [3]. Because of their hardware design, both approaches put a lot of effort on developing software algorithm in differentiating user's video image outline from a VR hardware system background. We believe our approach of using retro-reflective material and digital infrared imagery will contribute to solving this problem more efficiently and reduce the space that will be required.

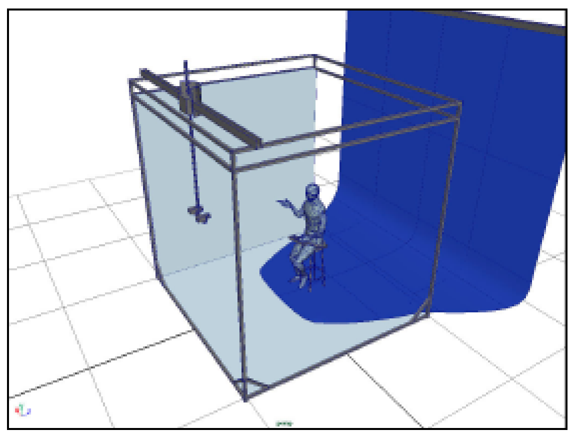

Figure 1. *Proposed layout of the Video Stereoscopic Avatar experimental setting for the CAVE™, with the retro-reflective backdrop extended to the CAVE™'s floor. A synchronized adjustable video cameras pair is in the user's front.*

2. Using "blue-screen" in the CAVE™

We propose to use a stereo video camera pair to capture stereoscopic live video of the users in the CAVE™. In order to do so, many factors need to be taken into consideration for achieving a visually comfortable final result that is usable for real applications. Above all, we need to find a way of extracting the user from the background in the video, in order to compose the user seamlessly into virtual environments. Setting up a normal blue-screen or green-screen next to the CAVE™ could be a solution. This motion picture and television special effects technique requires even and bright lighting that will very likely interfere with the CAVE™'s graphics display. We explore possible solutions next.

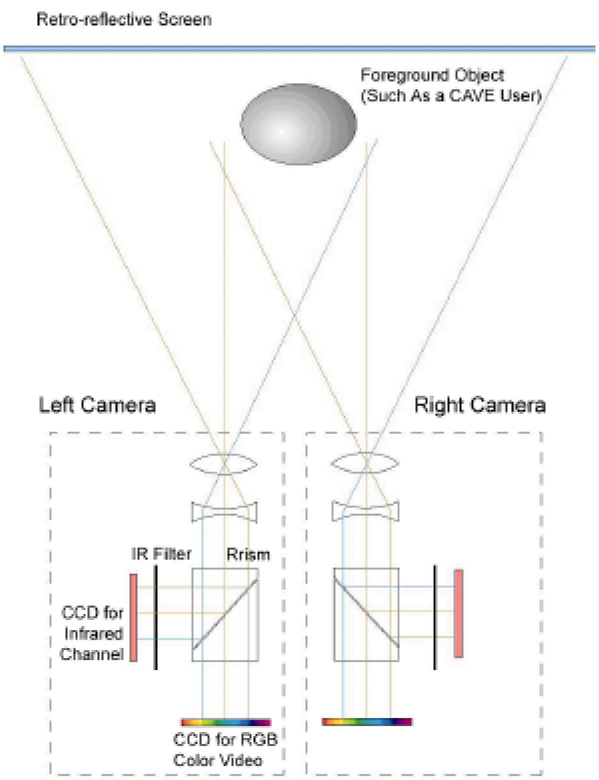

Figure 2. *Symbolic diagrams of the optical system design for using stereoscopic video inside the CAVE™.*

2.1. Retro-reflective Screen

Retro-reflective materials are designed to reflect light back to its incoming direction. If we replace normal blue-screen fabrics with a retro-reflective material, as shown in Figure 1., and place a blue or green light source close to a camera, from the camera's view, the retro-reflective screen will appear to be evenly lighted with blue or green as if in a traditional setting. Retro-reflective materials such as the "Scotchlite™" products made by 3M™ are widely available.

2.2. Infrared (IR) as invisible light source

With a retro-reflective screen, the need for intense, uniform lighting required by traditional blue-screen or green-screen can be eliminated. "White" (uncolored) Scotchlite™ fabrics are light gray, a color that matches well with the CAVE™'s projection wall and floor. But one light source of blue or green still needs to be placed near the video cameras and positioned between the user and front projection wall. In order to make the video cameras discreet, using infrared (IR) as an invisible light source is preferred. But to do so, a new set of issues must be considered.

Infrared video imagery with CCD camera and IR illuminator has been used in past VR and stereoscopic studies as ways to track user's head motion [4]. This gives us example of using IR imagery in VR settings, although our approach will only needs to track the outline of a user's image and could be algorithmically simpler.

2.3. Using IR as the alpha channel

Since we need an infrared channel as the alpha channel besides the normal visible light images in video, we will need to differentiate IR from visible light information. Normal CCD's convert full spectrum lights into red, green, and blue, three channels. They will also pick up infrared but how they interpret it into RGB channels still needs examination.

Our preliminary tests suggested that light emitted from several different kinds of infrared LED's are captured by normal CCD as "white" light. We have conducted several small-scale tests to evaluate the foundation of this design. In one of the setting, Four infrared LED's with wavelength of 880nm was positioned on to a Fujifilm® FinePix S1 Pro DDC still camera. Another setting uses night vision feature of a Sony DCR-TRV9 MiniDV hand-held video camera using the video camera's infrared illuminator as IR light source. Results from both settings suggested that the reflections from the retro-reflective backdrop are white in color and brighter than the IR reflections from the foreground objects. If the projected IR lights evenly covers the entire lens' viewing angle we should be able to get a gray-colored reflection from the backdrop, the brightness depends on the brightness of the camera-mounted light source. It is possible to generate a clean alpha channel from video images like this.

Our testing with both consumer CCD cameras and a Sony® XC HR50 progressive scan camera with a

Heliopan® RG830 infrared filter illuminated with a strobed array of 36 infrared LED's suggested that much better results could be achieved by using normal CCD video cameras to capture visible light imagery while using infrared video cameras to capture the IR reflections. We will consider hardware and software solutions according to further experiments. The goal is to find out if we can obtain video images of the retro-reflective backdrop to be relatively evenly lighted, constant in brightness, and easily differentiate from the foreground objects.

In order to make the system even simpler, we also have considered the possibility of finding special CCD's, if available, that can output "RGBI", four video channels. So far, we have concluded that a more practical approach, as shown in Figure 2., is, to use a prism to split the incoming light beam in two, one targeting a normal CCD camera that outputs RGB channels from visible light, the other targeting an infrared camera with an IR filter which outputs only IR imagery to be used for the alpha channel.

2.4. Differentiating multiple IR light sources

In theory, it is possible that extra IR lights will interfere with existing IR signals, such as those used to synchronize the shutter glasses with the stereo projection. So far, in our preliminary tests, the infrared sources we have tested do not interfere with the stereo shutter glasses in our CAVE™ system. In case interference occurs, possible solutions could be either to use IR lights of different wavelength, or to polarize different light sources using filters or prisms.

2.5. Possible alternative methods

We will continue to consider the possibilities of using alternative methods during our research and development. For example, one of the directions on our desk was to use a backdrop that evenly emits strong infrared light to the direction of the cameras. But technically, this could be even more challenging .We will need to build a backdrop that is either the IR light source or evenly transmits IR from other light sources. For safety reasons, this backdrop obviously cannot generate excessive heat.

3. Stereoscopic video imagery

Extensive research has been done in the area of stereoscopic photographic and stereoscopic computer imagery. Knowledge from these studies is extremely helpful for to customize a high quality stereoscopic video setting in the CAVE™.

3.1. Stereo camera calibration

For a standard size CAVE™, the distance from a user to the stereo camera is estimated to be between four and seven feet. Therefore, we will be working with near stereoscopic viewing, where stereoscopic depth distortions will likely be an important factor. The stereo video camera pair needs to be adjustable and allow precise control of its inter-viewpoint distance, convergence distance, and other system parameters [5]. Many other CAVE™-related factors, including variable ambient light levels and color temperature will need to be taken into consideration. Specific engineering design is very likely necessary.

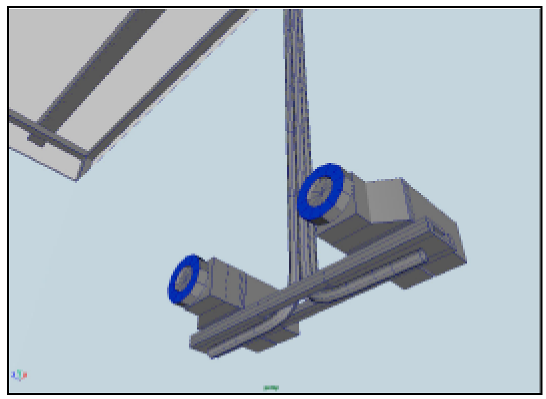

Figure 3. *A possible design of the fully adjustable stereoscopic video camera unit for the CAVE™ with infrared LED rings mounted in front of each lenses.*

3.2. Video resolution and lighting

Current CAVE™ wall projection resolution varies from 1280x942 pixels to 1024x764 pixels depending on the capabilities of the graphics engines and complexity of the applications. Normal NTSC video format is 648x486 pixels in resolution, which might need to be rotated by 90 degrees from "landscape" to "portrait" and stretched to compensate for the CAVE™ projectors' pixel distortion to fill a square screen.

A working CAVE™ is a low light environment with primary lighting from the imagery projections that are subject to change. In order to produce clear and stable video of the user, the cameras need to be very sensitive. We can arrange extra lighting for the user as long as it does not interfere with the image projections. A software filter that can dynamically compensate the stereo video based on graphics card outputs could also be helpful. However, the video avatar's brightness and color change resulted from local projection could be an interesting feature for certain applications.

3.3. Focal length and camera position

The camera-user distance will be mostly between four to seven feet. In order to cover the user from head to toe, wide-angle lenses will be needed. To allow a local user and an avatar to look at each other "in the eyes" without the cameras blocking their faces, the unit needs to be small enough and placed between the user and the avatar with a height slightly above local user's eye level.

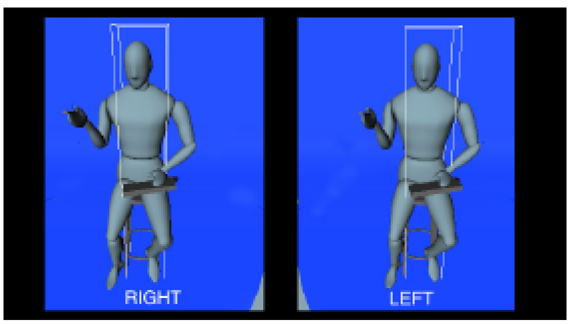

Figure 4, *Simulated left and right views form a stereoscopic video camera set with 28mm focal length. The light color box above the manikin's shoulder indicates a standing height of sit feet and two inches.*

4. Designing for Applications

With this project, we intend to find a way in virtual interaction that allow remote users virtually sit face to face and work as if they were locally in a studio setting. The primary function of the CAVE™ will be changed from content demonstration and evaluation to real time content creation. That means we need to design software and hardware interfaces to accommodate changes in the user's behavior. For example, we will consider ergonomic issues including designing seating and adding a wireless keyboard/mouse and other input devices so as to accommodate possible user demands during multi-hour "in-person" virtual collaborations. There will be plenty of new possibilities to design and develop new interfaces and functionality and challenges to understand the implications of users being immersed in a VE for longer times than in previous scenarios.

5. Possible future works

If proven workable, we believe this experiment will open the doors to many other possible developments including: mobile stereo panorama video camera units that can be dispatched to remote locations which allow user in the CAVE™ to observe from the camera unit's point of view, software and hardware interface design for control and navigation, composing computer synthesized 3D and 2D graphics with stereo video from multiple camera units, recording stereo video footage and editing in real-time, analyzing data from stereo video and use them to create synthesized graphics on top of stereo video. We believe studies in this direction could potentially be valuable to human-computer interaction and 3D imagery research.

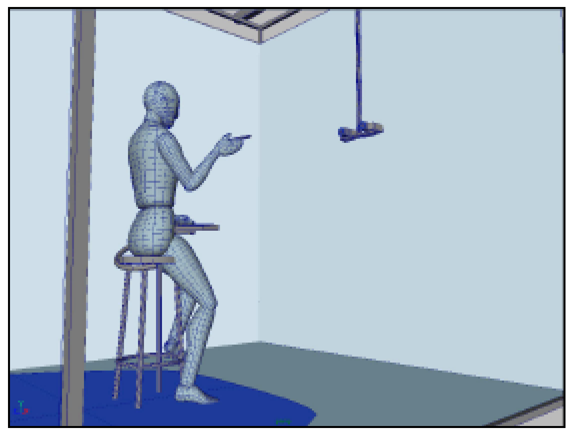

Figure 5. *A user will be working in front of a stereo video camera unit in the CAVE™, the live stereoscopic video footage of one or more remote users could be displayed on the wall to assist remote collaboration.*

6. References

[1] CAVE™ Overview. (n.d.). Retrieved July 18, 2004, http://www.evl.uic.edu/pape/CAVE/oldCAVE/CAVE.overview.html

[2] M. Gross, S. Würmlin, M. Naef, E. Lamboray, C. Spagno, A. Kunz, and E. Koller-Meier, "Blue-C: a Spatially Immersive Display and 3C Video Portal for Telepresence", *ACM Transactions on Graphics (TOG)*, Volume 22 issue 3, Association for Computing Machinery, New York, NY, July 2003.

[3] J. W. Davis and A. F. Bobick, "Sideshow: A silhouette-based interactive dualscreen environment." *Technical Report 457*, MIT Media Lab, August 1998.

[4] T. Hattori, T. Ishigaki, K. Shimamoto, A. Sawaki, T. Ishiguchi, and H. Kobayashi, "An Advanced Auto-stereoscopic Display for G-7 Pilot Project", *Stereoscopic Displays and Virtual Reality Systems VI, Proceedings of SPIE, Vol. 3639,* SPIE, Bellingham, Washington, January 1999.

[5] D.B. Diner, and D H. Fender, *Human Engineering in Stereoscopic Viewing Devices*, Plenum Press, New York, 1993, pp. 179.

Third-Person Wearable Design Considerations

Clint Heyer
School of Information Technology and Electrical Engineering
University of Queensland, 4072 Brisbane Australia
clint@itee.uq.edu.au

Abstract

Wearable design is often done from a selfish viewpoint – developing functionality to support a single person. This paper proposes to also consider those around the wearable user. Not only is there opportunity for a wearable to be a benevolent confluence hub, it is hoped that the increased understanding of the device will lead to greater acceptance and desirability of the platform. Additionally, increased focus on "unaugmented" external users by way of tangible or ambient interaction will provide new opportunities for social computing applications on a wearable platform.

1. Introduction

Much of wearable computing literature is concerned with how a wearable computing platform can enhance an individual's abilities, or how a group of wearable-equipped people can better interact or perform tasks. Little work however, has been done on the question of how the platform can be used to benefit non-wearable computer users, or indeed what are the differing design considerations that take these people into account.

Given that it will be some time before widespread adoption of the type of wearable computers seen in literature (such as those with a head-mounted display and mobile input), wearables need to be able to exist harmoniously in a world of mixed technologies. Furthermore, wearables are personal devices, and people are social beings. With the enhanced functionality of wearable computers should come enhanced focus on making wearables sociable: becoming a positive social influence rather than a disruptive influence.

This work proposes utilising user-centred design principles to investigate wearable design from the perspective of external users - those that will co-habit the social context of the wearer. It is hoped that the wearable platform that can bring people together and provide useful, accessible services, even if others only have simple devices available, or indeed any device at all.

2. Proposal

Within the rich social environment that people exist, wearables will also commonly exist. Studies of wearables already in the social environment have shown that observers have negative, or otherwise incorrect perceptions of the wearer or the wearable [1, 2]. However, once the functionality of the device is described perceptions change and "'geekiness' becomes 'cool'" [2], presumably due to the palpable value of the features. A possible reason for observers misjudging wearables in previously reported work is because the systems were essentially self-serving ones. These wearables are exclusively personal devices, recording private moments and thoughts, and were not to be shared physically or virtually. It is easy to appreciate observers' suspicions of wearables when their input and output are hidden from notice, and easier to appreciate how they form incorrect mental models of its functionality. Is it the case that wearables are self-serving by definition?

Providing publicly-accessible services (with or without the wearable user's required intervention) is one possible method of partially addressing these concerns. Services shared will depend on the user's own preferences and access control lists, however providing anonymous[1] access to services should be encouraged where appropriate. To enable services, or during service use there should be a minimum impingement on the wearable user. Availability domains can also be used to restrict where and to whom services are available. For example, location-aware services could be provided to those physically collocated with the wearable without necessarily being detrimental to the privacy of the wearable user. Remote clients however may only be able resolve the wearable's location to a very coarse degree – if at all. There is also the potential for some services to be provided to those without a device on their person, in an ambient or tangible manner, as well as potential for co-opting existing social interaction cues or behaviours as the interface to the system.

Additionally, making the functionality of the wearable plainer by utilising embodied and tangible design principles could also be beneficial. External observers can

[1] In the File Transfer Protocol sense of the word; that is, users don't require credentials to use the service

become anxious as they are not able to discern how the wearable is being used, or which features are enabled [1]. An example is that of personal privacy: as wearables are often used in private settings and can feature video and audio input, observers and those interacting with the wearable user are understandably concerned as to how their image and voice is used. Is what they say recorded? Is their image being broadcast live to a website? These questions are valid ones, and are currently answered on a case-by-case basis by the wearer, depending on the functionality and status of the wearable. It is a goal of this work to discover design methods for allowing observers to ascertain the answers themselves.

As an elementary example, a wearable could have an externally-facing panel or display that indicates - at an abstract level - the activity in which the wearable is currently being used such as *reading, recording*, *searching*, *broadcasting* or *note-taking*. Such a display could inform observers not only of the active task, but of potential tasks, thus providing the lacking affordance cues [3]. This goal of externalisation attempts to invert the antisocial, inscrutable designs that have come about through the common use of single-user, private input/output devices such as monocle head-mounted-displays

It is hoped that by increasing the perceived functional value of the wearable to both the user and those that encounter the wearable or its services, the desirability of the platform will increase, as demonstrated by [4]. With increased public acceptance and positive "aggregate social consequence" [5], it is hoped that the current negative perceptions that wearable users anticipate [6, 7] will dissipate.

3. Related work

Recent work discusses augmenting social networks with a goal to provide services that kick-start traditional social interaction rather than simply providing information exchange within the community [8]. The work assumed that there would be a community of users with wearable computers, however many of the services described could be implemented on a basic wearable device such as a PDA.

The proposed use of externalised design principles to give the system a degree of design transparency is diametrically opposed to the goals of the e-SUIT [5] which seeks to provide methods of covert interaction with the wearable. For signalling by a system to be effective, it generally has to draw the attention of the user. Even when this signal is exclusively localised to the user (such as the vibrators in the e-SUIT), the resulting loss of mental focus can degrade social interactions, disrupting the masquerade of covert interaction. Work however on subliminal visual signaling that has shown to be effective [9] and may be viable even during social interaction.

4. Plan

The plan for this work is as follows. A literature survey will be conducted in the areas of wearable, mobile and pervasive computing, CSCW as well as social psychology to inform subsequent work. It's to be noted that the bulk of the research method - and thus the solution - will be left open as the process will be driven by the exploratory participatory design method [10]. By the use of this method, it is hoped that the developed theoretical position and designs will have a concrete user-centred foundation.

Desired initial services to be provided by the prototype will be identified by way of rapid participatory design and ethnographic methods. Once potential services are identified, a hardware specification can be formulated, code written and a rough prototype assembled. It is envisaged that the system will be used day-to-day by the author once functional at a basic level.

A broad survey will be conducted to examine people's preconceptions towards wearable computers, and those that use them. As it is expected that few have had previous contact with a wearable computer, this study will be done whilst the prototype system is in use, so participants can judge the system first hand. It is hoped that further data can be gathered during the course of the work by inviting strangers to fill out a survey after a normal everyday interaction with the author and wearable.

Additional participatory design sessions with external users of the wearable will be held to investigate the form and functionality of an externalised component to the wearable. The exact method for these sessions will also have to be investigated.

During use the wearable will log user and client access details (as per university privacy and ethical policies), and a reflective journal kept to record experiences as they happen.

Numerous people will be close to the author over the study's period so periodic questionnaires will be performed to discover how perceptions change with increased exposure to the system, and informal feedback about the design of the system collected. This feedback will form part of an iterative design process to evolve the system based on actual usage. At the end of the design process, a final study across a broad selection of people will be performed to validate the design.

As well as the prototype, it is hoped that a theoretical foundation of "third-person" or social design considerations within the wearable context can be developed.

5. Acknowledgements

I would like to thank my supervisor, Dr. Margot Brereton for her instrumental guidance, and for enabling and supporting this method of study.

6. References

[1] T. Starner, B. Rhodes, J. Weaver, and A. Pentland, "Everday-use Wearable Computers," vol. 2004. Atlanta: Georgia Tech, 1999.

[2] J. G. Sheridan, V. Lafond-Favieres, and W. C. Newstetter, "Spectators at a geek show: an ethnographic inquiry into wearable computing," presented at Wearable Computers, Fourth IEEE International Symposium on, 2000.

[3] D. Norman, *The Design of Everday Things*. New York: Perseus, 2002.

[4] K. Bodine and F. Gemperle, "Effects of Functionality on Perceived Comfort of Wearables," presented at Wearable Computers, Seventh IEEE International Symposium on, White Plains, NY, USA, 2003.

[5] A. Toney, B. Mulley, B. H. Thomas, and W. Piekarski, "Social weight: designing to minimise the social consequences arising from technology use by the mobile professional," *Personal and Ubiquitous Computing*, vol. 7, pp. 309-320, 2003.

[6] G. Korteum, M. Bauer, and Z. Segall, "NETMAN: The Design of a Collaborative Wearable Computer System," *Mobile Networks and Applications*, vol. 4, pp. 49-58, 1999.

[7] J. F. Knight, C. Baber, A. Schwirtz, and H. W. Bristow, "The comfort assessment of wearable computers," presented at Wearable Computers, Sixth IEEE International Symposium on, 2002.

[8] G. Korteum and Z. Segall, "Wearable communities: augmenting social networks with wearable computers," *Pervasive Computing*, vol. 2, pp. 71-78, 2003.

[9] R. W. DeVaul, A. Pentland, and V. R. Corey, "The memory glasses: subliminal vs. overt memory support with imperfect information," presented at Wearable Computers, Seventh IEEE International Symposium on, White Plains, NY, USA, 2003.

[10] C. Floyd, Wolf-Michael Mehl, Fanny-Michaela Reisin, Gerhard Schmidt and Gregor Wolf, "Out of Scandinavia: Alternative Approaches to Software Design and System Development," *Human-Computer Interaction*, vol. 4, pp. 253-350, 1989.

Laser Triangulation as a means of robust Visual Input for Wearable Computers

Patrick de la Hamette, Marc von Waldkirch, and Gerhard Tröster

Wearable Computing Lab, ETH Zürich, Switzerland,
www.wearable.ethz.ch
pdelaham, waldkirch, troester@ife.ee.ethz.ch

Abstract

New kinds of input devices in the area of wearable computing try to capture movement and gestures of the user's hand, for human computer interaction. They often lack robustness in the wearable environment, especially in bright sunlight or they fail to capture more than a contour or a relative movement. In this paper, we present a methodology for 3D sampling of a human hand by a wearable active vision device using laser triangulation. We discuss how the method remains robust while operating under sunlight and how a low power consumption is achieved. We show how a hardware system could be built, estimate the power needed and show an experimental setup. Finally we conclude that this novel method for a wearable computer interface is feasible and promising.

1 Introduction

With the use of wearable computers, new kinds of human computer interaction methods are being employed. One area deals with using the bare hands of the user to communicate with the computer, capturing hand movements and hand signs and gestures [6]. Some projects use motion sensors to detect hand motions [2], others use vision, with one or more cameras and sometimes active vision [7].

The vision devices, like the *Fingermouse* [3], are normally meant to be worn on the body, filming the user's hand, as depicted in figure 1.

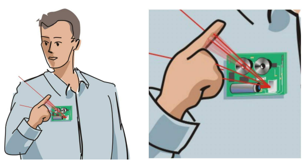

Figure 1. usage

However, most of those system have drawbacks of some kind. A major issue is robustness. Vision system have problems in the wearable environment such as moving background with clutter and extreme lighting conditions in sunlight. Other systems can only provide information about the two dimensional contour of the user's hand.

We propose a methodology that is robust and can provide a 3D sampled image of the user's hand, with a limited amount of $N \times N$ samples. This information can be used to recognize hand shapes or to assist other vision systems.

2 Approach

principle Our approach uses active vision, as shown schematically in figure 2. We project a grid of dots on the user's hand. A camera sensor captures a picture of the dot pattern on the hand. From the position (x,y) of the dots in this picture , we aim to calculate the real world coordinates (X,Y,Z) (Z being the range, or depth. c.f. figure 3) of these dots, thus representing a **3D sampling of the hand.** [5]

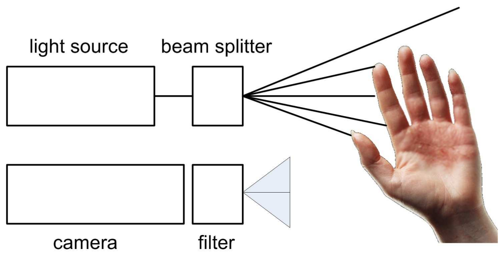

Figure 2. principle

system components and parameters For an easy extraction of the dots in the captured image, we want the dots to be brighter than the rest of the image. We use a laser diode, as a monochromatic light source. The wavelength is chosen to match a band, where solar irradiation is relatively low, thanks to atmospheric absorptions. The laser beam is split into a grid of $N \times N$ beams, using a diffractive optical element. A narrow bandpass filter in front of the camera attenuates light of other wavelengths [1]. To further save energy, the laser can be triggered to fire only during image capturing, and reduce its power in darker lighting conditions.

Thanks to the filtering, triggering and careful selection of the system parameters, we hope to be able to build an active vision system that works outdoors and has a low power consumption, as required for wearable applications. The required laser power will be roughly estimated in section 3.

3D sampling Since the geometry of the laser beams and the camera parameters are known, the real world coordinates (X,Y,Z) can be derived from the coordinates (x,y) of the dots in the acquired image.

We propose a special geometrical setup of the system, where the light source and the projected grid are parallel to the optical axis of the camera. For a given dot, its location is always on the same horizontal line in the image. The depth Z derives directly from its position x (if the camera is left to the laser, then the dots will move to the right, as the foreground object approaches, c.f. figure 3), thus reducing the depth mapping to a one-dimensional problem.

The problem of identifying the dots of the grid can be solved by using a separate range sensor, which identifies the approximate depth of a center dot. Once a dot's depth is known, its neighbors can be identified as well. This solves the problem, assuming that neighboring dots have a limited difference in their depth.

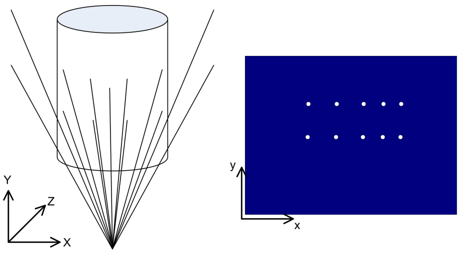

Figure 3. geometry of the triangulation: scene (left), captured image (right)

a possible hardware implementation A hardware system can be implemented as as single, integrated wearable device. This device contains the light source, the sensors and the processing unit which detects the dots. A possible architecture is shown in figure 4.

A black and white image sensor should be used, because single chip color sensors are relatively inefficient, in terms of resolution and light sensitivity, when capturing the monochromatic images. This is due to their color pixel arrays (e.g. bayer pattern), with only some of their pixels being sensitive to monochromatic light of a specific wavelength.

Although the beams are harmless at a certain power (c.f. section 3), we propose to build the system such that the laser beams point downwards, in order to avoid bothering people in the environment. To ensure the orientation, we integrate a tilt sensor. If the beams don't point downwards, the laser is turned off.

The dot detection can be done on the fly, meaning that the dot center coordinates (x,y) are measured during the camera sensor readout. Depending on the size of a laser dot in the captured image, a few line image line buffers are required, but there is no need for a frame buffer, thus saving energy and space (RAM chip). The task can be done by a microcontroller, an FPGA or an ASIC. Only the dot-coordinates will be transferred to the wearable computer, where high-level algorithms further use the information, e.g. to perform sign recognition. A detection of the (x,y)-image coordinates should be sufficient, as the conversion to (X,Y,Z)-real world coordinates can easily be handled by the external computer.

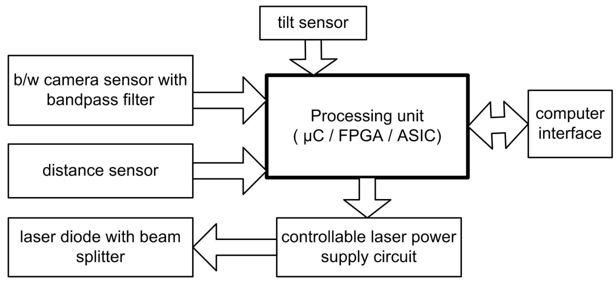

Figure 4. system architecture

3 Irradiance Estimation

The goal of this section is to estimate roughly the minimum laser power P_{laser} needed for a robust use of the triangulation method even in bright ambient lighting conditions. To this end, the laser irradiance I_{dot} (measured in $\mathrm{W/cm^2}$) of a laser dot on the human hand is directly compared to the disturbing irradiance I_{sun} which comes from direct sunlight. For this comparison we assume that both the laser light as well as the sunlight are diffusely reflected on the skin with a similar reflection rate. This assumption is realistic as only a narrow frequency band of the sunlight will pass the filter. As the laser light is nearly monochromatic, the dot irradiance I_{dot} can be expressed as follows:

$$I_{dot} = \frac{P_{laser}}{N^2 A_{dot}} \eta T(\lambda_0) \tag{1}$$

where A_{dot} is the cross-section of a laser dot on the hand and $T(\lambda_0)$ the transmission rate of the narrow-band filter at the laser wavelength λ_0. Further, the diffraction grating is assumed to provide a matrix of $\mathrm{N} \times \mathrm{N}$ dots of uniform intensity with a diffraction efficiency η.

In the similar manner the background irradiance I_{sun} is given:

$$\begin{aligned} I_{sun} &= \int_\lambda i_{sun}(\lambda) T(\lambda) d\lambda \\ &\approx i_{sun}(\lambda_0) T(\lambda_0) \Delta\lambda \end{aligned} \tag{2}$$

Here, $i_{sun}(\lambda)$ denotes the solar spectral irradiance (measured in $\mathrm{W/(cm^2\,nm)}$ of the sun on the ground and $\Delta\lambda$ the bandwidth of the narrow-band filter.

In order to apply the simpler threshold detection method and still guarantee a reliable detection the signal-to-noise ratio SNR should be larger than 1:

$$SNR = I_{dot}/I_{sun} \gtrsim 1 \tag{3}$$

Thus, the minimum laser power turns out to be:

$$P_{laser} \gtrsim \frac{i_{sun}(\lambda_0) A_{dot} N^2 \Delta\lambda}{\eta} \tag{4}$$

Eq. 4 shows that the laser power may be minimized by when reducing the filter's bandwidth or the number of dots in the matrix. A reduction of the latter, however, affects the accuracy of the triangulation. Another parameter that allows optimization is the frequency λ_0, it should be chosen in a band where solar irradiation is low, due to atmospheric absorptions.

To have an idea about the magnitude of P_{laser}, it is roughly estimated by means of eq. 4: The solar spectral irradiance at $\lambda = 670\,\mathrm{nm}$ is about $0.1\,\mathrm{mW/(cm^2\,nm)}$ [4]. For the narrow-band filter a bandwidth $\Delta\lambda \approx 10\,\mathrm{nm}$ is a realistic value while the diffraction grating was assumed to provide a 19×19 dot matrix with a diffraction efficiency $\eta \approx 70\%$ and dots with $1\,\mathrm{mm}$ in diameter on the hand. In that case the laser power turns out to be $P_{laser} \approx 4\,\mathrm{mW}$. The power for a single beam is less than $10\mu\mathrm{W}$, which poses no danger to skin or eyes, even for direct intrabeam viewing.

4 Experiments

As a proof of concept, we implemented an experimental setup of the described system. We use a $5\,\mathrm{mW}$ laser diode, emitting at $\lambda_0 = 670\,\mathrm{nm}$ (red light). The diffractive optics splits the beam into a grid of 19 x 19 beams. We used a bandpass filter with $\Delta\lambda = 20\,\mathrm{nm}$, centered at $670\,\mathrm{nm}$.

Using this setup, we acquired images of a model hand, under different light conditions, including rooms lit by bright sunlight. However, the system could not outperform outdoor sunlight falling directly on the hand, due to the width of our bandpass filter and the laser beams which were wider than we expected. Figure 5 shows captured images with a model hand.

5 Conclusions

Both the the theoretical power estimation and the experiments show that the method is not only feasible in low power, but also promising to be a robust means of implementing input interface to wearable computers. Projects like [3] show that such a system could be built as a compact, integrated device, as described in section 2.

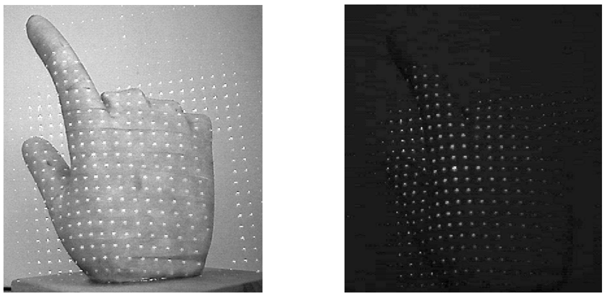

Figure 5. captured image, with (right) and without (left) bandpass filter

6 Outlook

For further research, more experiments with different laser diodes and diffractive optics are required, in order to find an optimal solution that works in all kinds of environments.

The second step would be to evaluate acquired pictures and implement algorithms to detect the dots in the image and convert them to real world coordinates, with as much precision as possible. Once the algorithms are known, the next step could be to design a wearable hardware prototype.

References

[1] T. Balch, B. Wilcox, and L. Matthies. Fast optical hazard detection for planetary rovers using multiple spot laser triangulation. In *Proceedings, IEEE International Conference on Robotics and Automation*, pages 859–866, 1997.

[2] W.-C. Bang, W. Chang, K.-H. Kang, E.-S. Choi, A. Potanin, and D.-Y. Kim. Fast optical hazard detection for planetary rovers using multiple spot laser triangulation. In *Proceedings, Seventh IEEE International Symposium on Wearable Computers, 2003*, pages 26 – 34, 2003.

[3] P. de la Hamette, P. Lukowicz, and G. Tröster. Fingermouse: A wearable hand tracking system. In *Adjunct Proceedings of the 4th International Conference on Ubiquitous Computing, 2002*, pages 15–16, 2003.

[4] C. A. Gueymard, D. Myers, and K. Emery. Proposed reference irradiance spectra for solar energy systems testing. *Solar Energy*, 2002.

[5] T. Pajdla. BCRF - binary-coded illumination range finder reimplemetation. *Technical report KUL/ESAT/MI2/950, Katholieke Universiteit Leuven*, 1995.

[6] J. P. A. Starner, T.; Weaver. A wearable computer based american sign language recognizer. *Digest of Papers, First International Symposium on Wearable Computers*, 1997.

[7] M. Turk. Computer vision in the interface. *Communications of the ACM*, 2004.

Projection Augmented Models: The Effect of Haptic Feedback on Subjective and Objective Human Factors

Emily Bennett
University of Portsmouth
Emily.Bennett@port.ac.uk

This paper outlines the PhD entitled 'Projection Augmented models: The effect of haptic feedback on subjective and objective human factors'.

1. Research question

Figure 1. PARIS with and without projection

A Projection Augmented model (PA model) is a type of projection based Augmented Reality display. It consists of a physical three-dimensional model, onto which a computer image is projected to create a realistic object [1]. An example of a PA model is the Projection Augmented Reality Information System (PARIS) (figure 1). It is worth noting that a defining characteristic of PA models is that the image is projected onto the three-dimensional physical model. Therefore PA models are different from systems that project an image onto a planar surface around physical objects.

People can touch the surface of a PA model with their bare hands; therefore PA models naturally support whole-hand haptic feedback. The concept of using real objects to provide haptic feedback has been shown to be beneficial to users in immersive virtual environments [2] [3].

Research surrounding PA models has focused on developing the enabling technology, for example the Shader Lamps project [4]. The Dynash project is the only project known to have focused on human factors [5]. However, the Dynash project does not investigate how the haptic feedback provided by PA models can benefit their users. This is an important issue because PA models provide a very high level of haptic feedback in comparison to other displays. Therefore an investigation into this issue may uncover the potential of PA models. This research project addresses this issue; it investigates the effect haptic feedback has on human factors.

The approach taken to investigate this issue was to focus on 'basic' human factors. This is because a main application area for PA models has not yet emerged. When conducting human centered studies on new displays that do not have a main application area, it is advisable to evaluate them in terms of 'basic tasks', such as perception and interaction speed [6]. Such tasks can be generalized to many situations and can guide application development. Both subjective and objective human factors were studied to ensure a broad picture was obtained. It should be noted that this approach differs to that of the Dynash project, which focuses on the use of PA models for design applications [5]. However, the two research projects complement each other by providing different perspectives on how people use PA models.

Any user interface system needs both a display and an interaction device, therefore both haptic feedback and haptic interaction were investigated. The type of haptic interaction devices that are of particular interest are those which allow the user to interact with a PA model by physically touching its surface with a tool, thus directly haptically perceiving the PA model. Such devices can be referred to as spatially-coincident devices. Therefore the research question is:-

What effect does touching a PA model and/or interacting with it using a spatially-coincident device have on subjective and objective human factors? In comparison to not touching it and interacting with it using a spatially-separated device.

2. Theoretical investigation

There are several types of PA models. They can either be front-projected where the image is projected from the front, or back-projected where the image is projected from behind. PA models can also be categorized depending on whether or not they have a mismatch between the visual and the haptic feedback. All current PA models have a mismatch because they have difficulty simulating physical texture, as the optimal projection conditions are a smooth, diffuse surface. However, it should be noted that it is possible to give the PA model a physical texture that matches its visual texture. This can only work effectively when simulating objects that in real life are made from a very similar material as the PA model.

To study the research question, initially front-projected PA models that have a smooth physical surface were focused on. This is because it is likely that different types of PA models may have different effects on human factors. Front-projected PA models were focused on because the aim of this research is to conduct user testing on the current technology, and current technology has focused on developing front-projected PA models. Smooth surfaced PA models were focused on because PA models that have an accurate physical texture can only be created effectively for a limited range of objects, which indicates this technique may have limited real world applications. Note, when the term 'PA model' is used in this section and the next section, it will be referring to front-projected smooth PA models.

PA models provide haptic feedback for geometric properties, such as size and shape. However, they do not provide haptic feedback for material properties, such as physical texture and temperature. It has been found that the visual sense dominates when perceiving geometric properties, and the haptic sense dominates when perceiving material properties [7]. This suggests that touching a PA model or interacting with it using a spatially-coincident device will have little effect on perception. However, there is evidence to suggest this may not be the case.

Recently it has been it has been found that touch and kinesthetic feedback can increase the accuracy of the perception of geometric properties, such as size [7]. This suggests a PA model may be more accurately perceived when it is touched or interacted with using a spatially-coincident device.

The subjective factor of object-presence is important because PA models aim to create realistic objects. Object-presence is the subjective feeling a particular object exists in a person's environment, when that object does not [1]. It has been shown that haptic feedback can increase presence [3], which suggests that touching a PA model or interacting with it using a spatially-coincident device may increase object-presence. However, the haptic sense is very sensitive to detecting an object's material properties [7], which suggests that the opposite may occur. All current of the spatially-coincident interaction devices that can be used with PA models are single-point-of-contact devices. Therefore they provide a lower level of haptic feedback of the PA models texture, compared to touching it with a bare hand. This suggests that object-presence may be affected differently by touch and interaction using a spatially-coincident device. An additional factor that may reduce object-presence is that when a PA model is touched or interacted with using a spatially-coincident device, the projected image appears on the back of the hand. Touch or interaction using a spatially-coincident device may also effect user preference factors, such as enjoyment and engagement.

A type of spatially-coincident interaction device that is of particular interest is a Tangible User Interface (TUI). A TUI allows a person to use small physical objects to interact with digital information. By using a TUI to interact with a PA model, an interface can be created that uses physical objects to both display and interact with digital information. TUIs utilize real world interaction skills, therefore they may increase subjective factors such as object-presence and user preference. TUIs that operate on a planar surface have been shown to support faster interaction than an equivalent GUI [8]. Therefore it is also possible that TUIs used with PA models will support fast interaction.

3. Empirical investigation: Study 1

Several questions arise from the discussion in the previous section:-

Does touching a PA model and/or interacting with it using a spatially-coincident device effect size perception, object-presence and user preference? In comparison to not touching it and/or interacting with it using a spatially-separate device.

Does interacting with a PA model using a TUI effect speed and accuracy of interaction, object-presence and user preference? In comparison to interacting with it using a mouse?

A controlled experiment was conducted to study both of these questions together. This is possible because a TUI is a type of spatially-coincident and a mouse is a type of spatially-separate device. The main conclusions of this study were:-

- The mouse supports faster interaction than the TUI.

- Accuracy of size estimates are higher when a PA model is touched, compared to when it is just looked at [9].
- User preference is highest when a TUI interaction device is used.
- Object-presence is highest when a PA model is not touched and it is interacted with using a spatially-separate device.

The last point is interesting because it may have been caused by the image being projected onto the back of the participant's hand when they touched the PA model. Alternatively, it could be because when they made contact with the PA models surface, they realized its physical texture does not match its visual texture.

4. Future empirical investigation: Study 2

This research projected started by focusing on front-projected smooth PA models. However, the issue raised in study one suggests that if a back-projected or accurately textured PA model was used, different results may have been obtained. Therefore study two will investigate the following question:-

When using a PA model, is object-presence affected by the image being projected onto the back of the hand, or by the mismatch between the visual and haptic texture?

The results from this study can be used to guide the direction in which technology needs to be developed if object-presence is to be maximized. A 2x2 factorial experiment will be used to investigate this question. The two independent variables (IV) are *direction of projection*, and *physical texture of a PA model*. A different PA model will be used in each condition, the four conditions will be:-

- Front-projected with a smooth surface.
- Front-projected with a physical texture that closely matches its visual texture.
- Back-projected PA model with a smooth surface.
- Back-projected with a physical texture that closely matches its visual texture.

Participants will complete the same task in all four conditions, which will require them touch the PA model. They will then give a rating of the sense of object-presence they feel.

5. Conclusion

PA models support a very high level of haptic feedback in comparison to other displays. Study one found that haptic feedback was beneficial to all factors except object-presence, when a front-projected smooth surfaced PA model was used. Study two will investigate why object-presence was reduced by haptic feedback in study one. The results from this study will suggest which type of PA model should be used if high object-presence is required.

6. References

[1] B. Stevens, J. Jerrams-Smith, D. Heathcote, and D. Callear, Putting the Virtual into Reality: Assessing Object-Presence with Projection-Augmented Models. *Presence: Teleoperators and Virtual Environments*, 2002, 11(1), pp. 79-92.

[2] A. Boud, C. Baber and S. Steiner (2000). Virtual Reality: A Tool for Assembly? *Presence: Teleoperators and Virtual Environments*, 2000, 9(5), pp. 486-496.

[3] B. Insko, M. Meehan, M. Whitton and F. Brooks, Passive haptics significantly enhances virtual environments. Presence 2001. In *Proceedings of the 4th International Workshop on Presence, 2001*.

[4] R. Raskar, G. Welch, K. Low, and D. Bandyopadhyay, Shader Lamps: Animating Real Objects With Image-Based Illuminations. In *Proceedings of the 12th Eurographics Workshop on Rendering,* Springer, 2001.

[5] J. Verlinden, W. Esker, L. Wind, and I. Horváth, Qualitative comparison of virtual and augmented prototyping in handheld products. In *Proceedings of International Design Conference – Design 2004*, Dubrovnik, 2004.

[6] A. Kirkpatrick, and S. Douglas, Application-based evaluation of haptic interfaces. In *Proceedings of 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 2002, 32-40.

[7] M. Ernst and H. Bülthoff, Merging the senses into a robust percept. *Trends in Cognitive Sciences*, Elsevier, 2004, 8(4), pp. 162-169

[8] Huang, C. (2004). Not Just Intuitive: Examining the Basic Manipulation of Tangible User Interfaces. In *Proceedings of CHI'04*. 2004, 1387-1390.

[9] E. Bennett and B. Stevens, Haptically perceiving and interacting with Projection Augmented models. In *Proceedings of ISMAR 2004.* (in press).

Wearing Embodied Space

Erik W. Conrad
University of California Irvine
Donald Bren School of Information and Computer Science
Arts, Computation and Engineering (ACE)
econrad@uci.edu

Abstract

Technologies are not mere exterior aids, but interior changes of consciousness that shape the ways the world is experienced [7]. As we enter the age of ubiquitous computing, where computers are worn, carried or embedded into the environment, we must be careful that the ideology the technology embodies is not blindly incorporated into the environment as well. Space is not merely a neutral background for human activity, culture is built into those forms [5]. As disciplines, engineering and computer science make implicit assumptions about the world that conflict with traditional modes of cultural production [8]. This paper introduces tactile/space*, an experimental location aware garment that seeks to address lacuna in the design process of computationally enhanced cultural artifacts and the spaces they inhabit.*

1. Experience

"Culturally and socially, space is never simply an inert background of our material existence" [5]. *tactile/space* is a large scale tactile installation that examines the influence of wearable computational media on the ways in which people perceive, understand and use space. Participants put on a vest outfitted with a wearable computer/global positioning satellite receiver and are instructed to wander about a space as they like. While exploring the area, participants encounter changes of vibro-tactile texture dependent on their location. Some phenomena seem to move, suggesting an unseen presence or palpable waves. Others are more static and may be perceived as gradients or landmarks. The everyday experience of space is augmented with an invisible tactile landscape, where one navigates less by vision and more by feel. New boundaries and thresholds are encountered as tactile textures move across the body shaping the experience of bodily orientation and navigation. *tactile/space* challenges the participant to think about the role other senses play in everyday spatial experience in a culture dominated disembodied technologies.

2. Background

The idea for *tactile/space* originates out of concerns about the ways in which new technologies shape our perceptions of space. The computer arises from Western scientific ideology which is built upon the assumption that the mind is separated from the body. The influence of this assumption is present at all levels of the technology, from the architectural level in the hardware/software split to the reduced set of body senses/movements engaged by its interface. The majority of research in human computer interaction has focused on the visual or aural modalities; very little has been focused on the cutaneous senses [10]. One possible explanation is that culturally, the senses of the flesh are regarded as the most lowly, the most animal. They are the farthest from the pure reason upon which computing is founded: the mind/body split. The mind/body split has been debunked time and again from all angles: philosophy, linguistics, biology, neurology, cognitive science, art, and even computer science [3]. However, embracing what embodiment means is difficult in a language that reifies the ideology. We talk of "my hand" or "my head" as if they are objects that we own, not inseparable parts of our being. To describe what is meant by embodiment, we must use neologisms such as the embodied mind, being-in-the-world, the life-world, etc. An embodied perspective that is not clouded by traces of duality is difficult, at best, in contemporary Western culture. In computer science, where the ideology is ingrained in the technology itself, it presents an even stiffer challenge.

3. Embodiment

It is important to briefly clarify what is meant here by 'embodiment.' It is not simply the fact that humans have bodies. It is not a notion of an embodied mind that still clings to sacredness of consciousness or abstract rational thought, and thus still leaves them outside the realm of the physical body. It has been shown that even a change in posture, while maintaining identical sensorial stimulation, alters neuronal responses [15]. Embodiment is the relationship between meaning and action, and “the process of grasping a meaning is performed by the body [6].” Thus, the methods for making sense of action and the methods for engaging in it are the same methods [3]. The focus is on the world as it is lived in and experienced, not the world as it appears when we stop acting and start thinking about it [4]. Understanding the world as the experience of an embodied mind creates a strong connection between person and their environment. From a biological perspective, Maturana and Varela have described this relationship as structural coupling. It is the process of an organism interacting with their environment where “the structure of the living system and the structure of the medium change together congruently as a matter of course.” [14] Therefore, the mind is inseparable from the body, and the body inseparable from the environment.

4. Technology

There are a number of tactile receptors that could be stimulated for use in wearable computing applications: thermal, pressure, electrocutaneous, humidity, air movement, vibrotactile, etc. The current state of the art points to vibrotactile as the modality for ubiquitous computing applications. Vibrotactile actuators are neither intrusive nor painful (problems that are possible with electrocutaneous actuators). They can be felt through clothing, are inexpensive, and have relatively low mechanical and power requirements.

A perceptual illusion, sensory saltation, holds promise for use in vibrotactile displays. Sensory saltation occurs across the senses resulting in the perception of apparent motion. Tactile sensory saltation was discovered in the early 1970s by Dr. Frank Geldard at the Princeton Cutaneous Communication Lab. In a typical setup for eliciting tactile sensory saltation, three mechanical stimulators are place equidistant from each other on the forearm. The stimulator closest to the wrist delivers three short pulses, followed by three more at the middle stimulator, and finally three more at the last stimulator. Instead of perceiving three pulses at each of the stimulator sites, the observer perceives that all of the pulses are distributed with approximately uniform spacing from the site of the first stimulator to the site of the third. The sensation is described as if a tiny rabbit was hopping up the arm from wrist to elbow, and is sometimes called the "rabbit" effect or the "cutaneous rabbit." An important feature of this illusion is that it is able to simulate higher spatial resolution than the actual number of stimulators, yet create the impression of a dense stimulator array, thus potentially reducing the overall weight and power consumption needs of a wearable device.

Experimentation with spatio-temporal patterns in vibrotactile displays has uncovered a group of concepts which can be perceived through the tactile sense. For example, a single point can be perceived as a direction. [13] Tactors arranged spatially on the body can create a relationship akin to vision where the ego center is perceived as one point and the stimulus at another, thus creating direction. Taking advantage of sensory saltation, lines can be perceived, as can their length, straightness, spatial distribution and smoothness. [2] There is also some more recent research [11] that suggests planes and three dimension forms can be perceived. Tactors on the body can create a 360 degree “field of touch” where lines and forms can be perceived not just on the surface of the body, but through the body. Finally, the tactile systems that have been discussed are very easy for users to learn, and require practically no training.

5. Conclusion

As computing produces increasingly influential cultural artifacts and becomes ingrained in culture, it can no longer disregard the ideologies that have been responsible for cultural production for the thousands of years of humanity's existence. The arts have traditionally emphasized body knowledge in the making of things. Embodied thought is important in the design of physical spaces as well as in the making of objects. As computing begins to shape physical spaces and their uses, it will be important to look to the arts and architecture to see where the underlying assumptions inherent in the discipline conflict with embodied experience. The goal of *tactile/space* is to render conscious the interaction between the space of computation and the space of the bodily senses.

Wearable vibro-tactile devices, especially those designed for the torso or entire body, are a rich medium for embodied communication with computational devices. Tactile displays shift attention, either focal or peripheral, to the body, which is an area that computing often ignores. The predominately visual nature of the computer interface mirrors the dominance of vision in Western culture. The dominance of vision can be traced back to the same ideological roots of computing: the separation of mind and body and thought from experience. Tactile interfaces may be an avenue for computing to embrace the body and work towards a rebalancing of the senses in design. Wearable vibro-tactile displays have been shown to "enhance environmental awareness" [2] and help maintain spatial awareness in unusual environments [11]. When this becomes an area of focus and concern for computing, the design of computational cultural artifacts that foster an embodied experience is possible.

6. References

[1] Cholewiak, Roger W., Collins, Amy A., and Brill, J. Christopher, "SpatialFactors in Vibrotactile Pattern Perception" Eurohaptics, 2001.

[2] Cholewiak, Roger W. and Collins, Amy A., "The generation of vibrotactilepatterns on a linear array: Influences of body site, time, and presentation mode." *Perception and Psychophysics,* 2000. pp. 1220-1235.

[3] Dourish, Paul, *Where the Action Is: The Foundations of Embodied Interaction,* MIT Press, Boston, 2001.

[4] Loren, L. and Dietrich, E., "Merleau-Ponty, Embodied Cognition and the Problem of Intentionality", *Cybernetics and Systems 28*, 1997. pp. 345-358.

[5] Hillier, Bill. *Space is the Machine A Configurational Theory of Architecture*. Cambridge University Press, Cambridge, 1996.

[6] Merleau-Ponty, Maurice. *The Phenomenology of Perception*, Routledge, London, 1962.

[7] Ong, Walter. *Orality and Literacy*, Routledge, London, 1982.

[8] Penny, Simon. "The Virtualisation of Art Practice: Body Knowledge and the Engineering World View." *CAA Art Journal* Fall 1997. pp. 30-38.

[9] Svanaes, Dag. "Kinaesthetic Thinking: The Tacit Dimension of Interaction Design" *Conmputers in Human Behavior*. *Vol 13, No. 4*, 1997. pp. 443-463.

[10] Tan, Hong Z. and Pentland, Alex., "Tactual Displays for Sensory Substitutionand Wearable Computers," *Fundamentals of Wearable Computers and Augmented Reality* (Woodrow Barfield, Thomas Caudell eds.) Lawrence Erlbaum Associates, Publishers, New Jersey, 2001.

[11] TNO Human Factors. 2003 Reports of department of Skilled Behavior, http://www.tm.tno.nl/

[12] US Navy Tactical Situation Awareness System http://www.namrl.navy.mil/accel/tsas/

[13] van Erp, Jan B. F., "Tactile Navigation Display," LNCS 2058, 2001. pp. 165-173.

[14] van Erp, Jan B.F. and van Veen, Hendrick A.H.C., "A Multi-purpose TactileVest for Astronauts in the International Space Station," *EuroHaptics* 2003.

[15] Varela, Francisco. "Autopoiesis and a Biology of Intentionality," ECAL 1997.

Using Context-Awareness to Improve Data Caching and Prefetching for Mobile Users

Ingrid Burbey
Bradley Department of Electrical and Computer Engineering
Virginia Tech
Blacksburg, Virginia 24061
iburbey@vt.edu

1. Introduction

A wearable computer with a strong wireless signal and an abundance of bandwidth can easily access information available on the Internet. However, problems arise when a user moves with his wearable computer, passing through areas where wireless resources are not as plentiful. The goal of this project is to explore two methods to allow a variety of wearable computers to function effectively while passing through changing environments and without complete dependence on external infrastructure. These methods include using location-awareness to prefetch data from a remote database, and exchanging data with other users when in areas without wireless coverage. The quantity and type of data requested from the Internet database is influenced by many factors: the limitations of the wearable computer, knowledge about the user's current and possible future locations, knowledge about upcoming areas of wireless coverage and the presence of other users.

The remainder of this paper is organized as follows. Section 2 discusses the issues involved with using context-awareness in wearable computing applications. Section 3 states the problem. Section 4 describes a scenario of a user walking across campus. Section 5 discusses some of the factors that affect data prefetching. Exchanging data with other users is covered in section 6, and section 7 lists the tasks involved in simulating the problem.

2. Background

This section discusses some of the issues in context-awareness that affect mobile wearable applications. These issues include the need for infrastructure to determine location and to provide wireless access, limitations of the wearable computer, the need to adapt to changes in the environment, and the difficulty of predicting a user's movements.

Location determination is critical to providing context-awareness to a mobile computing application. Many location systems require infrastructure to be added to the environment, which can not be assumed to be present in all the environments a mobile user will encounter. GPS is an excellent method of location determination, but only works outdoors, away from buildings and other tall structures. Indoor location determination usually requires built-in infrastructure, in the form of cameras, beacons, or tags [1]. A mobile application that seeks to function in diverse environments will need various ways to determine location and will need to know when to switch between methods.

Another infrastructure requirement is that of wireless support. A mobile application must be able to communicate with other devices and sources of information. Wireless access is not available in all environments. Ideally, the applications running on the wearable computer can either compensate for the lack of resources or degrade gracefully. Satyanarayanan calls this concept of hiding the variations in environments "masking uneven conditioning."[2]

Any wearable application needs to consider the capabilities of the wearable computer itself. Energy consumption, secondary storage, and processing power all need to be minimized. The mobile application should be implemented as a thin client, using as few resources on the wearable computer as possible. As the mobile user moves between resource-rich and resource-poor environments, the client needs to adapt to negotiate its way through a new space. One issue that arises is that of adapting the application to mask the uneven conditioning of the environment without overloading the device itself.

Another concept that affects accessing data is that of 'localized scalability.' [2] The addition of multiple wireless users interacting with wireless resources and each other can lead to an unmanageable number of interactions, devouring bandwidth and energy. The data prefetched should be limited to the device's local area to conserve bandwidth, memory and the power of the wearable computer.

Another goal of context-awareness is predicting the users' intentions, in order to adapt the application to the situation.[2] In this problem space, this involves speculation of a user's future movements.

3. Problem Statement

This project seeks to explore two methods of alleviating uneven conditioning; first, by prefetching data before entering an area without wireless coverage, and secondly, by exchanging information with other wireless users when in areas without wireless coverage.

Starner contends that one "way to alleviate coverage issues is to employ aggressive caching. [3]" Obviously, not everything can be cached. The application needs to be intelligent enough to predict what data might be needed and how much data can be downloaded and cached. This involves predicting the user's likely path, at least in the immediate future.

Nearby mobile users can serve as a resource in areas without wireless coverage. Passing users can exchange information about the area that they have just traversed, information that is needed by the other.

4. An Example

Imagine a visitor to the Virginia Tech campus. The visitor has a wearable computer with wireless communication capability and the ability to determine its location. The user can access a wired (Internet) database that contains both static and dynamic information. The static information includes maps and zones of wireless coverage. The dynamic information could include current events on campus or construction zones. As the user moves across campus, the application performs two tasks: it prefetches data about the paths before the user and it stores data about the user's recent surroundings (figure 1). If the application determines that a user is about to enter an area without wireless coverage, it will prefetch additional data.

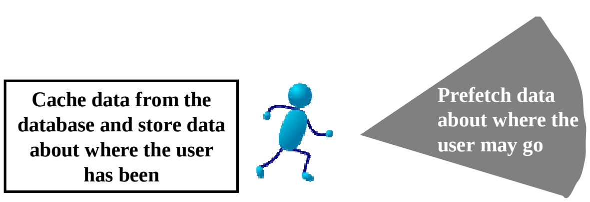

Figure 1. A single user prefetching data

Eventually our user enters an area of poor coverage. The primary data source is unavailable, so the application can utilize another resource: the data cached or collected by another user. Two users passing each have information useful to the other. If they can form an ad-hoc connection, they can trade useful information. This information may include forthcoming traffic congestion, areas with the best coverage or other items of interest (figure 2).

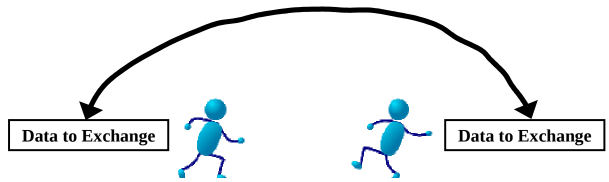

Figure 2. Users exchanging data with an ad-hoc connection

5. Constraining the Database Queries

When in an area with wireless coverage, the application connects to the database and requests the data necessary to describe the spatial area in front of the user. If the application notices an upcoming area of poor or no coverage, it begins prefetching data. The application needs to determine how much data and what data to request.

The prefetch area in front of a mobile user can be visualized as a cone of possible paths that the user could take. This area has two attributes: size and density. Each of these attributes represents different constraints on the queries made to the database (figure 3). The size correlates to the spatial area of interest in the database; and the density corresponds to the amount of data available. An example of a low-density query would be "Return only critical information."

The size of the prefetch area is also influenced by the user's state. The speed of the user changes its length. A user driving an automobile needs information several kilometers ahead, whereas a walking user only needs information for several hundred meters ahead. If the number of possible paths is small, there may be less information to request. For example, a vehicle traveling on a freeway is limited in its choice of paths; it either stays on the current path, or takes a predefined exit. A pedestrian can change direction at any time and short of walking into a wall, can go any conceivable direction, but will tend to follow a sidewalk or a hallway.

The size of the prefetch area is also determined by the environment and the device. If the application knows that the user is approaching an area with no wireless capability, the size of the cone increases to prefetch enough data to get the user through the quiet zone.

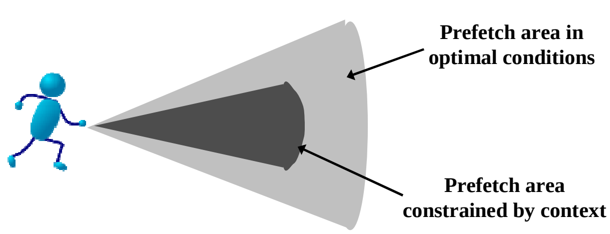

Figure 3. Reducing the prefetch area

The other attribute of the cone is its density. The density represents the number of data points that the application will request and is related to the relative importance of those facts. A resource-rich device can request all the information available about the area before the user. If resources are limited, the application may ask for a subset of the data. In the most limited case, the application would only display critical data such as emergency alerts, alarms or recent changes.

The most effective prefetch strategy is one that can predict the user's desires and actions. There are several approaches that may predict the user's path. The application could assume that the user is going to continue on the current heading (for example, heading North) [4] or, if the application has route data, it could assume the user is going to stay on the same route (a road, for example, or a sidewalk). The application could keep historical data and glean the user's usual routes. In addition, the application could use other resources, like the user's online datebook, to predict the user's destination.

6. Sharing Data with Passing Users

In areas of poor coverage, the data collected and cached by other mobile users is a valuable resource. This project explores the use of a peer-to-peer ad-hoc network to allow users to share their recent travels. This information may include the path, the time needed to traverse the path, areas of upcoming wireless coverage, alarms and any other data either cached or collected. The density of upcoming wearable computers could indicate traffic congestion.

The duration of the ad-hoc connection is dependent on the relative velocities and directions of the mobile users. The capabilities of the wearable computers and the quantity and quality of the information to be shared also affect the parameters of the exchange.

Security, trust, and privacy issues also have a part to play in the interaction. Users need to be encouraged to share data and assured that their personal information will not be divulged. The solution may be to exchange only data that are public knowledge, such as areas of wireless coverage, without including sensitive data, such as the identities or locations of the users. A security policy could be implemented [5].

7. Simulation

Initially, a simulation will be designed to explore the parameters of the problem. Part I of the simulation, which models a mobile user prefetching and caching data from an Internet database, will include a server with a database containing static and dynamic data of varying importance. The mobile devices will be modeled as clients with simulated location information. This portion of the simulation involves several issues:

- Map data. Should the database consist of route data or a set of PDF images? What is the best way to store route information on a thin client?
- Path prediction. Initially, we can simulate a known path in and out of areas of wireless coverage to implement the prefetching algorithm. Eventually, path prediction algorithms will need to be evaluated and implemented [6].
- Formation of queries to the database. What data and how much data should the application request? How should these queries be constrained by context?
- Presentation of the data. Given the limitations of the device, what is the best way to present the data?

The second part of the simulation models the interaction between passing users in areas without wireless support. These tasks include

- Identification of other users.
- Negotiation of the ad-hoc connection.
- Exchange of data and decision about what to exchange.

The simulation can answer questions about the use of context-awareness to efficiently prefetch and cache data, either from a database located on a server or a cache stored on another user's wearable computer.

This project seeks to use context-awareness to aid in travel through areas of poor coverage. An effective algorithm to prefetch and cache data and to share data with other users will allow wearable computing applications to move beyond areas with built-in infrastructure.

8. References

[1] J. Hightower, G. Borriello, "Location Systems for Ubiquitous Computing", *IEEE Computer*, Aug 2001, pp 57-66.

[2] M. Satyanarayanan, "Pervasive Computing: Visions and Challenges", *IEEE Personal Communications*, Aug 2001, pp 10-17.

[3] T. Starner, "The Challenges of Wearable Computing: Part II", *IEEE Micro*, July-August 2001, pp. 54-67.

[4] H. Kung, H. Ku, "A Real-time Mobile Multimedia Communication System for the Prevention and Alert of Debris-Flow Disaster", *IEEE Vehicular Technology Conference*, October 2003, pp 3400 – 3404.

[5] M. Massimi, U. Wolz, "Peer-to-peer policy management system for wearable mobile devices", *ISWC*, Oct 2003, pp 245-247.

[6] M. Kyriakakos, N. Frangiadakis, L. Merakos, S. Hadjiefthymiades, "Enhanced path prediction for network resource management in wireless LANs", *IEEE Wireless Communications*, Dec 2003, pp. 62-69.

Towards Socially-Intelligent Wearable Networks

Anmol Madan, Ron Caneel, Alex "Sandy" Pentland
MIT Media Laboratory
20 Ames Street, Cambridge 02139
{anmol,rcaneel,sandy} @ media.mit.edu

.

Abstract

We propose a wearable system that uses machine perception to quantify a user's social context and propagate this information to others in the user's social network. The social context is evaluated for the user's instantaneous, face-to-face interactions by evaluating proximity, collective speech features, head-movements, and galvanic skin responses. This information is then propagated to others within the user's social or work group who have pre-approved permission to `patch in' to interesting conversations. We believe that propagation of social context will allow distant users to become better integrated into ongoing projects or discussions, and thus improve distance-separated social interaction, teamwork, and social networking

1. Idea & Motivation

People usually operate in groups, and their needs and goals are often defined by their group context. Most current wearable systems, however, only consider individual preferences and physical context variables such as location. We believe that it is essential to consider the people in the user's face-to-face proximity, as well as the user's social or work network, in order to meaningfully model their needs and expectations.

Quantification of social context may have its largest impact on distance interactions, where social context can be propagated over physical distance in order to better integrate distant participants. One can imagine harnessing the computing and sensing power of today's nearly ubiquitous wearable devices to provide a `social intelligence' that improves the functioning of distributed work groups.

There has been past work to build mathematical understanding group behavior using speech feature analysis, physiology, activity classification and context mapping. Our feature space currently consists of speech features, galvanic skin responses (GSR) and head-movements, with which we attempt to capture quantitative elements of non-verbal communication like body language, prosody, empathy, and mirroring behavior. We add to this information about the user's location and their proximity with others.

Choudhury [4,15] was able to show how turn-taking in face-to-face conversation (modeled using a generalized coupled Hidden Markov Model) could be a measure of influence, which was shown to have an extremely high correlation with one measure of the social novelty (and thus presumably the interestingness) of the information being presented. Pentland et al. [14, 16] have attempted to generalize speech features as measures of activity, engagement, stress and mirroring behavior. Our early work with the influence model also indicates how it could be representative of direction of information flow in face-to-face conversation.

GSR has been used as a measurable parameter of a person's internal "state" [9,17], associated with fear, anger, startle response, orienting response and sexual feelings. Studies in psychophysiology have used galvanic skin responses (or electro-dermal activity) as a physiological index for multiple individuals exposed to the same cognitive or emotional events [12]. The role of head nodding and head movement has been acknowledged in conversational feedback for the speaker [11], mirroring behavior and the *chameleon effect* (where people mimic body movements of their conversational partners and that this is reflected more in the behavior of a seemingly more empathetic person) and in *isopraxism* [3]. In [13] we have shown how a combination of speech features and similar GSR epochs and head-movements for a group, could be used to build a *group interest index.*

By capturing all these features, it is possible to incorporate *group social awareness* in wearable devices (based on sensing and real-time machine learning). The propagation of just group location and activity awareness within peers has been shown to enhance distance separated collaboration [10] and has also been used to infer shared interests [19]. We believe that by adding mathematical understanding of behavior based on our feature analysis, we can enhance distance-separated teamwork, collaboration, decision-making and the sharing of social experiences. *Social intelligence*, captured using wearable devices and propagated over distance will give users some fashion of information they are accustomed to acquiring from body language, speaking style and sub-conscious understanding of behavior, if they were in the same physical-temporal space.

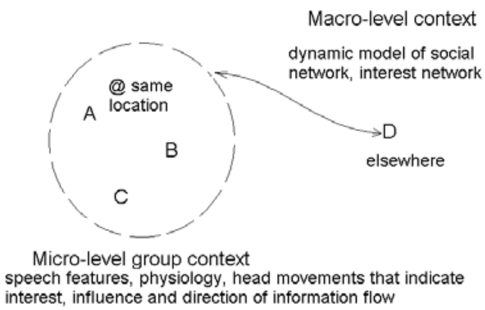

Fig 1: Propagating quantified social context over distance

Some scenarios to illustrate our broad vision of quantitative social intelligence in cell phones and PDAs are,

- Knowing the flow of information and influence for participants in a teleconference negotiation
- Sharing social experiences and common interests with separated friends and family – "Its Saturday night, who's on my interest network?"
- Next-generation Instant Messaging software where users do not manually type emoticons, but use wireless physiology sensors and speech-analyzing cell phones
- Mobile social software (like Friendster or Orkut) that captures state information – interest, influence, engagement and direction of information flow

2. Proposed Implementation & Early Work

2.1 Hardware and Software Platform

Our hardware and software platform is derived from the Mithril system [6]. The system core is the Zaurus SL6000 Linux PDA, capable of real-time data analysis, peer-to-peer wireless networking, bluetooth wireless sensing, full-duplex audio, local data storage, graphical interaction, and a keyboard/touch-screen input. A sensor hub connects it to an accelerometer and GSR leads, and also supports measurement of heart rate, EKG and other physiological signals. We use the Enchantment whiteboard and signaling software as a lightweight means of routing information transparently across processes or distributed devices, and propose bluetooth for proximity detection [8].

We have strongly felt the need for wireless sensing and an omnipresent form factor, and hence are integrating a wearable system with bluetooth accelerometers [5], bluetooth GSR sensors [18] and Motorola A760 Linux cell phones. We think that a robust, trendy system that sufficiently captures state information can be made using a Zaurus /cell phone and bluetooth accelerometers and GSR sensors.

Fig 2: The present system featuring a Zaurus PDA, Hoarder sensor board, bluetooth accelerometer (on hat), GSR leads and a microphone

2.2 Feature Analysis

Our proposed system calculates speech features (devoid of any linguistic content) of two different types - individual features and the group features. Individual features include fraction of speaking time, standard deviation of energy, standard deviation of formant frequency, standard deviation of spectral entropy and the voicing rate, and we use them as measures of speaker prosody and emphasis. We use a multi-level HMM structure [1] to classify the voiced/non-voiced sections and hence the speaking /not-speaking regions of the audio stream.

Group features represent the dynamics of interaction between individuals, and include the influence parameters and back-and-forth exchanges. The influence model is a Coupled Hidden Markov Models (CHMMs) designed to describe interactions between two people, where the interaction parameters are limited to the inner products of the individual Markov chains [2,4]. This model allows a simple parameterization in terms of the *influence* each person has on the other. We also label short interjections of a time scale less than 1 second as back-and-forth exchanges (typically single words like 'okay', 'aha', 'yup' or sounds of approval), and consider them as mirroring behavior [14], where one person unconsciously mimics the other's prosodic pattern.

Head-nodding is detected using a Gaussian mixture model on the frequency domain representation of head movements[7].

Galvanic skin responses are detected using slope detection.

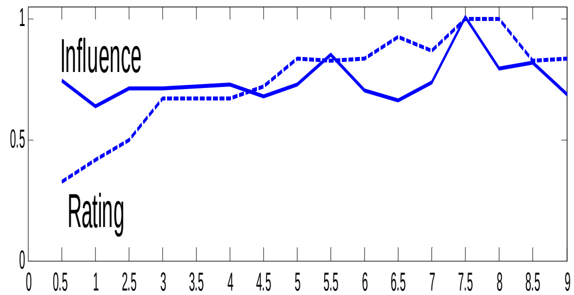

Fig 3: Graph of influence parameters and real-time group-interest feedback on a PDA

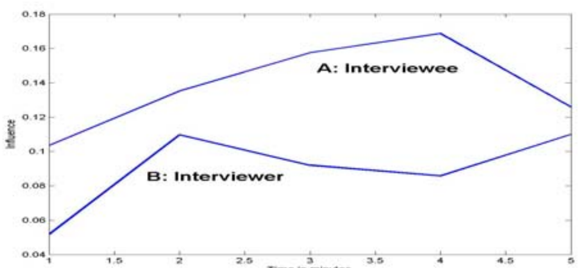

Fig 4: Graph of influence parameters indicating flow of information in an interview situation

2.3 Propagation of Social Context

Our goal is to develop lightweight, unobtrusive wearable systems that can identify face-to-face groups, capture collective social information, extract a meaningful group index, and transmit the group context to remote group members.

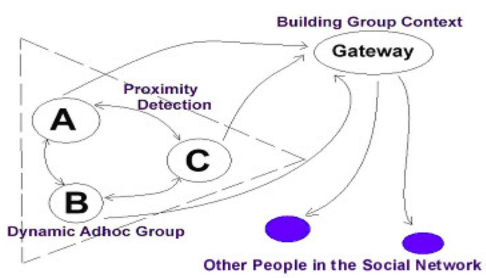

Fig 5: Group formation and information propagation

The gateway in Fig. 5 uses real-time machine learning methods to identify relevant group context changes, such as changes in influence parameters (indicating a change in information flow), or instances of group head-nodding (indicating changes in group interest) [13] or instances of group GSR concordance (indicating interesting group events).

Depending on the identities of the people taking face-to-face, the directionality of the information flow, and the interest levels, the gateway can then notify pre-approved distant users that they might want to `patch in' to the conversation. Upon receiving such a notification distance-separated group members can either subscribe to this information over the enchantment whiteboard and begin to receive the raw audio signal plus annotations of the social context, or they can choose to be notified by the system only in case of interesting events, like simultaneous GSR spikes in a number of people, or they can store the audio signal with social annotations for later review.

References

[1] S. Basu. "Conversation Scene Analysis", in Dept. of Electrical Engineering and Computer science. Doctoral, MIT, 2002

[2] M. Brand, N. Oliver & A. Pentland, "Coupled hidden Markov models for complex action," IEEE Conference on Computer Vision and Pattern Recognition (CVPR97), 1996.

[3] T.L. Chartrand, & J.A. Bargh, "The chameleon effect: The preception-behavior link and social interaction." Journal of Personality and Social Psychology, 76, 893-910, 1999.

[4] T. Choudhury, "Sensing and Modeling Human Networks." PhD thesis, Dept of MAS, MIT, 2003.

[5] Contexual Computing Group, Georgia Tech, GA Bluetooth accelerometer design (http://www.cc.gatech.edu/ccg/resources/btacc)

[6] R.W. DeVaul, M. Sung, J. Gips, S. Pentland, "MIThril 2003: Applications and Architecture", Proc. ISWC '03, White Plains, N.Y., October 2003

[7] R. W. DeVaul and S. Pentland, "The MIThril Real-Time Context Engine and Activity Classification", Technical Report, MIT Media Lab, 2003

[8] Eagle, Pentland, "Social Serendipity: Proximity Sensing and Cueing", Media Lab Technical Report 580, May 2004

[9] G. D. Fuller, "GSR History, & Physiology" taken from Biofeedbacks Methods and Procedures in clinical practice, 1977.

[10] Holmquist, Falk, "Supporting Group Collaboratin with InterPersonal devices"

[11] A. Kapoor & R. Picard, "A real-time head and nod shake detector". In IEEE PUI, Orlando, FL, 2001.

[12] C. Marci, "Psychotherapy Correlates of Empathy, An investigation of the patient-clinical dyad", 2002.

[13] A. Madan, R. Caneel, S. Pentland, "GroupMedia: Distributed Multimodal Interfaces", ICMI 2004

[14] Pentland A, Curhan J, Eagle N, Martin M., "Money and Influence", Media Lab Technical Note 577, May 2004 (http://hd.media.mit.edu)

[15] Pentland A, "Social Dynamics:The Voice of Power and Influence", Media Lab Technical Note 579, May 2004 (http://hd.media.mit.edu)

[16] Pentland, A. (2004) Social Dynamics: Signals and Behavior, ICDL, San Diego, CA Oct 20-23

[17] Picard, R.W., "Affective Computing", MIT Press,Cambridge

[18] Reynolds C. Strauss M. , Picard R.W., Bluetooth GSR sensor design, (http://arsenal.media.mit.edu/notebook/)

[19] Terry, Mynatt, Ryall, Leigh, " Social Net: Using Patterns of Physical Proximity over Time to Infer Shared Interests", 2002

Interaction Authoring for Adaptive User Interfaces in Ubiquitous Mixed Realities

Irma Lindt
Fraunhofer Institute for Applied Information Technology
irma.lindt@fit.fraunhofer.de

Abstract

This paper describes a new approach to adaptive user interfaces for ubiquitous Mixed Realities (MR). In the proposed work, user interfaces are authored in an abstract, device- and modality-independent manner to comprise a broad range of interaction devices and interaction techniques, respectively. Typical MR interaction techniques such as Tangible User Interfaces and 3D widgets, as well as interaction techniques from the mobile and desktop computing domain are supported. At run-time, user interfaces are generated for specific interaction devices. With the proposed interaction authoring approach, MR applications are decoupled from specific hard- and software configurations and may run with potentially any interaction devices.

1. Introduction

MR has the means and the potential to seamlessly integrate computing technology into our everyday world in the form of an unobtrusive user interface. First prototypes such as MARS [6] or Smart Roads [5] demonstrate Mobile MR applications that enrich our everyday world with additional information and smart services.

Extending Mobile MR to be accessible with a variety of possible interaction devices will lead to ubiquitous MR. This term follows the vision of Mark Weiser [10] about ubiquitous computing and expresses the idea of Mixed Reality technology becoming so easy to use, vital and widespread that it is vanishing into the background of a user's attention. Ubiquitous MR means a tight connection between the physical and the virtual world. In ubiquitous MR multiple sensors prevail throughout our physical environment to collect context information that allows for smarter and more supportive MR applications. Ubiquitous MR does not rely on a single interaction device, but works with a large variety of heterogeneous devices, interaction techniques and interaction modalities.

This paper proposes an approach that allows the creation of MR user interfaces independent of hard- and software configurations. The most advantageous interaction techniques and interaction modalities are chosen based on the current context, such as the available interaction devices and the current situation and tasks of a user.

2. Motivation and objectives

One objective of the proposed work is the *simplification of the development of user interfaces for ubiquitous MR*. Not only software experts but also domain professionals should be able to create the interactive part of their MR applications. This will allow a broader group of people to access and use MR technology according to their requirements. As user interfaces are a major part in the creation process of ubiquitous MR applications, the approach will concentrate on supporting various interaction techniques and interaction modalities for different interaction devices.

Another objective is the *support of a broad range of user interfaces*. Well-designed user interfaces are task and user-oriented and the employed interaction techniques are carefully chosen. Tangible User Interfaces, for example, as they are frequently used for MR applications are well suited for geometric transformation tasks, whereas 3D Widgets are better suited for representing complex functionality. Instead of a user interface concept that tries to comprise all user and application requirements, various user interfaces need to be supported.

The third objective of the proposed work is the *development of context-adaptive user interfaces*. As the idea is to have MR applications that are ubiquitously available, user interfaces should adapt to given interaction devices, techniques and modalities.

This will allow for a seamless integration of MR applications into our everyday world in the form of unobtrusive user interfaces.

3. Related Work

Creating MR applications is a rather complex process, as various components have to be developed and controlled. Some research projects try to simplify the application development with tools that allow for defining the user interface and simple application logic. Other projects focus on the creation of flexible and adaptive user interfaces. This section summarizes interesting projects in the field of interaction authoring and adaptive user interfaces.

The AMIRE project [11] is dedicated towards the efficient creation of MR applications. The underlying framework is component-oriented and designed for creating MR applications and authoring tools. The authoring process includes the selection of MR components, the adaptation of the selected components, the specification of the connections between the components and the calibration within the MR environment. Typically, the result of an authoring process is stored in an XML based application description that is interpreted at runtime.

APRIL [3] is an XML compliant high-level description language for authoring Augmented Reality presentations. APRIL describes media objects, their behaviors and possible user interactions with a story-centered approach. The description of the presentation flow ("the story") is based on UML state charts. States within the UML diagram represent behaviors, whereas transitions represent interactions.

MONA [8] supports multimodal user interfaces for low-end WAP-phones, Symbian-based smartphones, and PocketPC PDAs with a single implementation. MONA is based on the User Interface Markup Language (UIML) [1], a meta language allowing for defining device-independent user interfaces. Similar to the MONA project Ali et al. [2] define a generic vocabulary that can be rendered for multiple platforms. Although a uniform language for describing multi-platform applications seems to be obvious, only a few approaches do exist. There is no approach that comprises typical MR interaction techniques and interaction techniques from the mobile and desktop computing domain.

The growing interest in adaptive user interfaces is driven by the rapid emergence of an increasing variety of computing devices in the past few years, and the wish to access the same data and services on all devices. For MR, the adaptation to various aspects of the physical and synthetic environment has been realized in the form of information filtering, adaptive label placement and registration error correction [7] [4]. Existing approaches to adaptive MR user interfaces mainly focus on information presentation issues, though. Other possible adaptation aspects such as interaction techniques and interaction modalities have rarely been explored in the field of MR.

4. Proposed work

This section summarizes ideas and considerations regarding interaction authoring for adaptive user interfaces in ubiquitous MR which will form the basis of the proposed work.

4.1 Abstract user interface representation

An application running on various software and hardware platforms typically involves extra efforts and costs. Among existing approaches to the problem of user interface portability, such as binary emulations or ported APIs, user interface description languages (UIDLs) are quite promising.

UIDLs aim at user interfaces independent from underling devices, platforms and modalities. One objective of UIDLs is to enable automated generation of user interface code from a rather abstract user interface description. Typically, UIDLs are defined with the XML meta-language and allow for describing user interfaces for desktop and mobile PCs, PDAs and cellular phones [9]. However, vocabularies based on existing UIDLs do not offer explicit support for 3D user interfaces as they are used in MR applications. A thorough literature review did not reveal considerable efforts in the area of abstract user interface representations for 3D user interfaces.

In the course of the proposed work, existing UIDLs will be thoroughly investigated concerning their suitability for describing user interfaces for ubiquitous MR applications. A vocabulary based on a selected UIDL will be developed that allows for a device and modality-independent description of user interfaces.

4.2 Interaction authoring

Application authoring simplifies the application creation process by abstracting from the underlying programming language. Interaction authoring in the context of the proposed work describes the modifiable aspects of virtual artifacts and possible interactions in an abstract manner.

The envisaged interaction authoring approach will focus on the following interactions:

- Tangible interaction. Virtual artifacts can be connected to one or several physical objects to allow for creating tangible user interfaces. The quality of the linkage between the virtual and the physical object can be specified to allow for offsets, scaling and restricted movements.
- Widgets-based interaction. Widgets are geometry that can represent arbitrary functionality. A set of custom widgets will be developed that can be employed and connected to application-specific functionality.
- Multimodal interaction. Alternative modalities can be specified for interactions. As user interfaces will be described in an abstract manner, concrete multimodal renderings such as visual, speech and/or haptic user interactions will depend on the available software and hardware platform.

4.3 Adaptive user interfaces

A small set of interaction devices will be described with their characteristics, such as the input and output modalities they support. For each supported device a reasonable mapping of the abstract user interface description to one or several concrete user interfaces will be developed.

The actual user interface for a given application depends on the availability of interaction devices (input and output devices). A major contribution of the proposed work is a method that allows for selecting advantageous interaction techniques and modalities based on the current context. The context includes the set of possible interaction devices and also the set of possible interaction techniques and modalities. Based on this and possibly also on the preferences and the current situation of a user the most suitable user interface will be generated. As automatically generated user interfaces are often of poor quality, a research challenge is the generation of user interfaces that follow accepted user interface guidelines.

5. Conclusion and outlook

This paper presents a new approach to interaction authoring for adaptive user interfaces in ubiquitous MR and proposes using an abstract user interface vocabulary based on existing UIDLs. Relevant MR interaction techniques are identified and some ideas regarding adaptive user interfaces are described. The proposed work can provide valuable inputs for future scientific efforts in the field of interaction authoring and adaptive user interfaces for MR.

6. Acknowledgements

The author would like to thank her colleagues at Fraunhofer FIT, and especially Wolfgang Broll and Wolfgang Prinz for valuable discussions.

7. References

[1] Abrams, M., “Device-Independent Authoring with UIML”, in Proceedings of the W3C Workshop on Web Device Independent Authoring, 1999.

[2] Ali, M. F., M. A. Pérez-Quiñones, M. Abrams, E. Shell, “Building Multi-Platform User Interfaces with UIML”, in “Multiple User Interfaces” by A. Seffah and H. Javahery (eds.), John Wiley & Sons, 2003.

[3] APRIL Website, http://www.studierstube.org/april, 2004.

[4] Azuma, R., C. Furmanski. “Evaluating Label Placement for Augmented Reality View Management”, Second International Symposium on Mixed and Augmented Reality, ISMAR 2003, Tokyo, October 2003.

[5] Christian, V., A. Ferscha, W. Narzt et al., “Smart Roads in the Pervasive Computing Landscape”, in the Video Proceedings of the Second International Conference on Pervasive Computing, Vienna, April 2004.

[6] Höllerer, T., S. Feiner, et al., “Exploring MARS: Developing Indoor and Outdoor User Interfaces to a Mobile Augmented Reality System”, Computers and Graphics, 23(6), Elsevier Publishers, 1999.

[7] Julier, S., M. Livingston, J. Swan, Y. Baillot, D. Brown. “Adaptive User Interfaces in Augmented Reality”, in Proceedings of the STARS Workshop at ISMAR 2003, Tokyo, October 2003.

[8] Simon, R., M. Jank, F. Wegscheider, “A Generic UIML Vocabulary for Device- and Modality Independent User Interfaces”, in Proceedings of the 13th International World Wide Web Conference, May 2004, NY, USA.

[9] Trewin, S., G. Zimmermann, G. Vanderheiden, “Abstract User Interface Representations: How well do they Support Univeral Access?”, in Proceedings of CHI 2003.

[10] Weiser, M., “The Computer for the 21st Century”, Scientific American 265(3): 66-75, 1991.

[11] Zauner, J., M Haller, A. Brandl, W. Hartmann, “Authoring of a Mixed Reality Assembly Instructor for Hierarchical Structures”, Second International Symposium on Mixed and Augmented Reality, Tokyo, October 2003.

Usability of Electronic Maintenance Manuals on Wearables and Desktops

David Liu
School of Information Technology & Electrical Engineering, University of Queensland
LinkOne Product Team, Solutions Development, Mincom Ltd
s201140@student.uq.edu.au

Abstract

This paper describes a planned study comparing the effectiveness of Interactive Electronic Technical Manuals on wearable computers and desktop computers. It details the stages of the study, and how a usability test will be conducted with Mechanical Engineering students from the University of Queensland's Formula SAE-Australasia team. This paper concludes with the expected outcomes of the study and plans for future research.

1. Introduction

Modern mechanical equipment needs regular maintenance to prolong its service life and ensure correct operation. In the past, manufacturers distributed printed manuals that guided engineers through maintenance tasks. These manuals are being increasingly replaced by electronic systems known as Interactive Electronic Technical Manuals (IETMs).

Wearable computers are designed to provide access to information immediately at any time or place. IETMs viewed on wearable computers can potentially increase the performance of maintenance engineers by reducing the time spent looking up manuals, and increasing the number of faults detected during inspections.

This paper details a planned usability study comparing wearable computer-based IETMs with desktop-based systems.

2. Background

Maintenance activities. Equipment maintenance can be classified into four categories: inspection, preventative maintenance, troubleshooting, and repairs [1]. Typical maintenance manuals for a single equipment model often contain 100,000 pages, up to 625 items in individual inspection checklists, and over 50 steps in a single maintenance procedure [1]. Most engineers will work on several models of equipment across many manufacturers, so they often carry and refer to technical documentation to perform their work. IETM versions of maintenance documentation have significantly lower distribution costs, and are far easier for the engineers to transport.

Like the majority of desktop computer applications, most IETM software has been designed on the basis that the user is stationary at a workbench, with their full attention on using the computer. This is often not the case, as using a desktop or laptop computer can be impractical or even impossible. For example, mine shafts are highly constrained spaces, and engineers inspecting cranes need both hands for climbing.

When engineers need to refer to manuals in these scenarios, they have two choices. The first choice is to guess or ignore what they have forgotten, such as an inspection step. Alternatively, they can stop their work, go to the workbench with the computer, refer to documentation, and return to the equipment to continue working. If they are frequent, time consuming interruptions such as this will significantly reduce engineers' productivity.

Wearable IETMs. Wearable computers are often touted for these scenarios as they are ideally small and highly portable, supporting the user's primary task, maintenance. Engineers can have all of their required information immediately accessible, and they are able to document their work as it is performed.

By not having to continually referring back to the workbench for the technical manuals, engineers could spend more time performing maintenance procedures. Inspections may be completed more accurately because more information is instantly available to the inspector.

There are many usability issues that may affect the usefulness of wearable computer-based IETMs. For example, issues with head-mounted displays have been well documented [2, 3]. Few wearable computers have battery lives long enough to support full work days.

Wearable computers may also generate too much heat for the user to comfortably wear.

Previous studies. There have been several studies of wearable IETM systems in the past. They were evaluated for their effectiveness as on-the-job training tools [4, 56] and performance aids [6, 7, 8, 9]. In these studies, the baseline comparison was either paper-based manuals or no easily accessible reference material.

Technical documentation is increasingly being published only in electronic formats, so there is a need to study wearable IETM systems using traditional IETMs on desktops or laptops as the control condition. This ensures that any identified benefits of a wearable IETM system are due to the use of a wearable computer, not necessarily the proximity of support.

3. Investigative Method

In my undergraduate Software Engineering thesis I investigate the design and evaluation of an IETM system for wearable computers. Usability tests will be conducted using a Xybernaut Mobile Assistant® V wearable computer and MicroOptical SV-6 head-mounted display, provided on loan by Xybernaut GmbH. There are five phases to this study:

Phase 1 – User Requirements. Contextual inquiries will be performed by visiting worksites where technical manuals are routinely used. Engineers will be observed performing their work, with the goal of understanding the nature of their maintenance work, and how they reference technical manuals.

Phase 2 – User Interface Design. A set of user interfaces suitable for the users' needs and work environment will be designed. These interfaces will build on interface designs from previous studies [6, 7, 8]. Head-mounted displays, flat-panel displays, speech input, and rotary jog dials will be supported.

Phase 3 – Prototype Development. Mincom's LinkOne IETM software will be extended to support wearable computers. This allows existing technical manual content to be viewed without modification. However, it is expected that content published for viewing on a desktop will not be effective on a wearable computer.

Phase 4 – Usability Testing. The usability test subjects will be Mechanical Engineering students from the University of Queensland's Formula SAE-Australasia (Student Division) team. Each year, they design, develop and test a Formula-style race car that is entered into several national competitions.

Participants will perform a simulated preventative maintenance and repair scenario involving the front corner assemblies (which attaches the wheels to the chassis). Normally the procedure involves disassembly from the car's chassis, crack testing the A-arms and rods, replacing damaged parts and reassembly. In this experiment, the participants will assemble the front-right corner assembly, followed by the front-left corner assembly and then disassemble both corner assemblies.

Before testing, the technical manual content will be developed and will include pictures and video clips of each step in the procedures. Both groups will be given training immediately before the usability test. The tests will be conducted in the UQ Usability Laboratory. It is expected that there will be eight participants split evenly between the test and control groups. The participants' experience levels will be classified as either novice or expert.

The tests will measure the time required to assemble and disassemble each corner assembly as the main performance metric. If wearable IETM systems are effective, then less time will be needed to complete all of the tasks. As the left and right corner assemblies are mirror images of each other, the difference in their assembly times may show variations in learning rates between the test and control groups.

The entire test process from the initial training to performing the procedures will be videotaped for review. The participants will complete questionnaires and interviews before training, after training, and at the conclusion of the tests.

Phase 5 – Analysis of Results. With the limited number of participants, range of devices, and time frame for this study, the results will need to be carefully examined for their validity and impact on wearable IETM usability.

The major metric in the study (procedure times) may be significantly affected by factors relating to the testing process and not necessarily the usability of wearable computers in general. The test subjects have had years of experience with desktop computers, but will have used the MA® V only for a few hours. The students from the test group may focus too much of their attention on using the wearable and end up taking more time than otherwise to complete the maintenance procedures.

It is expected that the video footage, questionnaires, and interview transcripts will provide insights into the usability of the software and the wearable computer. The footage will show potential problems such as difficulty in reading text on the head-mounted display or using the system in general. The questionnaires and interviews allow the participants to express their thoughts and concerns regarding the system.

Progress to date. There has been some progress on the undergraduate thesis to date. Site visits for the contextual inquiries have been arranged for several local businesses. The prototype IETM system is under development and is running on the Xybernaut wearable computer. The UQ Usability Laboratory is being set up for the experiment and the technical manuals are being developed. All eight test participants from the SAE team have been selected.

4. Outcomes and Conclusions

There are three expected high-level outcomes from this study. First, the usability tests may show a significant reduction in the time required to perform the maintenance procedures and increase in the inspection accuracy with the wearable computer. This implies that there would be no major usability problems with the system. Given the results of previous studies, this is unlikely.

Second, there are no critical usability issues with the system, but navigating the system and referencing information is too slow to be effective. For example, it may be that the system is perfectly usable, but still faster for the user to refer to a desktop computer. In this scenario, further research would be conducted to identify bottlenecks and alternative solutions.

Third, accessing IETMs on wearable computers may be infeasible because of usability issues. Some may be solved with improvements in technology (short battery lives), whilst others will be more fundamental (head-mounted display rivalry and interference).

More research is needed after this study for general conclusions on wearable IETM usability. The work environments of engineers and IETM usage patterns need more investigation. The usability tests need to include a wider range of wearable computers and input/output devices under more representative conditions across all industries.

Even if wearable IETM systems prove to be effective, external factors may prevent their adoption. Wearable computers are significantly more expensive than desktop computers and require additional staff training. The workplace culture may be negative to the introduction of such pervasive technology. Finally, there may be benefits to engineers being interrupted, such as allowing them to stretch and take a short break from the work.

There is an enormous potential for wearable computers to provide faster and more effective access to IETMs, but further research is required before it can be realised.

5. Acknowledgements

This work is funded by Mincom Ltd and the Defence Science & Technology Organisation (Information Networks Division). Acknowledgements to Xybernaut GmbH for loaning the wearable computer. Prof Penelope Sanderson is the supervisor of the study.

6. References

[1] L. Bass, D. Siewiorek, M. Bauer, et al, "Constructing Wearable Computers for Maintenance Applications," *Fundamentals of Wearable Computers and Augmented Reality*, W. Barfield, and T. Caudell, ed., LEA Publishers, 2001, pp. 663-694.

[2] R.S. Laramee, and C. Ware, "Rivalry and interference with a head mounted display," *ACM Trans. on Computer-Human Interaction 9(3)*, ACM Press, New York, 2002, pp. 238-251.

[3] J.C. Kotulak, and S.E. Morse, "Oculomotor responses with aviator helmet-mounted displays and their relation to in-flight symptoms", *Human Factors 37(4)*, Human Factors and Ergonomics Society, 1995, pp. 699-710.

[4] E.S. Baudhuin, "Telemaintenance applications for *The WEARABLE*^TM^ *PC*," *Proc. 15th AIAA/IEEE Digital Avionics Systems Conf.*, IEEE Computer Society, Atlanta, 1996, pp 407-413.

[5] C. Thompson, L. Najjar, and J. Ockerman, "Wearable computer based training and performance support systems," *19th Interservice/Industry Training Simulation and Education Conf. Proc.* (I/ITSEC'97), 1997, pp. 746-752.

[6] D. Siewiorek, A. Smailagic, L. Bass, et al, "Adtranz: A Mobile Computing System for Maintenance and Collaboration," *Proc. 2nd Int'l Symp. Wearable Computers 1998* (ISWC'98), IEEE Computer Society, Pittsburgh, 1998, pp. 25-32.

[7] J. Ockerman, and A. Pritchett, "Preliminary Investigation of Wearable Computers for Task Guidance in Aircraft Inspection," *Proc. 2nd Int'l Symp. Wearable Computers 1998* (ISWC'98), IEEE Computer Society, Pittsburgh, 1998, pp. 33-40.

[8] J. Siegel, E. Hyder, J. Moffett, et al, "IETM Usability: Using Empirical Studies to Improve Performance Aiding," *Computer Science Technical Reports 2001*, Carnegie Mellon University, Pittsburgh, 2001.

[9] M. Boronowsky, T. Nicolai, C. Schlieder, et al, "Winspect: A Case Study for Wearable Computing-Supported Inspection Tasks," *Proc. 5th Int'l Symp. Wearable Computer 2001* (ISWC'01), IEEE Computer Society, Zurich, 2001. pp. 163-164.

Implementing an AR-Search Database

Seng Heng Foo
Singapore Polytechnic
500 Dover Road
Singapore 139651
foosh@sp.edu.sg

Angelina Novacheva
Stuttgart University of Applied Sciences
24 Schellingstrasse
70174 Stuttgart
noapg111@mars.rz.fht-stuttgart.de

Abstract

This paper presents a plan to carry out an experimental project on applying augmented reality within the compound of the Stuttgart University of Applied Sciences. The purpose is to build an augmented search database to search and locate the professors of the university using a webcam and a laptop. The plan focuses on the impact of varied lighting condition, size and pattern of markers, and multiple markers in implementing augmented reality. These factors influence the performance of ARToolkit in detecting the markers in a real-world scenario. ARToolkit [1] will be adopted in this project as it is popularly used and its manual is well documented. The results from the project will be analysed and appropriate recommendations will be put forward.

1. Introduction

During the last couple of years with the advances in hardware development and the availability of free software, Augmented Reality becomes an increasingly popular research area. Most of the research projects are confined to controlled laboratory conditions. Even though their scientific contributions are considerable, more evidence is required to confirm how a contemporary AR system could perform in a real-life situation.

The authors' experiment aims at solving a real-world problem - helping new students and visitors to navigate through the buildings of the Stuttgart University of Applied Sciences quickly and effectively. Prior to writing a program to perform this task, the authors proposed to investigate the factors that influence marker detection and these factors are varied lighting condition, size and pattern of markers, and multiple markers in the same vicinity. For the purpose of this project, a USB 2.0 web camera, a laptop, and the freely available software, ARToolkit, will be adopted. At the end of the project, the results and the respective analysis including the steps in implementing an AR-Search Database will be documented.

2. The AR-Search Database Project Plan

The goal is to provide the user with the possibility to download the AR-Search software from the Internet and install it on his personal laptop prior to his first visit to the Stuttgart University of Applied Sciences. As an initial step only a few rooms of the professor in Building 2 of the university were included in the search. However, following the same principle the trial test could be extended to encompass all the employees and facilities in the augmented search. Version 2.52 of ARToolKit with VRML library proved to be the best choice at the time of writing this plan for the purpose of marker detection. It is stable and fast in detecting a single marker within acceptable range and good lighting condition.

The experimental project will involve the following activities:

- Creating patterns [2] for the markers to be used in addition to the samples provided in ARToolkit. These patterns will be tested under various lighting condition and range to determine the suitability of employing them. In other words, these patterns need not be distinguishable to the human eye but they must be easily detected and distinguished by the image processing algorithms implemented in ARToolkit.
- Training the patterns directly at the places where they are deployed. Some patterns may be repeatedly used at many places and tests will be carried out to determine the ease of detecting the same pattern at various places.
- Grouping of well distinguishable markers in the same vicinity. The test will need to determine the ability of ARToolkit to differentiate these markers within the preview of sight of the webcam.
- The results of the above activities will be collated and analysed. Appropriate recommendations will

be provided for implementing an augmented search within the compound of the university.

- Based on the recommendations, an interface to ARToolkit will be created. This interface program will allow the user to select a staff member of the university for the augmented search. The photograph, e-mail, building, room number and research interest of the selected staff will be displayed to confirm the selection. The AR-Search will then begin. A draft interface as shown in Fig. 1 will be adopted.

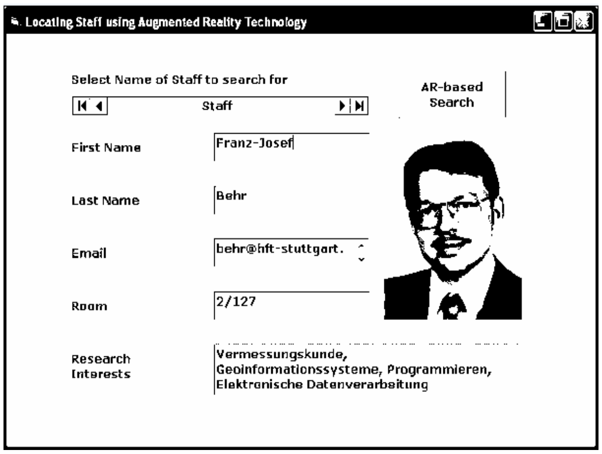

Figure 1: The main application window

3. Conclusion

The overall plan of the experimental project has been drafted and the authors agreed that the proposed activities to be executed are sufficient and meaningful. The results collected from the activities will be analysed and appropriate recommendations for improving marker detection and identification will be put forward. These recommendations will be adopted in implementing the augmented search.

4. References

[1] Kato, H., Billinghurst, M., and Poupyrev, I. (2000). ARToolKit version 2.33, on-line documentation, November 2000<http://www.hitl.washington.edu/>.

[2] Owen, C., Xiao, F., and Middlin, P. (2002). What is the Best Fiducial?, First IEEE International Augmented Reality Toolkit Workshop, Darmstadt, September 2002.

Context from multiple sensors for activity recognition

Georg Ogris
Institute for Computer Systems and Networks
UMIT, Innsbruck, Austria
georg.ogris@umit.at

Abstract— **This paper outlines the basic concepts of our current work on activity and context recognition using multiple sensors. It gives an overview which sensors are used - from simple motion sensors to digital video cameras - for which tasks and describes how they are used. Having multiple sensor information this paper states that the best way of sensor fusion is to interpret the information of a single sensor in the context of the knowledge that the other sensors provide. The paper also points out the benefits of multiple sensors for the data segmentation task. At the end this paper presents test suits and planned applications for the presented technique.**

Index Terms— **context, multiple sensors, sensor fusion, probabilistic modeling, activity and gesture recognition**

I. Introduction

The main applications for wearables are maybe intelligent, comfortable interfaces (communication between human and machine) and providing the user useful contexual information (communication between environment, machine and user). To judge what is useful in a certain moment the computer has to know both the state of the user as well as the state of the environment. So important research fields within wearable computing are (1) sensors - sensing human's whishes, needs, and activities and sensing what is going on in the (close) environment, and (2) mathematical concepts for understanding and modeling the *real world*.

The goal of our work is to improve the outcomes of well known (probabilistic) recognition algorithms, e.g., Dynamic Bayes Networks, by establishing contextual information via multiple, (body worn) sensors; with the following objectives:

– monitoring of workflows (especially in maintenance and repair services)
– providing contextual information, whereas contextual information means information that could be needed according the current user context.
– "wearable" interface / on demand human-machine interaction possibilities

The main challenges to be met within this work are both on the sensor side and on the computational side. A sensor environment has to be build up that must not disturbe the user. The user needs to act like usual, without feeling uncomfortable because of the equipment. So the common requirements for wearable and ubiquitos computing have to be met.

For the sensors itself there are no restrictions which types will be used. Starting with low level (ultrasonic) sensors fixed regarding both user and environment, the final setup may also use a high level, solely body worn sensor like a digital video camera.

After an outline of previous work in section II this synopsis starts with a description of the sensors in section III including techniques to fuse the sensor outputs (subsection III-C). Section IV gives an overview of the models that are used. After a summary of possible applications in section V this paper comes to an end with a short conclusion (section VI).

II. Related work

A lot of work has been done on activity recognition especially with accelerometers [13], [2], [5], [15], [3] and gesture recognition using different sensors for maintenance monitoring has been investigated by [14], for American sign language recognition by [4]

Work on ultrasonic sensors for user localization has been done by [11], [20], [19]. They describe the use of an ultrasonic location system for context aware computing. A more general overview of automatic location sensing techniques in the field of wearable computing give [12] and [18].

Smart environments have been build up in different forms: In [6], [9] the authors use image processing and sound analysis to interact with a smart office environment via gestures and voice. Helal *et al.* [11] presents a smart home that monitores user position via a ultrasonic location system. A more complex spatial monitoring system for context aware computing was presented by [10].

III. Sensors

Various types and various combination of types of sensors will be evaluated for their useability. The subjects of investigations are:

– reliability of the sensors
– reliability within changing environments
– quality of information content
– recognition rate
– convenience

The sensors may be body worn, e.g., motion sensors, or may be fixed on objects in the surrounding, e.g., devices for sensing the environmental context, or can have both devices fixed according a certain space (machine, room) and body worn devices, e.g., position sensors.

A. Sensor types

For the motion detection a combination of accelerometers and gyroscope will be used. We are using the orientation sensors MT9B from Xsens[1]. The Ultrasonic sensors system from hexamite[2] will be applied for the user localization, light and sound for close environment information extraction and change detection. Active infrared badges[3] will also be tested as an alternative for ultrasound, accelerometers, and gyroscopes within indoor applications.

B. Communication

Just a short note on how the sensors communicate their data: static sensors communicate via cable and serial port, body worn sensors need a second device that transmitts the serial data via bluetooth to avoid too heavy cabling.

The communication is done asynchron. Synchronisation is managed by the streamnet toolbox presented in [1] running on a central PC.

C. Sensor fusion

The different sensor information can either be seen as discrete features or discrete feature vectors of the input signal. That in return would result in two possible systems. The first possibility is a system that does the activity recognition on a single model that uses the output of the different sensors as inputs. And the second system would train a model for each sensor and fuse the recognition results in the end. Such systems result in a more accurate recognition rate as single sensor systems as long as the sensors deliver non redundant information.

Another approach for fusing the information of different sensors is evaluting the results of each sensor in the context of the outcomes of the other sensors. Information about the position can give prior knowledge about what gesture is likely to be expected. The same motion (e.g., raising the arm) will be interpreted in the context of the user position (e.g., phoning in case the user is standing close to the phone, and drinking in case the user is sitting or standing close to a table, or just a nonrelevant movement in case the user is not standing close to any interesting point). The other way round a specific motion can be used to refine the position information, because specific gestures are more likely to appear at specific positions in the room. In such a way multiple sensors can make recognition rates more accurate as well as in the first approach.

The advandage of the contextual method over the first approach is that it can eventually provide a solution for the segmentation problem. A hand gesture is maybe starting when the user reaches a certain point or when a specific sound appeares. The user reaches a certain point when a hand gesture starts or a specific sound reaches a specific intensity. This approach of course demands an exhaustive planning of the models used for the recognition tasks. Though the planning task should be feasible in case the number of activities that has to be modeled is not too high.

[1] http://www.xsens.com/

[2] http://www.hexamite.com/

[3] http://www.lukotronic.com/

IV. Modeling

Hidden Markov Models (HMM) [17], [7], [8] became a sort of state of the art method for (online) analysis of sequential, time series data. HMMs have first been used in the field of speech recognition [17] and are now used in many other domains. HMMs are probablistic models that are defined over a finite number of states, their observations, and state transition probabilities. The advantages of HMMs are time invariance and their ability to deal with real world data - data that has been recorded under real world conditions.

But with an increasing number of states, the complexity of such models can result in ineffectiveness due to explosion of the number of parameters unless the parameters can be reduced through an exhaustive planning, which demands an exact knowledge of the system that should be modeled. Dynamic Bayesian Networks [16] as a generalization of HMMs should overcome this problem.

For our test suits open source software will be used and extended for our needs. We decided to go for Murphy's BNT[4] and Intel's PNL[5] respectively. For a more applicable Software - applicable due to speed (online analysis) and slimness (wearable and hence low performance computer) - the tested software will be ported to our own framework [1].

V. Applications

A number of applications are thought of as test field for the outlined technique.

1) Monitoring of maintenance and repair activities: Well suited Sensors both in the environment as well as fixed on the clothing and equipment of the worker will monitor the progress of a specific activity. As starting test suit a simple maintenance and repair activity of, e.g., a bicycle will be monitored with about three type of sensors. Previous work has already been done at ETH Zuerich [14].

2) Teaching and learning workflows: A similar application as the one above would be a automated teacher for workflows. That can be a task like bycicle maintenance again or maybe a more complex task like a sports trainer. Whereas sports training brings up new interesting challanges in case the location is outdoors. Which increases especially the requirenments for the sensors.

[4] http://www.ai.mit.edu/ murphyk/Software/BNT/bnt.html

[5] http://www.intel.com/research/mrl/pnl/

3) Interface for smart office/home environments: The test suites for an interface will be in the begining simply a number of switches situated somewhere in an office, that can be turned on and off via focusing on them and via gestures. At the end an entire office environment shall be controlled by gestures and motions. An important requirement of such an smart office/home environment also includes contextual information allocation. Which can be helpful especially for the following scenario.

4) Information services in hospitals and fireguards.: Information at the right time at the right place with respect to the contextual situation of the user is essential, especially in life critical services. Evaluation of the usability of wearable systems in this fields will also be part of our work.

VI. Conclusion

This paper has presented a concept for a sensor environment for activity recognition. It furthermore outlines how this activity recognition can be made more accurate by using this multiple sensor scenario and what is more comes to the conclusion that contextual information should give essential hints for segementation and classification.

Our future work will be verifiing these concepts and ideas at the basis of the results of the test and application scenarios that were presented in section V.

References

[1] D. Bannach. An online sensor data processing toolbox for wearable computers. In *ISWC04, Eighth IEEE International Symposium on Wearable Computers*, Octobre 2004.

[2] L. Bao. Physical activity recognition from acceleration data under semi-naturalistic conditions. Master's thesis, Massachusetts Institute of Technology, 2003.

[3] C.V. Bouten, K.T. Koekkoek, M. Verduin, R. Kodde, and J.D. Janssen. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Transactions on Bio-Medical Engineering*, 44(3):136–47, 1997.

[4] H. Brashear, T. Starner, P. Lukowicz, and H. Junker. Using multiple sensors for mobile sign language recognition. In *Proceedings of the 7th IEEE International Symposium on Wearable Computers*. White Plains, NY, October 2003.

[5] J.B. Bussmann, W.L. Martens, J.H. Tulen, F.C. Schasfoort, H.J. van den Berg-Emons, and H.J. Stam. Measuring daily behavior using ambulatory accelerometry: the activity monitor. *Behavior Research Methods, Instruments, & Computers*, 33(3):349–56, 2001.

[6] C. Le Gal, J. Martin, A. Lux, and J. L. Crowley. SmartOffice: Design of an intelligent environment. *IEEE Intelligent Systems*, 16(4), 2001.

[7] Z. Ghahramani. *Adaptive Processing of Sequences and Data Structures. Lecture Notes in Artificial Intelligence*, chapter Learning Dynamic Bayesian Networks, pages 168–197. Springer Verlag, Berlin, c.l. giles and m. gori edition, 1998.

[8] Z. Ghahramani. An introduction to Hidden Markov Models and Bayesian networks. *International Journal of Pattern Recognition and Artificial Intelligence*, 15(1):9–42, 2001.

[9] D. Hall, C. Le Gal, J. Martin, O. Chomat, and J. L. Crowley. MagicBoard: A contribution to an intelligent environment. *Robotics and Autonomous Systems*, 34, 2001.

[10] A. Harter, A. Hopper, P. Steggles, A. Ward, and P. Webster. The anatomy of a context-aware application. In *Proceedings of the Fifth Annual ACM/IEEE International Conference on Mobile Computing and Networking, MOBICOM'99*, pages 59–68. Seattle, Washington, USA, August 1999.

[11] S. Helal, B. Winkler, C. Lee, Y. Kaddourah, L. Ran, C. Giraldo, and W. Mann. Enabling location-aware pervasive computing applications for the elderly. In *Proceedings of the First IEEE Pervasive Computing Conference*. Fort Worth, Texas, June 2003.

[12] J. Hightower and G. Borriello. Location systems for ubiquitous computing. *IEEE Computer*, 34(8):57–66, August 2001.

[13] Nicky Kern and Bernt Schiele. Multi-sensor activity context detection for wearable computing. In *Proceedings of the European Symposium on Ambient Intelligence*. Eindhoven, The Netherlands, November 2003.

[14] P. Lukowicz, J. A. Ward, H. Junker, M. Stäger, G. Tröster, A. Atrash, and T. Starner. Recognizing workshop activity using body worn microphones and accelerometers. In *Pervasive Computing: Proceedings of the 2nd International Conference*, pages 18–22. Springer-Verlag Heidelberg: Lecture Notes in Computer Science, April 2004.

[15] J. Mantyjarvi, J. Himberg, and T. Seppanen. Recognizing human motion with multiple acceleration sensors. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 2001.

[16] K. P. Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, University of California, Berkeley, Fall 2002.

[17] L. R. Rabiner. A tutorial in Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[18] J. A. Tauber. Indoor location systems for pervasive computing. Technical report, MIT Laboratory for Computer Science, Cambridge, MA, August 2002.

[19] A. Ward. *Sensor-driven Computing*. PhD thesis, University of Cambridge, August 1998.

[20] A. Ward, A. Jones, and A. Hopper. A new location technique for the active office. *IEEE Personal Communications*, 4(5):42–47, Octobre 1997.

Towards an ultra low power user activity recognition system

Niroshan Perera, Julian Randall and Gerhard Tröster
Wearable Computing Laboratory, Department of Electronics
Swiss Federal Institute of Technology, ETH Zürich, Switzerland.
perera@ife.ee.ethz.ch

Abstract

Human activity recognition systems comprise of body worn sensors and signal processing units to process the sensor data. Such systems need to be comfortable to wear i.e. small and unobtrusive and ultra low power to provide continuous service for several months. An ASIC as a dedicated processing unit will achieve the lowest possible power consumption and smallest size for such systems. We propose a systematic methodology to design an ASIC for activity recognition. Each design phase of the ASIC will focus on achieving the lowest power consumption and reasonable activity recognition accuracy based on experimental evaluations and optimizations.

1 Introduction

Human context and activity recognition [2, 4] is one of the key applications of wearable systems in the future. Context and activity recognition has a wide range of applications, such as human-machine interaction [7, 10], health monitoring [6], context based information retrieval [9] etc., where the wearable system automatically adjusts its configuration and functionality to provide services to the user.

Depending on the user's location, state and environment, human activities may have different forms ranging from simple hand gestures to complex and non-recurring actions. Although recognizing every human activity in every situation is a challenging task, basic human activities (standing, sitting, walking, opening a door, handshaking etc.) within specific scenarios can be recognized with a high degree of accuracy by processing data collected by multiple sensors (audio, acceleration etc.) worn on the body.

From a user wearability point of view, these sensors and processing nodes must be small, unobtrusive and low power. In addition, they can be seamlessly integrated into clothing and need minimal or no maintenance during their service life.

Current hardware platforms for sensor data processing use relatively large and power intensive digital signal processors (DSPs) [3, 8]. These hardware platforms provide a very flexible and versatile signal processing environment to run feature extraction and classification algorithms on multiple sensor data streams. However, due to their bulkiness and high power demand, these standard platforms are incompatible with the vision of wearable low power context sensing systems. Therefore, we need to explore and design systems with dedicated hardware architectures (ASICs - Application Specific Integrated Circuits) to achieve the lowest power consumption and smallest size.

2 Hypotheses for the proposed system

We expect the proposed wearable activity recognition system to fulfill the following criteria:

Wearability: the system needs to be small (less than 3cm^3 in volume) and unobtrusively integrated in to everyday clothing. In addition, the system would need little or no maintenance (e.g. battery change or recharging) for many months.
Low Power consumption: In order to minimize maintenance and prolong the service life, the system needs to achieve the lowest possible power consumption. Therefore, the system should comprise of low power consuming devices. In addition, the system will operate with the lowest possible duty cycle, which still gives sufficiently accurate results.
Reconfigurability: The system needs to adapt to situations to provide useful information to the user. Furthermore, the system needs to have the capability to shutdown and wake up automatically depending on user activity to save energy.
Autonomous Operation: The system needs to operate autonomously by harvesting ambient energy. e.g. using Solar cells, Inertial generators etc.

3 System design approach

It would be intractable to design and optimize a general activity recognition system to fulfill all the criteria mentioned in section 2. Therefore, we will focus on systems for specialized applications and scenarios where activity recognition is useful and moderately complicated. Furthermore, as a first step, we will investigate methods and possibilities to minimize the size and power consumption of such systems.

Since signal proceeding hardware is the biggest and most power consuming element in an activity recognition system, our investigation will concentrate on designing an ultra low power, miniaturized hardware (an ASIC) to implement feature extraction and classification algorithms.

We propose a systematic approach to design an ASIC for a low power activity recognition system as follows.

Scenario Selection: Evaluate scenarios where activity recognition is useful and feasible with reasonable computing effort.

Sensors: Find the most useful and energy efficient sensor or sensor combination to collect data for a given scenario.

Sensor placement: Determine the ideal locations for the placement of sensors.

Data Resolution and sampling rate: Determine the minimum sampling rate and bit resolution of sensor data.

Feature selection: Determine which features from sensor data are useful for further processing by classification algorithms.

Algorithms: Simple and efficient algorithms need to be found for data fusion and classification. They need to be tested with selected features from sensor data to determine recognition accuracy and computing complexity. Furthermore, we explore the possibility to approximate the chosen algorithms with minimum impact on accuracy and performance.

Scheduling: The computing tasks need to be scheduled to minimize overhead.

Duty cycle: Estimate the lowest possible duty cycle for the system to achieve a reasonable performance.

Validation: The selected feature extraction and classification algorithms will be implemented on a PCB prototype with sensors and a DSP as the processing unit.

ASIC design: An ASIC will be designed to implement the algorithms with the optimum operating parameters. We focus on a low power ASIC design flow and use multiple techniques to efficiently implement our algorithms. Different architectures, logic optimizations and voltage scaling schemes will be evaluated and compared to measure the impact on power consumption. We will also make provisions to run the ASIC with minimum duty cycle with clock gating techniques. The multi algorithm ASIC will have the capability to adapt itself to different situations by shutting down and powering up different blocks as required to minimize power consumption.

In general, we demonstrate that the proposed design methodology puts emphasis on low power operation of the ASIC in every step in the design process by choosing appropriate scenarios, sensors, algorithms, ASIC design methods and optimizations.

As an extension of our project we will investigate the following topics to build the complete activity recognition system including the ASIC. An initial study on these topics are reported in [1].

Ambient energy harvesting: Evaluate different ambient energy harvesting methods and determine the power generation capability of each method. By comparing power available from such schemes and typical power consumption of the system, we will determine whether it is feasible to make the system autonomous.

Hardware component selection: Find smallest and lowest power consuming *off-the-shelf* peripheral devices for the final system. The system needs sensors, A/D converter, power supply regulator and an RF transceiver. Special care is needed to ensure that components selected will work together without seriously affecting the performance or the power consumption.

System Integration: Optimizations for the final system for small size (low volume) and power consumption. This final step needs expertise in packaging solutions for micro assembly.

The hardware architecture for the proposed system is shown in figure 1.

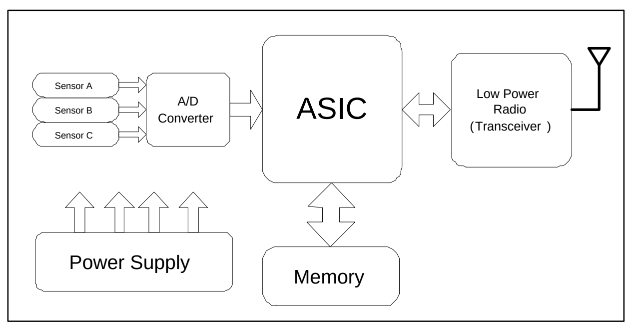

Figure 1. Hardware architecture for the proposed system

4 Experiments and design methodology

Experiments and simulations are necessary for the overall optimization towards the design of the ASIC. Therefore, we focus on each aspect mentioned in section 3 to find optimum parameters with several iterations.

Firstly, we need to closely examine human activities and select common, recurring and useful everyday scenarios for activity recognition for ultra low power implementation. Some interesting and useful scenarios such as typical office activities, workshops, assembly lines, repair and maintenance of vehicles as reported in [11] will be evaluated. Once the scenarios are finalized, we need to compare which combinations of sensors will give us the best information in order to detect activities. Furthermore, we will evaluate the ideal placement of sensors on the body (wrist, elbow, torso, knee etc.) to collect information.

Next, different features from each sensor data stream will be extracted. As described in [5], we will extract features both in time and frequency domain. We will evaluate fluctuation of amplitude, mean value and variance in the time domain and frequency centroid, bandwidth, spectral roll-off point and fluctuation of amplitude-spectra in the frequency domain.

In the next step, we will evaluate different classifiers such as neural networks, decision trees, gaussian model-based classifiers, hidden markov models with the above mentioned features. Since not every feature or combination of features works well with every classifier, we need to find the best classifier for given set of features which gives sufficiently accurate results.

The process of feature selection and classification will be repeated for different senor combinations and placements until we find several promising alternatives with sufficiently accurate results. During this iterative process, minimum sampling rates and bit resolution of sensor data will also be determined.

A PCB prototype will be built (with sensors and a DSP) to implement the algorithms. This prototype will serve to validate our approach by testing our algorithms with real time data. Furthermore, we will be able to evaluate the impact of noise on the accuracy of results.

The ASIC design process will commence once the best algorithms and operating parameters are determined experimentally.

It is essential that a considerable effort is made to optimize the ASIC to make it as low power as possible. Optimizations can be done at architectural, RTL (Register Transfer Level) and gate level. The design will be implemented in VHDL, simulated and synthesized. After synthesis, the gate list is used to generate the switching activity file for power estimation. In addition, task scheduling and voltage scaling options will also be explored to keep the power consumption to a minimum.

Finally, we will explore the possibility to integrate the ASIC with sensors, RF transceiver and other components to build the fully functional activity recognition system.

5 Conclusions and outlook

We have demonstrated the need for ultra low power human context and activity recognition systems in future wearable systems. Furthermore, we have also shown that an ASIC as the dedicated signal processing unit for such a system will minimize the power consumption and size. In order to design an ASIC for activity recognition system, we have summarized the performance requirements, a comprehensive and systematic design approach and an experimental study to validate our approach. We have also demonstrated how low power considerations can be applied to every design aspect of the proposed ASIC as well as the proposed system.

References

[1] N. B. Bharatula, S. Ossevoort, M. Stäger, and G. Tröster. Towards wearable autonomous microsystems. In *Pervasive Computing: Proceedings of the 2nd International Conference*, pages 225 – 237. Springer-Verlag Heidelberg: Lecture Notes in Computer Science, Apr. 2004.

[2] B. Clarkson, K. Mase, and A. Pentland. Recognizing user context via wearable sensors. In *ISWC 2000: Proceedings of the Fourth International Symposium on Wearable Computers*, pages 69 – 75, Oct. 2000.

[3] J. Farringdon, A. Moore, N. Tilbury, J. Church, and P. Biemond. Wearable sensor badge and sensor jacket for context awareness. In *The Third International Symposium on Wearable Computers, 1999*, pages 107 – 113, Oct. 1999.

[4] S. Lee and K. Mase. Activity and location recognition using wearable sensors. *IEEE Pervasive Computing*, 1(3):pages 24 – 32, July-Sept 2002.

[5] D. Li, I. Sethi, N. Dimitrova, and T. McGee. Classification of general audio data for content-based retrieval. In *Pattern Recognition Letters, 22(2001)*, pages 533–544, 2001.

[6] P. Lukowicz, U. Anliker, J. Ward, G. Tröster, E. Hirt, and C. Neufelt. Amon: A wearable medical computer for high risk patients. In *ISWC 2002: Proceedings of the 6th International Symposium on Wearable Computers*, pages 133–134, Oct. 2002.

[7] P. Lukowicz, J. A. Ward, H. Junker, M. Stäger, G. Tröster, A. Atrash, and T. Starner. Recognizing workshop activity using body worn microphones and accelerometers. In *Pervasive Computing: Proceedings of the 2nd International Conference*, pages 18–22. Springer-Verlag Heidelberg: Lecture Notes in Computer Science, Apr. 2004.

[8] S. Park, I. Locher, A. Savvides, M. Srivastava, A. Chen, R. Muntz, and S. Yuen. Design of a wearable sensor badge for smart kindergarten. In *ISWC 2002: Proceedings of the sixth International Symposium on Wearable Computers*, pages 231– 238, Oct. 2002.

[9] B. Rhodes. Using physical context for just-in-time information retrieval. In *IEEE Transactions on Computers, Volume: 52 , Issue: 8*, pages 1011 – 1014, Aug. 2003.

[10] M. Stäger, H. Junker, M. von Waldkirch, and G. Tröster. Using wearables in maintenance: A modular test platform. In *TCMC 2003: Workshop on Wearable Computing*, Mar. 2003.

[11] M. Stäger, P. Lukowicz, N. Perera, T. von Büren, G. Tröster, and T. Starner. SoundButton: Design of a Low Power Wearable Audio Classification System. In *ISWC 2003: Proceedings of the 7th IEEE International Symposium on Wearable Computers*, pages 12–17, Oct. 2003.

Interactive Mixed Reality Rendering in a Distributed Ray Tracing Framework

Andreas Pomi and Philipp Slusallek
Computer Graphics Group, Saarland University
Im Stadtwald - Building 36.1, 66123 Saarbrücken, Germany
{pomi, slusallek}@cs.uni-sb.de

Abstract

Photo-realistic rendering methods are required to achieve a convincing combination of real and synthetic scene parts in mixed reality applications. Ray tracing is a long time proved candidate, but when it comes to interactive frame rates GPU based rendering has usually been preferred. Recent advances in distributed interactive ray tracing suggest to explore also its potential for mixed reality rendering applications.

In this paper we show how simple extensions to the OpenRT ray tracing framework enable interactive applications from different ends of the mixed reality spectrum. We point out the benefits of a ray tracer over traditional GPU approaches when it comes to flexibility and complex compositing operations.

1. Introduction

The spectrum of mixed reality applications spans from inserting real objects or actors into a virtual scene (e.g. a virtual TV studio) up to augmenting live video with rendered objects (AR). In order to provide a convincing result for the viewer, photo-realistic rendering methods are needed for taking care of visual cues like shadows and reflections.

Current GPU hardware can be used to create photo-realistic mixed reality scenes (e.g. [3]), but seems rather inflexible for this task. Recent advances in ray tracing [10] provide a new basis for photo-realistic rendering at interactive frame rates. Ray tracing enables simple implementation of complex compositing methods ([4]) that are hard to achieve with GPU based approaches.

In the following we give a brief introduction to interactive ray tracing, and some basic extensions for mixed reality. We show application examples from different ends of the mixed reality spectrum and point out benefits of ray tracing over other approaches.

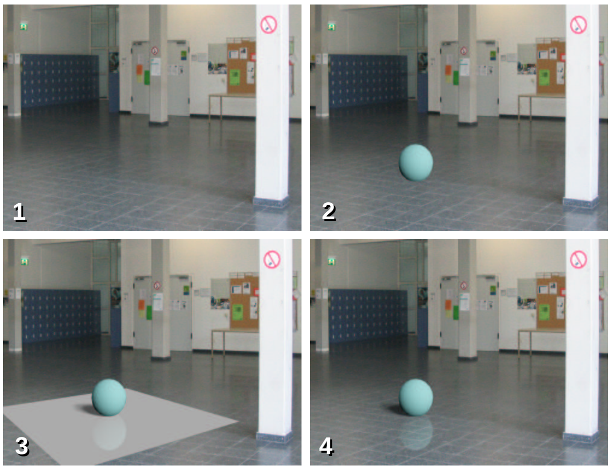

Figure 1. Simple AR compositing example with OpenRT. 1) Background view. 2) Inserted object, without a shadow it seems to float. 3) A stand-in object for the floor to generate a soft shadow and reflection from real captured illumination. 4) Final view after applying a differential rendering method.

2. Distributed Interactive Ray Tracing

There were a number of attempts to perform ray tracing at interactive frame rates e.g. on GPUs [9]. A pure software (CPU) based distributed approach [10] has proven its potential and provides a much more flexible framework than GPU approaches in general.

Ray tracing is capable of arbitrary complex geometry and advanced shading methods e.g. using interactive global illumination [2]. It is an ideal candidate for photo-realistic mixed reality rendering. An OpenGL-like API (OpenRT [5]) ensures easy implementation of applications. Compared to GPU based approaches pure software ray tracing does not suffer from hardware resource conflicts.

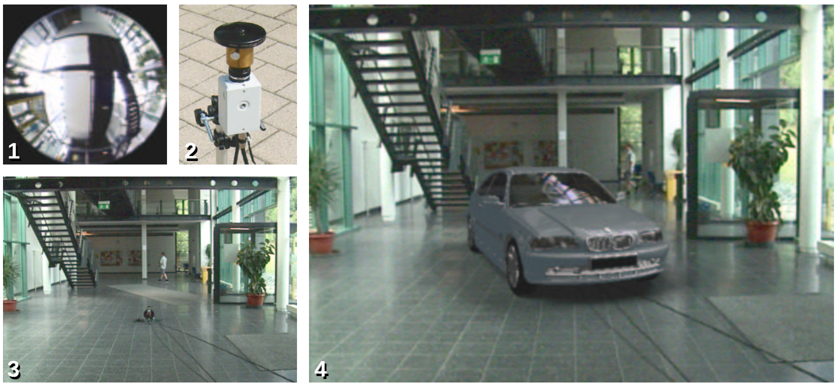

Figure 2. Car model inserted into a live captured background. The car is lit by the incident light captured with a HDR video light-probe at the car position. 1) A light-probe HDR frame. 2) HDR camera with 180 degree fish-eye lens. 3) A background frame. 4) Final video output with soft shadows cast onto the real floor. Note the transparent car windows.

A typical software implementation of OpenRT consists of a number of computers connected by a commodity network. The application runs on a server on top of the OpenRT API, which hides the distribution issues from the user. The rendered image is split into tiles (e.g. 32x32 pixels), which are scheduled for rendering on the clients.

Each clients runs a highly optimized ray tracer implementation that is roughly a factor 30 faster than traditional ray tracing systems [10]. The rendering performance can be easily scaled by adding more clients, and is only limited by the available network bandwidth.

2.1. Mixed Reality Extensions to OpenRT

A simple approach to mixed reality is to include the real world in the rendering process by using live video streams. In [8] we came up with two extensions to the OpenRT framework: *Streaming video textures* and *in-shader view compositing*.

Video textures allow for synchronized streaming of live video from dedicated video texture servers with low latency. A multicast networking approach ensures scalability in the number of rendering clients.

In-shader view compositing is a variation to video textures for inserting a video background behind rendered objects. Compared to a video texture, we only stream the colors of the appropriate tile from the background instead of the whole image. This yields lower network bandwidth and latency. The background color for each primary ray can be accessed in the shader for performing *differential rendering* methods [4].

3. Application Examples

3.1. Virtual Object Insertion into Live Video

Figure 1 shows a simple AR compositing example: A sphere is rendered in front of a video background. A stand-in object for the local floor is used to create a soft shadow and a reflection. Differential rendering [4] integrates the effect into the background video.

A more sophisticated example using image-based lighting is shown in Figure 2. A virtual car is placed into a live video background. A HDR video camera with a 180 degree fish-eye lens is used to capture the incident light. The light-probe image is streamed as a video texture, and is used for both, lighting the car and creating a soft shadow on the real floor.

3.2. Actor Insertion into Synthetic Scenes

To insert live actors into a virtual scene, billboards can be used. A video texture streams the video of the actor captured in front of a green or blue background. Chroma-keying is performed inside the billboard shader (*in-shader*, see Figure 3.1).

However billboards have a number of drawbacks due to their 2D nature (see e.g. [7]). Our framework enables another approach: A number of video textures with different views of the actor, combined with a 3D image-based visual hull [6] reconstruction algorithm provides correct shadows and reflections compared to 2D billboards (see Figure 3.2). A visual hull shader assigned to a box provides a simple integration into the scene. Shadows and reflections are computed automatically in a ray tracing framework [7].

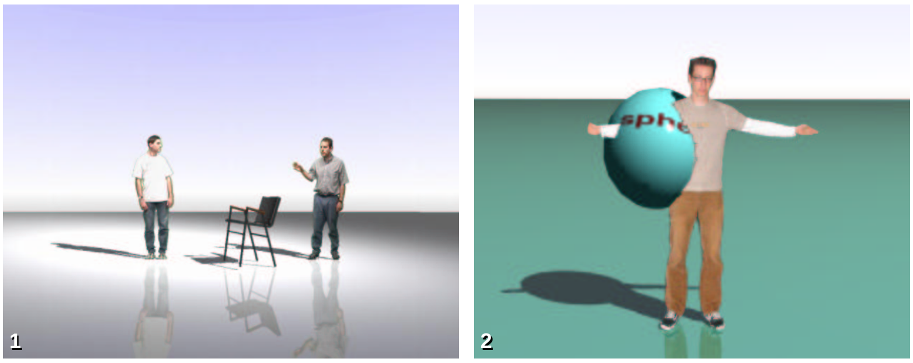

Figure 3. 1) 2D video texture billboards with in-shader compositing. 2) In-shader 3D visual hull reconstruction with exact occlusion by a sphere, lighting effects and reflection.

4. Results

Table 1 shows the frame rates we currently achieve for our examples at a video resolution of 640x480. Note the number of samples used for lighting and soft shadow generation. The rendering clients were dual Athlon MP 1800+ PCs connected via 100Mbit/s Ethernet to a central switch. The rendering server is connected via Gigabit. The video delay is about 3-4 frames.

5. Conclusion and Future Work

We have shown that interactive ray tracing is indeed suitable for mixed reality rendering, and there still remains a large space for further exploration.

Compared to rasterization based approaches we still achieve lower frame rates, but a pure software ray tracer allows for much simpler implementation and combination of rendering effects than current GPUs since no hardware resource conflicts can occur and arbitrary complex rendering algorithms can be implemented.

For the future, we aim for providing an ARToolkit [1] version for OpenRT. We plan farther research on interactive image-based lighting (using stand-in geometry [3]), and to integrate the lighting process in our OpenRT based interactive global illumination system [2].

Example	Samples	#CPUs	fps
Sphere, soft shadow (Fig. 1)	40	8	9.1
Sphere, hard shadow	1	8	20.5
Car, 208259 tris (Fig. 2)	40	24	4.5
Billboards (Fig. 3.1)	1	8	16
Visual hull shader (Fig. 3.2)	1	16	15.5

Table 1. Frame rates

References

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

[2] C. Benthin, I. Wald, and P. Slusallek. A Scalable Approach to Interactive Global Illumination. *Computer Graphics Forum*, 22(3):621–630, 2003. (Proceedings of Eurographics).

[3] J. Cook, S. Gibson, T. L. Howard, and R. J. Hubbold. Realtime photorealistic augmented reality for interior design. In *ACM Siggraph 2003 Conference Abstracts and Applications*. ACM Press, July 2003.

[4] P. Debevec. Rendering Synthetic Objects into Real Scenes: Bridging Traditional and Image-based Graphics with Global Illumination and High Dynamic Range Photography. *Computer Graphics (Proceedings of SIGGRAPH 98)*, 32(4):189–198, 1998.

[5] A. Dietrich, I. Wald, C. Benthin, and P. Slusallek. The OpenRT Application Programming Interface – Towards A Common API for Interactive Ray Tracing. In *Proceedings of the 2003 OpenSG Symposium*, pages 23–31, Darmstadt, Germany, 2003. Eurographics Association.

[6] W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, and L. McMillan. Image-based visual hulls. In K. Akeley, editor, *Siggraph 2000, Computer Graphics Proceedings*, pages 369–374. ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 2000.

[7] A. Pomi, S. Hoffmann, and P. Slusallek. Interactive In-Shader Image-Based Visual Hull Reconstruction and Compositing of Actors in a Distributed Ray Tracing Framework. In *1. Workshop VR/AR, Chemnitz, Germany*, 2004.

[8] A. Pomi, G. Marmitt, I. Wald, and P. Slusallek. Streaming Video Textures for Mixed Reality Applications in Interactive Ray Tracing Environments. In *Proceedings of Virtual Reality, Modelling and Visualization (VMV)*, 2003.

[9] T. J. Purcell, I. Buck, W. R. Mark, and P. Hanrahan. Ray Tracing on Programmable Graphics Hardware. *ACM Transactions on Graphics*, 21(3):703–712, 2002. (Proceedings of SIGGRAPH 2002).

[10] I. Wald, T. J. Purcell, J. Schmittler, C. Benthin, and P. Slusallek. Realtime Ray Tracing and its use for Interactive Global Illumination. In *Eurographics State of the Art Reports*. Eurographics, 2003.

Location by solar cells: an experiment plan

J. Randall, O. Amft, G. Tröster
Wearable Computing Lab., ETHZ, Zurich, Switzerland
randall@ife.ee.ethz.ch

Abstract

Present indoor navigation techniques are relatively costly and little widespread. We believe that solar cells might be able to provide a partial solution by minimising hardware investment, providing useful *data* about the user environment and acting as an energy harvesting *power* source. In this paper the related state of the art is briefly reviewed, a vision of the implementable systems is provided, hypotheses with regard the use of solar cells as location aids are developed and planned experimentation is described.
We conclude that obtaining both "*power and data*" would be a novel use of an energy harvesting technology in the wearable computing domain. Such a combination would support the case that other energy harvesting technologies could also be used for providing energy and data. Applications can also be forseen beyond wearable computing.

State of the art

Indoor location can be achieved with a variety of technologies. These in the area of wearable computing include, but are not limited to:

- ultrasound (e.g. AT&T Laboratories BAT, MIT Cricket),
- radio frequency (e.g. WLAN: Microsoft Research RADAR, RFID: University of Washington SpotON, mobile telephone triangulation),
- infra-red (e.g. AT&T Laboratories Active Badge),
- (CCD) camera (MIT Patrol: Starner et al.).

Combinations of the above technologies such as radio frequency and ultrasound have also been studied (e.g. University of Bristol Low Cost Indoor Positioning System, University of Tokyo DOLPHIN: Fukuju et al.).
Details of all the above technologies can be found using a standard internet search tool such as Google, so for this reason and due to the limited pages of this document, no references to particular publications are provided.
Position has also been measured with lateral effect photodiodes (LEP) for example in robotics [1] and automated vehicles as well as with photodiode arrays in the case of atomic force microscopy [2]. However, both the latter required either a point source light (laser beam) or lens focusing of the light onto the LEP as in [3]. Narrow light beams are atypical indoors.
There are other technologies that are usually used as a complement to the location technologies already mentioned such as accelerometers, magnetic sensors (compass) and gyroscopic sensors [4].

Vision of system that is proposed

None of the above approaches to location explicitly use overhead lighting as the principal location reference. However a number of environments such as office corridors and manufacturing plants are lit for the majority of the day by fixed electrical light sources.

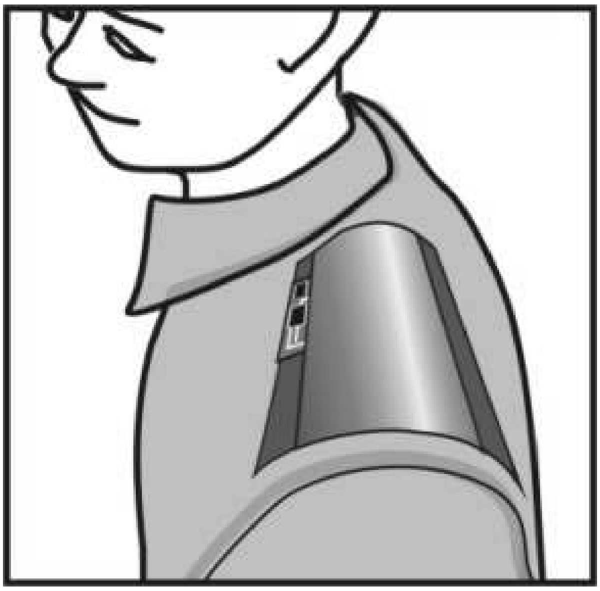

Figure 1: Solar cell location epaulette concept

A location tracking device based on existing technologies might look like Figure 1. The main components are a flexible solar module, a peak current detector and a pulse RF communication device. Not shown in Figure 1 is the RF receiver for the collected data that might be on-body or in the environment.

It is an important hypothesis of this project that shoulder mounted solar cells satisfactorily detect indoor lights. The level of accuracy would need to be sufficient for certain applications. If this were to be the case a number of advantages, all of which are adapted to wearable computing, would include:

- solar photovoltaic modules are relatively cheap sensing devices (around 1US dollar/module)
- as the modules are cheap, it may be feasible to have a number of them for each user in order to improve detection and location accuracy
- the modules deliver a relatively low data rate compared with other systems such as CCD camera; this suggests that wireless communication of the data would be low power
- solar modules can be prepared in thin films that would be very low weight and flexible (i.e. can be integrated into clothing)
- with sufficient surface, the solar module might provide power to location tracking devices or for other on-body computing purposes
- people wear a number of garments that cannot necessarily be electrically connected; the “power and data” concept may support wearable autonomous wireless sensor node solutions

Furthermore, research is already in progress on photovoltaic fibres [5]. These might be woven into fabrics thus allowing effective solar modules to be better integrated into clothing.

Conversion of the light detection data into usable location or navigational information generally requires one or more algorithms. Research in this area, both in robotics and ubiquitous computing, often apply Bayesian techniques that include Kalman filters, Hidden Markov models (HMM), dynamic Bayes nets and particle filters. It is proposed in an initial experiment in a hallway with regularly positioned fluorescent tubes that the periodic intensity variation detected from walking under the tubes be used to determine movement and walking speed. Simple time warping for template matching can be used to recognize the light intensity pattern curve. A HMM can then be developed and tested to associate the intensity peak of each sensor to a state probability. Assuming a large training database can be used, the classifier will be more robust against feature variations than the time warping approach that relies on a single sample model.

Hypotheses

<u>Location and activity tracking</u>

The basic requirement of each solar (photovoltaic) module is to “recognize” light sources. Detection is expected to be dependent on light type, size, height and position; the resulting accuracy should be comparable to a CCD camera with the advantage of reduced resources. It should also be possible to automatically distinguish daylight and fluorescent light by frequency and intensity. Changes in the ambient lighting will indicate user activity.

It is expected that the light levels experienced by a user moving down a corridor can be modeled and the models validated by the experiments mentioned in the next section.

It is hypothesized that increasing the number of solar modules per shoulder will allow more detailed information to be collected. For example, two solar modules on the same shoulder oriented in the axis of walking may be used to determine speed of movement. With three or more photovoltaic modules positioned equidistant from one another on the shoulder, it is hypothesized that orientation detection will be possible.

Provided the above basic data and a suitable algorithm implementation, it is proposed that location accuracy can be improved by using complementary devices. Such devices may include, but are not limited to, on-body sensors (e.g. step counter) and off-body devices (e.g. beacons or via fluorescent tube broadcast [6]). Location accuracy could also be improved with complementary information such as a map of the light sources (daylight and electrical) in a building, typical usage patterns of the lights and typical habits of the user(s) using the location devices. Finally, it would be convenient if the location system could learn from mistakes.

<u>Energy harvesting (or towards an autonomous sensor)</u>

We estimate that 0.7mW power can be collected from 80cm^2 of active photovoltaic area positioned on an adult shoulder with a corridor fluorescent tube directly overhead. It is expected that sufficient average power is available to run a low power microprocessor that is able to handle signal conditioning, filtering, artifact removal and simple feature extraction. If this is the case, then pulse RF communication should also be feasible. Having sufficient average power for these functions implies a design trade-off between maximizing energy collected (i.e. by having the solar modules in close proximity and in a horizontal plane) and maximizing the location data collected, typically be separating the solar modules which suggests they will not be oriented in a plane that is perpendicular to the incident light.

Proposed experiments

For reasons of limited space, this section does not cover all above mentioned hypotheses. The focus is how to design the experimental system, what initial tests should be carried out and how to validate the concept.

Build first experimental system

The first iteration of hardware will consist of three photovoltaic solar modules per shoulder of around 27cm^2 each that can be individually monitored for current at minimum 200Hz. These six modules should be calibrated before and after use. The modules should be easily fixed on the shoulder and have some flexibility in their exact location so that the influence of position on the shoulder may be investigated. We expect to use "Velcro" for this. The modules shall be wired to a data storage device (e.g. laptop) that should be carried by the user. A system will be required for effective collection, annotation and data base access of the photovoltaic module data-sets. These data-sets will be used to validate models of the user moving down a corridor for example. Whilst in general peak light intensity is expected under each tube, special cases should also be catered for such as an intensity peak created by super-position of two adjacent lights.

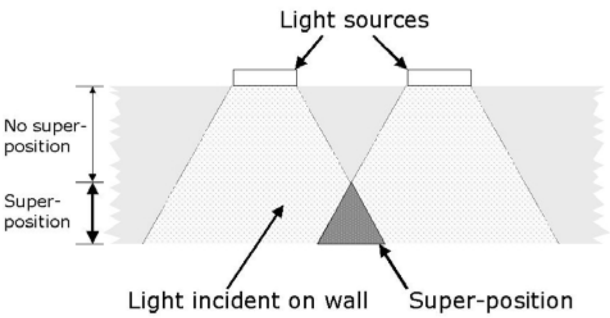

Figure 2: A special case: light source superposition

Initial experiments

Data will be collected for light intensity and frequency using the shoulder data collection system in various indoor environments, including near windows during daylight, under normal activity conditions. The shoulder system will also be used for testing a variety of individuals with characteristics representative of children, women and men performing everyday activities. Individuals should be tested who have long hair, wear hats and other related light obstacles to determine the related light detection deterioration and to determine the ideal location of the photovoltaic modules for each class of individual; potentially linear discriminant analysis will be useful for this. Data will also be collected for less normal activity such as swaying under a light or running back and forth near a light.

Based on these data-sets, algorithms will be developed that are able to locate an adult male walking at 1ms^{-1} down a predefined corridor within 1m using an off-line classification. For the corridor example, three algorithms based on different Bayesian approaches such as a HMM will be compared. The possibility of mixing algorithms to provide more accurate location data will be investigated, as well as the extent to which a map with skeleton data (walls and light positions) can improve location detection. The potential for complementary on-body sensors or off-body beacons to improve location accuracy will also be investigated.

All the above will lead to the building of a second iteration of the hardware based on what has been learnt through the experiments and using flexible solar modules.

Conclusion

A novel technique for low-cost location tracking has been proposed. It holds a number of advantages including the possibility of multi-functionality (sensing and a power source from the same device) and low cost. A number of hypotheses have been presented and initial experiments based on these have been described. The latter experiments are scheduled to take place at the Wearable Computing Lab. from November 2004 onwards; results are anticipated by February 2005.

References

[1] An intelligent active range sensor for vehicle guidance: system overview, Pears, N.E.; Proceedings of the International Conference on Intelligent Robots and Systems, 4-8 Nov. 1996, Volume: 1, pp 81 - 88

[2] Position measurement using atomic force microscopy, Doi, O.; Torii, A.; Ueda, A.; Proceedings of 2001 International Symposium on Micromechatronics and Human Science, 9-12 Sept. 2001, pp 143 – 148

[3] Accuracy of position detection using a position-sensitive detector, Schaefer, P.; Williams, R.D.; Davis, G.K.; Ross, R.A.; IEEE Transactions on Instrumentation and Measurement, Vol. 47, Issue 4, Aug. 98, pp 914 - 919

[4] Location Estimation Indoors by Means of Small Computing Power Devices, Accelometers, Magnetic Sensors and Map Knowledge, Vildjiounait, E.; Malm E-J.; Kaartinen J.; Alahuhta P.; Proceedings of Pervasive 2002 Conference, pp. 211 - 224, 26 - 28 Aug. 2002

[5] Photovoltaic effect in threads covered with CdS, Shur M.S., Gaskiene G., Rumyantsev S.L., Rimeika R., Gaska R., Sinius J.; Electronics Letters 37, 16, 1036-1038

[6] http://www.talking-lights.com/

Real-Time Photorealistic Rendering for Augmented Reality in Unprepared Environments

Stephen DiVerdi
Department of Computer Science
University of California, Santa Barbara, CA 93106
sdiverdi@cs.ucsb.edu

Abstract

Our goal is the creation of an augmented reality system which can be used in arbitrary settings with little to no preparation or setup, while maximizing realism of augmentation. The final result should provide real-time interaction with virtual objects that are seamlessly integrated with the real world. Towards this goal, we will investigate techniques to automatically or semi-automatically acquire dynamic scene data – geometry and lighting conditions. Depending on the limitations of the acquisition techniques, scene video data will be used for image-based error correction of the rendered scene.

1 Introduction

In recent years, significant attention has been given to increasing realism in augmented reality applications. Improvements in desktop computing capabilities, especially in the area of real-time 3D graphics rendering, has made powerful systems available for relatively low cost. The importance of realism is driven by increasing demand from the entertainment industry for more realistic computer generated imagery, as well as from the simulation and medical sectors. Applications such as simulation training for fire fighters or pilots require realism to give trainees experience as similar as possible to what they will see in the field. Video games and other forms of interactive entertainment also benefit from increased realism, to create a better sense of immersion in participants.

A number of systems have been developed, which make progress towards improving the visual fidelity of virtual objects in augmented reality. Each of these systems has made necessary tradeoffs of flexibility and preparation cost to achieve more realistic results. Some of the earliest work was done by Fournier et al. [2], creating a system that could add virtual objects to an image of a real scene with a carefully constructed scene model and a progressive radiosity renderer. Loscos et al. [7] improved on the concept with an incremental radiosity renderer that could achieve framerates of approximately one frame per second, which allowed for interactive manipulation of virtual objects. Both of these systems simplify the problem by only augmenting a single, static image. Gibson and Murta [3] made the same simplification, but approached the rendering problem with a variety of different techniques to achieve framerates of up to fifteen frames per second. However, since they do not calculate a full global illumination solution, the system only simulates a limited set of light transport mechanisms. Agusanto et al. [1] focuses on lighting dynamic virtual objects with static real world illumination, using prepared light probe data and image based lighting techniques. Their work allows objects to move, deform, and change materials, but changes in the lighting environment will not be reflected.

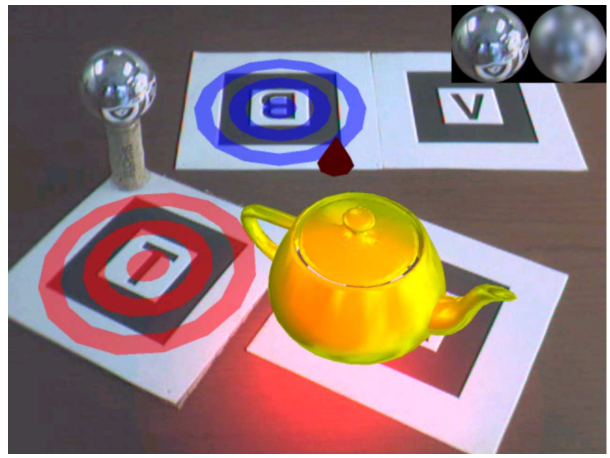

Figure 1. An early prototype system, with real world illumination of virtual objects, and virtual illumination added into the physical scene.

The goal of our research is to create an augmented reality system that builds on the results of those previously mentioned, but especially focuses on the requirement of responding to changes in the physical environment in real-time. The system should be usable in a new environment with little to no preparation or instrumentation necessary, so it is easy to start using immediately. Once the system is running, changes in the lighting environment or motion of physical objects should not require recalibration or careful measurement to maintain accuracy when rendering. Real-time rendering of realistic virtual objects, achieving seamless integration with the physical scene, is still a top priority. The results of an early prototype utilizing real-world illumination can be seen in Figure 1. The long term goal is to achieve fully interactive, indistinguishable physical and virtual worlds.

To better understand what is required for such a system, first assume that data acquisition is not an issue - that is, what problems remain when all dynamic geometric and lighting information is known for an environment? The problem becomes a rendering issue, for which interactive global illumination algorithms have already been extensively researched [8, 10]. In the case of a static scene, complete knowledge has already been achieved by systems such as [7] and [3], via extensive measurement and calibration as a preparation step. No system has been created with full knowledge of a dynamic physical environment, though one could be constructed with extensive instrumentation (affixing trackers to rigid bodies, monitor light probes, etc.).

The difficult work of measuring and calibrating a static environment, or heavily instrumenting a dynamic environment, is a serious impediment to the ease of use of such augmented reality systems. Instead, we focus on lightweight, automatic or semi-automatic acquisition techniques. The necessary information includes the local lighting conditions, to properly shade virtual objects, and the scene geometry, to add virtual light and shadows to physical objects. Fully automatic techniques to acquire this data would be ideal, though some small amount of user interaction is acceptable as long as it remains simple.

The particular rendering techniques used depend heavily on how the scene data is acquired. For example, if environment light sources are carefully measured, they can be modeled as point, area, or directional light sources with a traditional rendering technique. On the other hand, acquired light probe data can be converted into a light field [6] or irradiance volume [4] with a significant amount of processing, or it can be filtered into a material reflectance map and applied directly to geometry [9]. The geometry acquisition technique may yield a point cloud, polygon sets, or voxel data, which influences the internal data representation and choice of rendering algorithm. Also, limitations of the geometry acquisition technique may cause certain types of errors such as noisy edges or misregistration. To combat these errors, we can use the video of the physical scene to improve results by tweaking geometry to match image features.

2 Planned Contributions

The focus of this work will be on automatic acquisition and maintenance of dynamic scene information, as well as a unified data structure and rendering techniques that produce high visual fidelity results in real-time from the available data. The planned contributions are

- a means of acquiring and maintaining dynamic light environment data,
- improvement in rendering techniques for image-based lighting, specifically shadow generation from environment maps,
- combining automatic geometry acquisition techniques for more accurate results,
- and an image-based technique for correcting geometry errors in rendering.

For a unified rendering solution, an aggregated data structure will be necessary to maintain the dynamic lighting and geometry information. An overview of how these contributions are related can be seen in Figure 2.

For lighting of virtual objects, we will focus on environment mapping and image-based lighting techniques, for their speed and demonstrably realistic results. The limitations of environment mapping are well understood, and some techniques are available for alleviating them via filtering, warping, and blending multiple maps. An AR system as we have described will require an easy to maintain structure for storing lighting data to use for environment mapping. A regular sampling of data could be acquired with an omnidirectional camera mounted on the user, though the equipment would diminish the user experience. Alternately, a sparse sampling of light may be sufficient, as users could provide light information for any particular region where lighting is visibly inaccurate. For example, a user could use a tracked light probe affixed somewhere in the scene, or manually sweep it through problem areas when necessary. Another possibility rather than storing a sparse structure of lighting values, is to combine the lighting information with known scene geometry, and project the light probe image onto the geometry, storing lighting values as textures. Then, the environment map for any point in the scene could be generated on the fly by reprojecting the scene onto a sphere. This is especially appealing, as it stores lighting information only at the source polygons, rather than at points within the scene volume. The difficulty of these techniques will be maintaining a data structure with a small incremental cost, to handle dynamic updates in real-time.

For realistic image-based lighting with environment maps, filtering to create material reflectance maps is necessary. Real-time frequency space techniques have been developed for exactly this purpose [9], which will allow virtual objects of arbitrary material parameters to be shaded. However, the generation of shadows from environment map data is currently an open problem. One important piece of information for shadow generation is depth, which environment maps do not contain. However, since we have a model of scene geometry, we can compute an environment map with an associated depth map, providing the necessary information to create high visual fidelity soft shadows. We plan to adapt existing real-time soft shadow techniques [5] to use environment maps for lighting information, as well as decomposing environment maps into a small set of point and area light sources by segmenting out the brightest regions. These patches will then be used with traditional soft-shadow algorithms. Alternately, we will investigate the direct use of environment maps with image processing techniques for shadow generation. The procedure here will be to prototype implementations of different algorithms

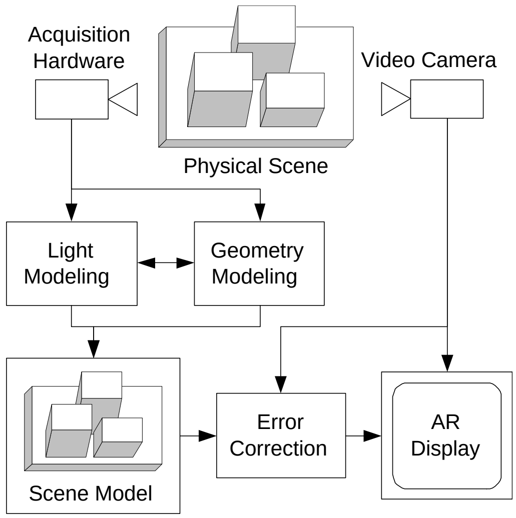

Figure 2. The main components of the planned contributions. Acquisition hardware will feed input to the light model and geometry model, which will be aggregated into a unified scene model. The error correction will use a video signal to tweak rendering for final display.

and compare the visual fidelity of the results, to see what artifacts can be dealt with efficiently. For instance, extracting individual area light sources may be adequate for environments with low ambient illumination, but might fail in bright scenes. Once the artifacts of each technique are understood, they can be addressed.

To acquire real-time geometry information, established techniques will be examined, such as a stereo camera with a real-time depth correspondence algorithm or vision-based shape from shading or motion. We will also consider semi-automatic techniques that require some minimal amount of user input. One possibility would be a small widget the user can move around in the environment to sweep out volumes and define planes. Using standard tracking technology, the user can quickly define a rough model of the scene without the need for a complex 3D modeling program or detailed measurements. Such a technique would still need to be combined with some fully automatic acquisition method such as stereo vision to handle fully dynamic geometry. The unification of selected acquisition methods into a final scene model will be a major component of the aggregated data structure. Geometry acquisition also raises the issue of noise in the acquired model, which will adversely affect the quality of the final results. Stereo vision often yields noisy depth boundaries, while a user-swept widget will be unable to resolve shapes below a certain size threshold.

When rendering the contribution of virtual light sources or virtual shadows cast onto the physical scene, the scene model's accuracy becomes extremely important. Edge discontinuities generate sharp features in lighting or shadows which will not line up properly if the model is inaccurate or misregistered. Unfortunately, fully automatic real-time techniques for geometry reconstruction are not accurate enough to avoid these sorts of errors. The acquired video is a useful source of information for error correction, as it has intensity edges which mostly correspond to geometry edges in the scene model. Misregistration can then be partially corrected for by subtly adjusting the rendered lighting in image-space to line up with the video's edges. This could be accomplished by computing an image warp function that maps projected scene model edges to video edges via a nearest neighbor search. Rendering with the applied warp in real-time is then a simple matter with new programmable graphics hardware.

The overall procedure is to prototype the desired acquisition techniques and evaluate exactly what artifacts have the most significant effect on perceived visual fidelity. The various approaches for addressing them can be implemented and evaluated at that point. An aggregated scene model data structure will be developed concurrently to support the various input devices. We will evaluate our system in a variety of applications including real-world simulation training and desktop augmented reality.

References

[1] K. Agusanto, L. Li, Z. Chuangui, and N. W. Sing. Photorealistic rendering for augmented reality using environment illumination. In *ISMAR*, October 2003.

[2] A. Fournier, A. S. Gunawan, and C. Romanzin. Common illumination between real and computer generated scenes. In *Proceedings of Graphics Interface*, pages 254–262, 1993.

[3] S. Gibson and A. Murta. Interactive rendering with real-world illumination. In *Proceedings of 11th Eurographics Workshop on Rendering*, pages 365–376, 2000.

[4] G. Greger, P. Shirley, P. M. Hubbard, and D. P. Greenberg. The irradiance volume. *IEEE Computer Graphics Applications*, 18(2):32–43, 1998.

[5] J.-M. Hasenfratz, M. Lapierre, N. Holzschuch, and F. Sillion. A survey of real-time soft shadows algorithms. *Eurographics State of the Art Report*, 22:753–774, 2003.

[6] M. Levoy and P. Hanrahan. Light fi eld rendering. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42, 1996.

[7] C. Loscos, G. Drettakis, and L. Robert. Interactive virtual relighting of real scenes. *IEEE Transactions on Visualization and Computer Graphics*, 6(4):289–305, 2000.

[8] M. Nijasure, S. Pattanaik, and V. Goel. Real-time global illumination on the GPU. *Journal of Graphics Tools*, 2004. To appear.

[9] R. Ramamoorthi and P. Hanrahan. Frequency space environment map rendering. In *Proceedings of the 29th Conference on Computer Graphics and Interactive Techniques (SIGGRAPH-02)*, volume 21, 3, pages 517–526, July 21–25 2002.

[10] I. Wald, T. Purcell, J. Schmittler, C. Benthin, and P. Slusallek. Realtime ray tracing and its use for interactive global illumination. In *State of the Art Reports, Eurographics 2003*, 2003.

Fluid Beam - A Steerable Projector and Camera Unit

Lübomira Spassova
Department of Computer Science
Saarland University
Saarbrücken, Germany
mira@cs.uni-sb.de

Abstract

This paper presents an approach of using a steerable projector and camera unit in order to augment instrumented environments with projected virtual objects. The environment appears to be covered with a kind of invisible virtual layer on which images, texts, videos, computer desktops and other virtual objects can be placed. By steering the projector beam on particular surfaces the objects placed on them are made visible as if the projector were a virtual torch light.

1. Introduction

With the ongoing development of ubiquitous computing the need of omnipresent displays is rising. Although due to the technological progress physical screens are becoming larger and more affordable it is still not possible to cover a whole room with such devices. Therefore projection is increasingly regarded as a new means of ubiquitous visual output.
In the last few years different efforts have been made to develop projectors that display images free of distortion on arbitrary surfaces. A homography-based method using a non-calibrated remote camera has been presented in [5]. It has mainly been developed for slide presentations with stationary projectors and its significant disadvantage is the need for the camera to be able to detect the screen boundaries.
The Everywhere Displays (ED) Projector [3] enables projection in arbitrary directions within a cone using a rotating mirror placed in front of the projector. Image deformation is avoided by projecting the source image as seen from the point of view of a virtual camera with the same orientation and optical parameters as the projector. In this way the distortion caused by oblique projection is compensated. Unfortunately due to the mirror-projector configuration the projection range is restricted to a cone so that not every surface can be reached.

Figure 1. Fluid Beam hardware

This drawback can be relieved by placing the projector in a movable unit like the one described in [1]. This device enables projection in almost any direction. In this paper a method for distortion-free projection will be described that combines the approaches of [3] and [1] in the sense that we have a steerable projector and camera unit similar to the one in [1] which can project on arbitrary planar surfaces correcting for the distortion using the virtual camera method presented in [3].

2. Hardware

The device consists of a 3.3000 ANSI lumen projector and a high resolution digital camera placed in a moving yoke produced by a stage equipment manufacturer, controlled via a USB/DMX interface, and mounted in the center of the ceiling of the instrumented environment (see figure 1). The brightness of the projector is important for the images to be seen even in daylight. The camera can also deliver a low resolution video stream. The steerable unit can be rotated both horizontally and vertically.

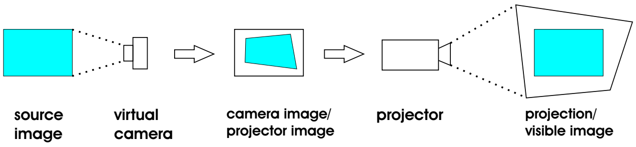

Figure 2. Correction for image distortion by means of a virtual camera

3. Creating a display continuum

In order to correct for the distortion due to oblique projection we create a virtual model of the instrumented environment in Java3D. A virtual camera is placed in this model at a position corresponding to the posture of the projector in the environment. Virtual objects like images, videos or live video streams can now be put as textures on arbitrary surfaces in the model, and by moving the projector and the virtual camera synchronously the objects appear at the corresponding surfaces in the instrumented environment (see figure 2). In this way we obtain a virtual layer on which virtual objects can be located at certain positions or even moved from one position to another.
In order to achieve high spatial accuracy both the position and the optical parameters of the virtual camera must be calibrated very precisely. Currently we are working on methods to calibrate the exact position of the projector in the instrumented room and to compute the precise transformation between projector and camera coordinates for the Fluid Beam device.

4. Current applications

4.1. SearchLight

The Fluid Beam device has been integrated into the SearchLight [2] application which implements a physical search function for an instrumented environment. It uses the movable unit of the Fluid Beam to scan the environment for optical markers and to highlight searched objects within a projected spot (see figure 3). This functionality of locating physical objects is analogous to file search on a common PC, and helps blur the distinction of physical world and virtual data.

4.2. Combination with wipe gesture recognition

One application that was shown at the international trade fair CeBit this year combines the Fluid Beam functionality with simple wipe gesture recognition. In this scenario the user can select images by dragging them from the PC

Figure 3. SearchLight

screen to a projection surface on a desk. Then by performing a simple gesture the user can "wipe" the images from the desk to a big projection surface on the wall, where a slide show of the selection is shown (see figure 4). For the wipe recognition we use the video stream delivered by the camera of the movable unit and compute the pixel differences between two successive frames. Thus we can detect the direction of the hand movement.

4.3. Combination with spatial audio

In our instrumented environment a spatial audio system is integrated, with which virtual soundsources can be created to appear at particular locations in the room. Combining this system with Fluid Beam enables virtual objects to be associated with their own sounds that can move together with their visual appearance. This increases the impression that the user is working with real objects.

4.4. Moving virtual character

Currently we are using Fluid Beam to display a virtual room inhabitant driven by a presentation planner, that plays the role of a virtual assistant for the instrumented environment. The implementation of the virtual room inhabitant combines a character engine with the spatial audio system described above and the Fluid Beam application. In this way the virtual character can "move" freely across the whole instrumented room, while at the same time spatial audio is generated which seems to come from the position of the character.

5. Future work

The central goal of my Ph.D. work will be to find and explore novel interaction options for the Fluid Beam, thus upgrading it to an Aware Projector [4].

Figure 4. Wipe gesture recognition scenario: a selection of two images is displayed on the desk; the white surface on the wall is used for the slide show

5.1. A system for self-calibration of the camera-projector unit

Using methods similar to those mentioned in [5] it is possible to automatically locate potential projection surfaces. By projecting a certain (rectangular) pattern on a surface, probably not perpendicular to the projection beam, one receives a distorted image. This distortion can be observed by a camera with sufficient distance to the projector, and analyzing it the Fluid Beam will not only be able to locate projection surfaces but also to detect its own position in the environment. Thus the user will no longer need to specify a virtual model of the environment but this will be done fully automatically by the Fluid Beam itself.

5.2. Different interaction options like gesture or speech recognition

As in the area of ubiquitous computing the use of keyboard and mouse will play a less prominent role, the user should be able to steer the Fluid Beam by gestures or speech. One interesting research topic deals with multimodal interaction. One can imagine a situation in which the user is pointing to a certain wall in the room saying “Create a display!”, and after analyzing his gesture and the sentence the Fluid Beam will display the user’s desktop at the specified surface.

5.3. Combining several steerable projectors

If the instrumented environment is very large (e.g. warehouses, supermarkets, etc.) a combination of several projectors will enable the creation of larger display surfaces and the compensation of shadows on them. It will also allow passing on virtual objects from one projector to another in order to move them over large distances. In this way - if the user is tracked - he can be guided through the environment.

5.4. Interaction between stationary displays and steerable projection

In our instrumented environment there is a large plasma touch screen and several smaller screens. The interaction between these sceens, providing islands of higher resolution, and a projected image is also an important research topic. Potentially the projected display continuum can be used to transfer virtual objects from one screen to another.

5.5. Performing user studies

Users will be invited to work with the Fluid Beam in order to evaluate its usability. I can imagine a scenario in which a projected character appears as an assistant moving through the environment helping the user by giving him advice, explaining the handling of the devices in the room, answering questions, etc. The user will be able to interact with the character applying speech and gesture, thus learning how to work with the instrumented environment.

References

[1] S. Borkowski, O. Riff, and J. L. Crowley. Projecting rectified images in an augmented environment. *ProCams Workshop, IEEE Computer Society Press*, 2003.

[2] A. Butz, M. Schneider, and L. Spassova. SearchLight: A lightweight search function for pervasive environments. *Proceedings of International Conference on Pervasive Computing*, 2004.

[3] C. Pinhanez. The everywhere displays projector: A device to create ubiquitous graphical interfaces. *Lecture Notes in Computer Science*, 2001.

[4] R. Raskar. Invited talk for PROCAMS-2003. *http://www.procams.org/procams-raskar.html*.

[5] R. Sukthankar, R. G. Stockton, and M. D. Mullin. Smarter presentations: Exploiting homography in camera-projector systems. *Proceedings of International Conference on Computer Vision*, 2001.

Wearable Speaker Separation and Tracking System

U. Anliker, J. Randall and G. Tröster
Swiss Federal Institute of Technology
Wearable Computing Lab
{uanliker, jr, troester}@ife.ee.ethz.ch

Abstract

Humans can differentiate one speaker from many. However automating this faculty is recognized as being non trivial. We present a system to track individual speakers during simultaneous discussions. We propose to split the problem into three tasks: audio source separation, location tracking, and speaker tracking and identification. For each task the state of the art and performance parameters are provided in this paper. We outline possible hardware and algorithms. We present different experiments, that could be used to validate the capabilities of our system.

1 Introduction

Humans have the ability to focus on person speaking among many simultaneous audio sources. The goal of our work is to copy this ability to track speakers and to follow discussions. We think such a system should to perform three tasks: audio source separation, location tracking, and speaker tracking and identification.

The scope of this paper is wearable speaker separation and tracking. We give a short overview of the state of the art in section 2. Section 3 introduces performance parameters for our system and provides qualitative goals. A hardware and algorithm outline followed by planned experiments are presented in section 4.

2 State of the Art

2.1 Source Separation

When the number of sources is greater than the number of sensors the problem is degenerate and traditional matrix inversion demixing cannot be applied. A wearable system favors few microphones, thus a degenerate situation is likely. The linear mixing matrix is non-square, and it is not possible to linearly invert the system. The basic approach to the problem of degenerate demixing is (1) determine an appropriate disjoint representation of the sources and (2) determine the partitions in this representation which demix. The performance of source separation algorithm depends strongly on noise and reverberation levels. Relatively few algorithms presented in the literature can be used with more sources then sensors.

For example, the following two algorithms assume that at every time-frequency instance only one source is active, i.e. temporal and frequency diversity of the source signals and that we have convolutive mixing, i.e. sources have different amplification and time delays in different mixtures. S. Ikdea [5] estimates the mixing matrix in the time-frequency domain based on a decorrelation algorithm. S. Rickard [9] calculates two mixing parameter for each time-frequency point. Points from one source will have similar parameter and will cluster around the true mixing parameter. Each clusters is then assumed to represent one source location.

2.2 Speaker Tracking

Several approaches have been presented for online unsupervised speaker change detection and speaker tracking for broadcasting and audio archives. The main drawback of published speaker change detection algorithm is that they need sufficient data to train a speaker model. Sufficient in this case is utterances of several seconds.

The algorithms presented in the literature use either a metric decision, the difference between two features sets have to be above a threshold or an information theoretic criterium, e.g. Bayesian Information Criterion (BIC) to decide if one or two model parameter sets describe two adjacent segments.

2.3 Communication

We are not aware of any Speaker Tracking system, that uses communication to improve performance. Many wearable systems have proposed to use communication between different users, but this has been rarely implemented in software, e.g. [3]. On-body communication and sensor data fusion has been used by different research groups, e.g. [7], and [6].

2.4 System

T. Choudhury and S. Basu at the MIT Media Lab. have developed the *Sociometer* [4] to sense and model human communication networks. They ignored segments when overlapping sound occurred. The system includes a microphone and IR transmitter/receiver. Face-to-Face communication is detected by the IR module and by investigations on the time patterns of the audio stream. The data is analyzed off-line for each device, no data exchange takes place to improve the situation awareness.

Ajerma et al. [2] addressed the speaker change detection problem during meetings. Short segments, many speaker changes and overlaps occur. They use location based features and Mel-Frequency Cepstrum Coefficients (MFCC) to detect speaker changes, and to identify speakers. Their recording hardware consists of a circular 8-microphone array and of 4 lapel microphones on a table.

3 Hypotheses

The envisioned system should achieve the following wearable speaker tracking benchmarks for the three proposed tasks.

3.1 Source Separation

The performance of source separation can be measured from the accuracy of the direction of the separated sources, the minimum separation distance between two sources, and the maximal number of sources that can be separated. We think that a separation system should have a time difference of arrival (TDOA) accuracy of one sample and to be able to separate two speaker having an TDOA difference of two samples or more seen by the user. Situations with more then four simultaneous active sources are rare. Thus a separation algorithm should be able to separate up to four sources. The speaker tracking steps needs clean speaker signals. Percentage of recovered signal energy and cross-talking level are possible parameters for the cleanness. A separation system should be able to extract 80% of the original signal energy and to have a signal to noise ratio (SNR) of 10dB for the separated signals (cross-talking would be considered as noise).

If the noise or reverberation level goes up, the separation performance degenerate. A speaker tracking system should be stable for SNRs down to 15dB. We should be able to deal with reverberation which occurs in an office/meeting environment, i.e. maximal reverberation time [1] below one second.

3.2 Position Tracking

In a wearable environment speakers and users can move freely. We consider both speaker movement and user movement as essential information. The tracking system should be able to follow a speaker when the TDOA estimation changes slowly, i.e. by up to 1 sample per second. A rotation of the recording system should be compensated by a complementary sensor. We anticipate rotation speeds up to 20 degree per second, the system must be able to handle this.

During a discussion, persons can move together down a corridor, not changing there relative position. The system should be able to tracking speakers while walking side by side.

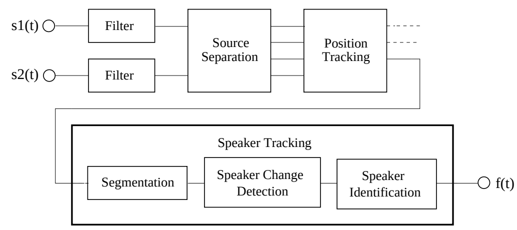

Figure 1. Data Flow

3.3 Speaker Tracking

Key benchmarks of an audio speaker tracking system are insertion (detection of not-existant speaker changes), deletion (missed speaker change) and wrong identification of speakers. For experimental purposes we propose to summarize these three as detection of number of speaker changes and detection of duration of utterance. The literature shows, that it is difficult to deal with short utterances (<2s). This problem is reflected by the number of changes. Longer speech segments can be detected better which is reflected by the duration. We think a speaker tracking system should be able to detect 80% of the speaker changes and 90% of the duration. If we take the location information into account, it should be possible to get 90% of the changes and 95% of the duration. If we consider information by other users, which might have a better position for speaker separation the performance might reach 100% of both measures.

An other interesting point to investigate is the influence of online processing. An off-line system could use an iterative approach, e.g. refinement of the speaker model. We expect a performance of 80% of the speaker changes and 90% of the duration using only the audio tracks.

4 Proposed Approach

In a wearable environment as few microphone as possible are preferred. Two mono omnidirectional lapel microphones is the lower bound. As a complementary sensor for rotation compensation an XSens module placed between the two microphones could be used. The module contains magnetic field, gyroscope and acceleration sensors to provide absolute orientation.

4.1 Algorithm Outline

Our envisioned system performs three tasks, see Fig. 1. For the first task, the different audio sources needs to be separated. For this task DUET, a blind source separation (BSS) algorithm introduced by S. Rickard [9], could be used. The algorithm functions for two or more mono microphones. The sources are separated by the delay and by the damping between a microphone pair. The performance is comparable to other established BSS/Independ Component Analysis (ICA) methods [8]. The DUET algorithm estimates directly the source separation parameter, no iterative optimization such as gradient search is needed. The algorithm is thus suitable for an online separation system.

In the second task, the system needs to track source position. The tracking algorithm might be based on microphone orientation and the DoA of a sound source based on a TDOA estimation. It seams reasonable to assume that positions with a small TDOA difference (below one sample) belong to the same audio source. The audio data received from one direction could then be merged to one audio stream to get larger audio segments.

In the third task, different speakers are tracked. The task could be split into three subtask. First, the stream would be split into voice and non-voice segments. Second, for each voice segments a speaker model would be trained. Based on the model speaker changes in the audio stream would detected. Third, the speaker model would be used to identify different speakers. For each identified speaker, information about a conversation is recorded; for example time, numbers and duration of utterances.

Communication between devices should increase the estimation accuracy and provide additional information. Source position estimations, time pattern and speaker models are interesting information. To exchange information a low bandwidth communication channel should be sufficient. Blue tooth or a simple RF channel could therefore be used. A person not speaking, cannot be detected from the audio stream itself. Communication by the wearable device of a listener would add him to the participants list. Face-to-Face discussion could be identified when two microphone pairs are pointing together and the audio streams alternate most of the time or by matching audio streams.

4.2 Experiments

For each hypothesis we need to set up an experiment to measure the performance of our envisioned system. The experiments are grouped by the three processing tasks.

4.2.1 Source Separation

The separation algorithm should be tested with different input signal qualities. A first test would include anechoic mixing of a increasing number of sources and/or smaller TDOA difference. The input data could be taken form a audio data base and the mixing would be simulated. The same testing should be done in a meeting and office room with real individuals. We think the noise robustness testing should be done with the data from the last experiments. The noise is then added in a simulation environment.

4.2.2 Position Tracking

The tracking capability would be tested via two approaches. First, in the simulation environment by changing the mixing parameter as function of time. Second, we plan do to recordings in our office with moving speakers and/or users.

4.2.3 Speaker Tracking

We propose to first test the speaker tracking system independently of the two other processing steps. The algorithms would be tested with labelled audio streams. The proposed performance parameter, accuracy of number and of duration of the speaker utterances, would be calculated for different noise levels.

The last planed step in the single system testing, is to apply the proposed algorithms to group meeting recordings. If the system performance requirement are satisfied, we would investigate the performance gain of a communication channel.

5 Conclusion

We have presented a system that should track an individual speaker from within a noisy environment. The envisioned system would perform three tasks: audio source separation, location tracking, and speaker tracking and identification. This paper gave a short overview of the state of the art for each task. We have introduced performance parameters for our system and have given qualitative goals. We have proposed hardware and outlined algorithms; we have shown which experiments would be needed to validate the performance capabilities of our system. We conclude that by splitting the job into three task, considering the state of the art of each task, it should be possible to satisfactorily solve the tracking problem. By the end of this year, we intend to have simulations validated by experiments.

References

[1] Reverberation time. http://hyperphysics.phy-astr.gsu.edu/hbase/acoustic/revtim.html.

[2] J. Ajmera, G. Lathoud, and I. McCowan. Clustering and segmenting speakers and there locations in meetings. Research Report IDIAP-RR 03-55, IDIAP, Dec 2003.

[3] S. P. and. I. Locher and M. Srivastava. Design of a wearable sensor badge for smart kindergarten. In *ISWC*, pages 7–10, Oct 2002.

[4] T. Choudhury and A. Pentland. The sociometer: A wearable device for understanding human networks. In *Conference on Computer Supported Cooperative Work*, November 2002.

[5] S. Ikeda and N. Murata. An approach to blind source separation of speech signals. In *ICANN'98*, 1998.

[6] E. Jovanov, A. O. Lords, D. Raskovic, P. Cox, R. Adhami, and F. Andrasik. Stress monitoring using distributed wireless intelligent sensor system. *EMB Magazine*, pages 49–55, May/June 2003.

[7] P. Lukowicz, H. Junker, M. Stäger, T. von Büren, and G. Tröster. WearNET: A distributed multi-sensor system for context aware wearables. In *UbiComp*, 2002.

[8] S. Rickard, R. Balan, and J. Rosca. Blind source separation based on space-time-frequency diversity. In *ICA*, 2003.

[9] O. Yilmaz and S. Rickard. Blind separation of speech mixtures via time-frequency masking. *IEEE Transactions on Signal Processing*, 52(7):1830–1847, Jul 2004.

Input, Control, and Application Design for a Useable Wearable Computer

Brian Mila
Virtual Reality Application Center
Iowa State University
bmila@vrac.iastate.edu

Abstract

Although wearable computers have been around for over a decade now we can see that their use is not pervasive. We cannot walk down the street and readily see somebody using a wearable computer but we can see any number of people using or carrying a laptop. Even though wearables cost about the same as laptops and provide the same functionality they have not enjoyed the same growth in popularity that laptops have. Perhaps it is because wearables have nothing more to add above and beyond what a user can get on a laptop. In fact, wearable computers rarely have attached CD drives or floppy drives and almost always have a smaller, lower resolution display. In addition, the input and control of a wearable computer is usually much more difficult than a laptop computer. Whereas a laptop computer has a built-in keyboard and mouse a wearable computer does not. Even though the idea of carrying a small, lighter weight computer initially sounds promising there are several factors which have stymied its mainstream acceptance.

Looking at the PDA we can see that users are willing to sacrifice processing power for a smaller form factor. Even though the PDA is several generations behind a laptop computer in terms of processing power it still has a rich suite of applications available to a user. In addition, the PDA provides several different modes of input, all of which are easily learned by a novice user within a few minutes of tinkering. A wearable computer's input on the other hand can be difficult to learn or physically obtrusive to the user [1,5]. Furthermore, wearable computers do not provide a rich suite of applications above and beyond what one could get with either a PDA or a laptop. While many researchers have attacked this problem, many solutions rely on the availability of special hardware i.e. sensors, tracking units, or other specialized hardware not readily available to the end-user [2,4]. Also many solutions expect that the world in which they are operated in is also instrumented with sensors or electronic tags [8]. These are good forward-looking approaches, but are not well suited to the here and now. If wearable computers are to survive the transition from research and development to mainstream use they must provide a function or service that is not available on a laptop or a PDA and also must be easily controlled by a novice user with no experience. In short, the two major hurdles are: input & control of the wearable, and application functionality and useability. Both of these problems are crucial and must be solved before widespread acceptance and usage of wearables can occur. In my current research I am attempting to solve both of these problems.

I propose to use an off-the-shelf PDA to serve as my input & control device. This choice was made for various reasons: current input devices like chorded keyboards are difficult to learn [1]. Others only exist as prototypes [4,5]. Also, most of these devices are wired. This means that there is another wire coming out of the pack the wearable is being carried in. The PDA solves these problems because it is a) common enough that most people are familiar with how to input text with it. b) wireless. c) easily obtained. Coupled with a wireless card in the wearable, both devices can always communicate with each other, either by using an existing wireless network if it is available, or by forming an ad-hoc network if not. Also, the

PDA can be used as a control for the pointer (mouse) as well. The user can move the stylus on the PDA, these coordinates can then be scaled and sent to the wearable. The result is an interaction that is very similar to a touchpad, again preserving the familiarity that a user would have with such a device. Finally, the PDA can also serve as a secondary display device. An individual screenshot can be taken and then sent to the PDA, or through VNC, the entire desktop can be viewed and controlled interactively.

With the input and control defined let me now discuss the “big” problem: applications. So often an application written for a wearable computer is extremely domain-specific. There are applications for airline mechanics, police, and the health care industry [3,6,7]. But there aren't many generalized applications that could be used by anyone (again, without needing specialized hardware). The real problem is this: a computer display is in front of you at all times, what will it be used for? Surfing the web? Possibly, but that requires wifi availability. In reality, you find that as you move about you encounter “pockets” of availability. This means that interactive applications like web browsing or instant messaging are going to be extremely frustrating to use if the user is mobile. What we need is an application that can capitalize on the limited wifi availability by doing bulk downloads of data when the network is available and then displaying this data in a controlled fashion. This would give the user the illusion of constant data availability. Some applications that come to mind are news feeds, stock quotes, email, and calendar appointments. While all of these are available on a PDA or a laptop, those devices require the user to physically pull out the device and turn it on. A wearable computer, on the other hand, can display this information in an unobtrusive fashion in the background, similar to the way a news ticker displays information without distracting the viewer from the main channel content. In some ways the news ticker is the ideal information source because it has a very low demand for attention. Whereas most applications attempt to usurp the user's attention when information is available, the ticker is constantly displaying new information and it is up to the user to decide when to spend attention on that information. I plan to recreate the news ticker idea to have it display information from all of the sources mentioned above. Also, the sources of information can be designated to have certain levels of demand for attention. A possible ordering from highest demand to lowest might be appointments, email, stock quotes, news feeds. While all of the information is displayed in the same medium they can be color coded to indicate their relative levels of importance. Finally, when a user decides that a piece of information is worth looking at it can be selected and viewed either on the wearable display or on the PDA. Reading small amounts of text on the PDA is easier in some ways because text can be difficult to read on a small head-up display. Also, displaying information on the PDA can be useful for showing to more than one person while reading on the head-up display would be useful for private information.

Accessing new content is only half the battle. Another obstacle is finding a way to access any content that we have previously stored. People spend a lot of time acquiring information from various sources and in various formats and its only natural to assume that they would want to access that information from various locations and at various times. However, to get access to the data on a different computer requires knowledge of the other computer's network address and protocols. Furthermore, if a user does not know the exact filename they would have difficulty in finding it even if they had remote access to that computer. To solve this problem my colleagues and I are developing a program we call the Tangle which is capable of indexing a user's entire hard drive including the contents of many popular file formats. The user can then search their computer without needing to know specific filenames. After a particular file is found it can be easily transferred through the Tangle to a remote computer. In addition, the user can add “friends” much in the same way that an IM user would add a “buddy” to his or her list of contacts. Adding a friend in this way establishes a level of trust between the two parties. After this trust has been established, they can search for and download information from each other. Finally, the transport was built on top of a instant messaging architechture which means that the user never needs to know their IP address or anyone elses. Through the IM protocols they are notified when friends go online and offline. Because file transfers resume where they left off a user can transfer files without needing constant network connectivity. A incomplete transfer will resume as soon as a new network connection is available. Once the transfer is complete the data has been physically copied to the wearable computer so it can viewed or edited anytime, regardless of network availability.

Bringing all of these ideas together will create a system that supports the user by minimizing the amount of attention needed to use it. It will give wearable users access to a constant stream of new information as well as any digital content that they have already acquired. This will allow them to use a new technology without giving up anything or having to spend valuable time reacquiring information. All of their information is easily accessible and the PDA should allow them to control and display that information with little to no training.

References

1. Lyons, K., Starner, T., Plaisted, D., Fusia, J., Lyons, A., Drew, A., Looney, E. W. “Twidder Typing: One-Handed Chording Text Entry for Mobile Phones”, *Proceedings of the 2004 conference on Human factors in computing systems*, pp. 671-678, 2004.
2. Bauer, M., Heiber, T., Kortuem, G., Segall, Z. “A Collaborative Wearable System with Remote Sensing”, *Proceedings of the 1998 International Symposium on Wearable Computers*, 1998.
3. Haniff, D., Baber, A. “Wearable Computers for the Fire Service and Police Force: Technological and Human Factors”, *Proceedings of the 1999 International Symposium on Wearable Computers*, 1999.
4. Farringdon, J., Moore, A., Tilbury, N., Church, J., Biemond, P. “Wearable Sensor Badge & Sensor Jacket for Context Awareness”, *Proceedings of the 1999 International Symposium on Wearable Computers,* 1999.
5. Perng, J.K., Fisher, B., Hollar, S., Pister, K. “Acceleration Sensing Glove”, *Proceedings of the 1999 International Symposium on Wearable Computers,* 1999.
6. Matsushita, S., Ogawa, K., et. al. “A Wearable Sense of Balance Monitoring System towards Daily Health Care Monitoring”, *Proceedings of the 2003 ISWC*, 2003.
7. Siewiorek, D. “Wearable and Context Aware Computers”, *Workshop on Ubiquitous Computing Environment*, 2001.
8. H. Kato and M. Billinghurst. “Marker tracking and HMD calibration for a video based augmented reality conferencing system”.*Proc. IWAR 1999*, pages 85-94. 1999.