Proceedings of the
Second International Workshop on
Location Awareness for Mixed and Dual Reality

LAMDa'12

In Conjunction with the International Conference on Intelligent User Interfaces
IUI'12

February, 14th 2012
Lisbon, Portugal, Europe

Organizers:

Tim Schwartz, M2CI, Saarland University, Saarbrücken, Germany

Gerrit Kahl, DFKI GmbH, Saarbrücken, Germany

Sally Applin, Centre for Social Anthropoloy and Computing, University of Kent, Canterbury, United Kingdom

Eyal Dim, MIS Department, University of Carmel, Haifa, Israel

Petteri Nurmi, Helsinki Institute for Information Technology, Helsinki, Finland

Teemu Pulkkinen, Helsinki Institute for Information Technology, Helsinki, Finland

A typology of augmented reality applications based on their tracking requirements

Jean-Marie Normand
LUNAM Université
Ecole Centrale de Nantes
jean-marie.normand@ec-nantes.fr

Myriam Servières
LUNAM Université
Ecole Centrale de Nantes
myriam.servieres@ec-nantes.fr

Guillaume Moreau
LUNAM Université
Ecole Centrale de Nantes
guillaume.moreau@ec-nantes.fr

ABSTRACT
In this paper we review existing augmented reality (AR) applications taxonomies and we propose ours, which is based on the number of degrees of freedom required for localizing the user, as well as on visualization type, allowing us to cover location-based services as well as more traditional AR applications. Other rendering modalities are also covered by the same degree-of-freedom system.

Author Keywords
augmented reality, taxonomy, degrees of freedom, multimodal

INTRODUCTION
Virtual Reality (VR) has been developing for more than two decades and resulted in numerous conferences and publications in the field [4]. But even though this research field is still very active, VR has faced from the beginning the problem of its relation to the real world which has been formalized by Milgram [9] in the reality-virtuality continuum.

The concepts, functionalities, usages and technologies of VR have been classified many times and we will not propose on a new taxonomy of VR applications, instead we will focus on trying to classify Augmented Reality (AR) applications.

AR is based on techniques developed in VR [1] and interacts not only with a virtual world but has a degree of interdependence with the real world. As stated in [5], “augmenting” reality is meaningless in itself. However, this term makes sense as soon as we refocus on the human being and on his perception of the world. Reality can not be increased but its perceptions can be. We will however keep the term of Augmented Reality even if we understand it as an “increased perception of reality”.

With the term “Mixed Reality”, Milgram [9] groups both AR and Augmented Virtuality (AV). The main difference is that AR implies being immersed in reality and handling or interacting with some virtual “objects”, while AV implies being primarily immersed in a virtual world increased by reality where the user mainly manipulates virtual objects. Nevertheless, the boundary between the two remains tenuous and will depend on applications and usages.

The main challenges of AR consist of the introduction of artificial objects at a location specified in real world coordinates. This requires determining the location of the AR interface in the real world (and not only the user position with respect to the interface as in VR) and including artificial objects in the field of view of the observer.

In the rest of this paper, we will give an overview of existing AR taxonomies, discuss their limitations and propose our own typology.

EXISTING AUGMENTED REALITY TAXONOMIES
Existing taxonomies differ in the criteria they use to classify applications. We chose to divide them into technique-centered, user-centered, information-centered and target of the augmentation taxonomies even if all of existing work will not fit into these four categories.

Technique-centered taxonomies
In [9] the authors propose a technical taxonomy of Mixed Reality techniques by distinguishing the types of visual displays used. They propose three main criteria for the classification: Extent of World Knowledge (EWK), Reproduction Fidelity (RF) and Extent of Presence Metaphor (EPM). EWK represents the amount of information that a MR system knows about the environment. The RF criterion represents the quality with which the virtual environment (in case of AV) or objects (in case of AR) are displayed ranging from wireframe object on a monoscopic display to real-time 3D high fidelity, photo-realistic objects. Finally, the EPM criterion evaluates the extent to which the user feels present within the scene.

In [8], the Reality-Virtuality continuum and some of the elements presented in [9] lay the groundwork for a global taxonomy of mixed reality display integration. The classification is based on three axis: the reality-virtuality continuum, the centricity of the type of display used (egocentric or exocentric) and the congruency of the control-display mapping.

Based on the proposal of a general architecture of an augmented reality system presented in [13], Braz and Pereira [3] developed a web based platform called TARCAST which aimed at listing and characterizing AR systems. It does not propose actual criteria but offers a long list of features for each system, hence is not really discriminative. TARCAST does not seem to be maintained anymore.

Submitted to LAMDa workshop

The technique-centered taxonomies presented here do not take into account any of the mobile AR techniques commonly used nowadays. Milgram's work was innovative at the time it was published and the authors could not predict how mobile AR would arise. Besides, we believe that presence cannot exactly be a common discriminative criterion as it does not refer to the same concept in virtual and real worlds.

User-centered taxonomies

Lindeman and Noma [6] propose to classify AR applications based on where the mixing of the real world and the computer-generated stimuli takes place. They integrate not only the visual sense but all others as well, since their "axis of mixing location" is a continuum that ranges from the physical environment to the human brain. They describe two pathways followed by a real world stimulus on its way to the user: a direct and a mediated one. In the direct case, a real world stimulus interacts through the real environment before reaching a sensory subsystem where it is translated into nerve impulses and finally transmitted to the brain. Those places are called "mixing points". In the case of AR applications, some computer graphics elements can be inserted into this path in order to combine the real world and the computer generated elements into one AR stimulus on its way to the brain. In the mediated case, the real world stimulus travels through the environment, but instead of being sensed by the user, it is captured by a sensing device (e.g. camera, microphone, etc.). Then, the stimulus might be post-processed before being merged with computer generated elements and then displayed to the user at one of the mixing point through appropriate hardware (depending on the sense being stimulated).

Wang and Dunston [14] propose an AR taxonomy based on the groupware concept. They define groupware as: computer-based systems that support groups of people engaged in a common task (or goal) and that provide an interface to a shared environment. The goal of groupware is to assist a team of individuals in communicating, collaborating and coordinating their activities. Based on generic groupware concepts, they isolated three main factors for classifying AR systems for construction use: mobility, number of users and space.

Hugues et al. [5] propose a functional taxonomy for AR environments based on the nature of the augmented perception of reality offered by the applications and on the artificiality of the environment. The authors divide augmented perception into five sub-functionalities: augmented documentation, reality with augmented perception or understanding, perceptual association of the real and virtual, behavioural association of the real and virtual, substitution of the real by the virtual or vice versa. The functionality to create an artificial environment is subdivided into three main sub-functionalities: imagine the reality as it could be in the future, imagine the reality as it was in the past and finally, imagine an impossible reality.

While the first axis of the taxonomy proposed by Hugues et al. covers most of the goals of AR applications, the second axis based on the creation of an artificial environment is less convincing since it does not take into account any alteration of the "present" reality. Moreover their taxonomy is limited to vision based approaches and does not handle other modalities. The groupware taxonomy of Wang and Dunston only takes into account collaborative AR and limits itself to construction-based AR applications. Finally, Lindeman and Noma propose an interesting taxonomy based on the integration of the virtual stimuli within multi-modal AR applications. Nevertheless, their proposal might not be discriminative enough, since very different methods like mobile see-through AR can be classified in the same category as a projector-based AR application. Furthermore, it only deals with each sense individually and does not offer any insight on how to merge them together.

Information-centered taxonomies

In [11], Suomela and Lehikoinen propose a taxonomy for visualizing location-based information, i.e. digital data which has a real-world location (e.g. GPS coordinates) that would help developers choosing the correct approach when designing an application. Their classification is based on two main factors that affect the visualization of location-based data: the environment model used (ranging from 0D to 3D) and the viewpoint used (first person or third person perspective to visualize the data). Based on these two criteria, the authors define a model-view number MV(X,Y) that corresponds to a combination of the environment model (X) and the perspective (Y) used. Each MV(X,Y) class offers different benefits and drawbacks and the authors suggest to choose a class depending on the final application targeted, the available hardware or sensors on the targeted devices.

In [12], Tönnis and Plecher divide the presentation space used in AR applications based on six classes of presentation principles: temporality (i.e. continuous or discrete presentation of information in an AR application), dimensionality (2D, 2.5D or 3D information presentation), registration, frame of reference, referencing (distinction between objects that are directly shown, information about the existence of concealed objects) and mounting (differentiates where a virtual object or information is mounted in the real world, e.g. objects can be hand-mounted, head-mounted, connected to another real object or lying in the world, etc.). This work-in-progress taxonomy is currently being tested with nearly 40 publications taken from ISMAR's recent conferences.

Suomela and Lehikoinen propose a taxonomy that can only be applied to location-based applications, hence is oriented towards mobile AR. Moreover they do not tackle multi-modal mobile AR applications. Tönnis and Plecher propose a complete taxonomy but they do not deal with low dimensionality (e.g. 0D in vision) in AR, nor with the multi-modality that can be used in AR applications.

Taxonomy based on the target of augmentations

Mackay [7] proposed a taxonomy which is neither based on the technology used nor on the functionalities or the application domain. The criterion used to classify AR approaches is rather simple: the target of the augmentation. Three main possibilities are listed in the paper: augment the user, when the user wears or carries a device to obtain information about physical objects; augment the physical object, the object is changed by embedding input, output or computational devices on or within it and augment the environment surrounding the user and the object. In the latter case, neither the user nor the object is affected directly, independent devices provide and collect information from the surrounding environment, displaying information onto objects and capturing information about the user's interactions with them. This taxonomy is not very discriminative. For example, one can notice that every single mobile AR technique falls into the first category, while the last category regroups only projection based methods. As in most of the taxonomies presented here, this work does not tackle the multi-modality issue.

PROPOSAL

We now propose our own taxonomy, based on three axis:

- the first axis is based on the tracking degrees of freedom required by the application and the tracking accuracy that is required. Frequency and latency of tracking can also be taken into account.
- the second axis is representing the application type, whether it is merely visualization/navigation or if it can imply interaction with the observer.
- the third axis covers other rendering modalities that go beyond visual augmented reality. It remains rather limited today but it can be taken into account by the same degrees-of-freedom system.

Tracking

The first axis can be divided into 4 classes:

1. *0D* applications: although it is questionable whether these kind of applications can be considered as AR applications, we find in this class applications that detect a marker (such as a QR-code) and display additional information about this marker. For this category of application, the displayed information has no relation with the real world position and orientation of the marker. Tracking accuracy is very limited since it only requires correct marker detection in one frame, indeed, once detected the marker is not tracked in the following frames. As a consequence of this lack of tracking, latency and update rates are no issues.

2. *2D* applications: this is the class for so-called Location-based services, i.e. applications that provide information about a given location, such as nearby restaurants, etc. Tracking accuracy is generally decametric and the tracking method is often an embedded-GPS (altitude information is not used, updates rates around 1Hz). A typical example of a 2D application is a Google Maps[1] like application which only uses a 2D map in order to help the user finding his way in a city.

3. *2D+θ* applications: this class is also for location-based services that include an orientation information which allows to show a relative direction to the user. All navigation systems are based on this principle, accuracy is most often metric. Note that a GPS alone cannot provide an orientation in static position. Orientation can be computed by differences between positions or can be given by a embedded magnetic compass as in modern smartphones. Required accuracy is also metric, update rates typically ranging from 1 to 10Hz. A typical example of a 2D + θ application is the Metro Paris[2] application which helps you locating nearby metro stations and other points of interests (restaurants, bars, etc.).

4. *6D* applications: this last class covers what is traditionally called augmented reality by computer vision scientists who usually work on tracking technologies. Several types of sensors can be used individually or all together (optical cameras, depth cameras, inertial sensors, etc.). Various precision classes exist depending on application types (e.g. marker-based vs. markerless) and on the working volume size (e.g. indoors vs. outdoors) and accuracy is relative to this size. Update rates are much more critical here, a minimum refresh rate would be around 10Hz, and can go up to 100Hz. At this point, continuous tracking must be distinguished from initial localization for which there exists fewer works.

Application type

For this second axis, we distinguish between application types. The first one is dedicated to (active) observation applications. It includes two main categories depending on the used device:

- Optical see-through applications: there are mostly found in head-up displays (HUD) where they are mostly in the 2D+θ class (for HUDs fixed to a vehicle) or in the 6D class where optical information are projected on lenses of see-through glasses (or for worn HUDs). These applications remain lab prototypes (centimetric accuracy) or can be found in the army (fighter pilots helmet based displays) where they are used to display relative position and speed of opponents as well as some navigational aid.
- Video see-trough applications where a device equipped with a back-located camera (such as a tablet or a smartphone) is filming the real environment and the video is reproduced on its display augmented with artificial images. These applications are often called *magic windows* or "video see-through" [9]. The *magic mirror* is a specific case where the camera and the screen point in the same direction.

At last, Spatially Augmented Reality (SAR) [2, 10] consists in adding information to the real world, not simply adding information onto the observer's eye. These applications have a better potential for being multi-user. They are often large scale applications where the projectors usually do not move.

Rendering modalities

Although the visual sense is by far the most important when talking about AR, some work has been carried out in order to mix the real world and computer graphics images across multiple modalities [6]. While the addition of sound in AR applications seems quite straightforward and common, it is

[1] http://www.google.com/mobile/maps/

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

much more unusual to see AR applications that provide with real 3D sound. Haptic feedback integration for augmented reality is also relatively common, especially for medical or training based applications, although, for mobile AR it is difficult to be able to give the user a better haptic feedback than the one provided by a vibrator (e.g. on a mobile phone). Olfactory and gustatory senses are much more rarely used in AR applications.

Nevertheless, we believe that multi-modality should be taken into account in a typology of AR-based applications, and that their integration could also be based on our degrees-of-freedom approach. Indeed, as for sound, we stipulate that a simple monoscopic sound such as a signal represents 0D sound, stereoscopic accounts for 1D (azimuth) and binaural corresponds to location-based sound (distance and azimuth). Hence, our degrees-of-freedom based classification would take into account the audio modality. But it has to be noted that in the presence of moving objects or user, real-time feedback becomes very complex.

As for the haptic modality, we take a similar approach. A simple vibration, (e.g. provided by a mobile phone vibrator) corresponds to 0D while the use of specific devices could account for higher dimensions of the haptic modality.

Concerning the olfactory and the gustatory modalities, we assume that a non-directional stimulus (or at least a stimulus whose origin cannot be determined such as an ambient smell) is also 0D. As gustatory senses are only touch-based sensors, we limit our typology here for them. If a smell direction can be identified, it is only in azimuth and we call it 1D. Other sensors (thermal sensors on the skin for example) available in the human body could also be classified this way. At the moment, it is technically impossible to directly stimulate proprioceptive sensors, they remain absent from our classification.

As mentioned before, the integration of real multi-modal user feedback requires some extra devices that presently prevent them from being used in most mobile AR applications.

CONCLUSION

In this paper, we have briefly surveyed and discussed existing taxonomies of augmented reality applications. We have then proposed ours based on application tracking requirements, application type and rendering modalities. The originality of our proposal is that it merges location-based taxonomies, such as [11], with classical AR vision-based applications into the same classification, broadening the spectrum of applications fitting into a single taxonomy. Moreover, unlike most existing taxonomies, we included multi-modality as a classification criterion although vision remains by far the most important sense.

During the workshop, we will try to demonstrate how augmented reality applications fit into that classification through a graphical presentation and discuss its compared advantages and drawbacks with respect to other existing taxonomies.

REFERENCES

1. Azuma, R. T. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments 6*, 4 (aug 1997), 355–385.
2. Bimber, O., and Raskar, R. *Spatial Augmented Reality: Merging Real and Virtual Worlds*. A K Peters/CRC Press, July 2005.
3. Braz, J. M., and Pereira, J. a. M. Tarcast: Taxonomy for augmented reality CASTing with web support. *The International Journal of Virtual Reality 7*, 4 (Dec 2008), 47–56.
4. Fuchs, P., Moreau, G., and Guitton, P., Eds. *Virtual reality: concepts and technologies*. CRC Press, 2011.
5. Hugues, O., Fuchs, P., and Nannipieri, O. New augmented reality taxonomy: Technologies and features of augmented environment. In *Handbook of Augmented Reality*, B. Furht, Ed. Springer, 2011, ch. 2, 47–63.
6. Lindeman, R. W., and Noma, H. A classification scheme for multi-sensory augmented reality. In *Proceedings of the 2007 ACM symposium on Virtual reality software and technology*, VRST '07, ACM (New York, NY, USA, 2007), 175–178.
7. Mackay, W. E. Augmented reality: Linking real and virtual worlds: A new paradigm for interacting with computers. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '98, ACM (New York, NY, USA, 1998), 13–21.
8. Milgram, P., and Colquhoun, H. J. A taxonomy of real and virtual world display integration. In *Mixed Reality – Merging Real and Virtual Worlds*, Y. Ohta and H. Tamura, Eds. Ohmsha(Tokyo) and Springer Verlag (Berlin), 1999, ch. 1, 1–16.
9. Milgram, P., and Kishino, F. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information Systems E77-D*, 12 (Dec 1994), 1321–1329.
10. Raskar, R., Welch, G., and Fuchs, H. Spatially augmented reality. In *In First IEEE Workshop on Augmented Reality*, IWAR'98 (1998), 11–20.
11. Suomela, R., and Lehikoinen, J. Taxonomy for visualizing location-based information. *Virtual Reality 8* (September 2004), 71–82.
12. Tönnis, M., and Plecher, D. A. Presentation Principles in Augmented Reality - Classification and Categorization Guidelines. techreport, Technische Universität München, 2011. Version 1.0.
13. Vallino, J. R. *Interactive Augmented Reality*. PhD thesis, Department of Computer Science, University of Rochester, Rochester, NY, USA, 1998.
14. Wang, X., and Dunston, P. S. Groupware concepts for augmented reality mediated human-to-human collaboration. In *Proceedings of the 23rd Joint International Conference on Computing and Decision Making in Civil and Building Engineering* (2006), 1836–1842.

Assessment of Broader Attention-focus Perspective

Eyal Dim
Information Systems Department
The University of Haifa
Mount Carmel, Haifa, 31905, Israel
dimeyal@bezeqint.net

Tsvi Kuflik
Information Systems Department
The University of Haifa
Mount Carmel, Haifa, 31905, Israel
tsvikak@is.haifa.ac.il

ABSTRACT

Mobile personal devices nowadays are equipped with variety of sensors that can be used for context-related measurements such as location, orientation, motion and more. These measurements may allow systems to fuse, reason and abstract context related data into meaningful contextual cues, and use them for personalized services delivery. An example for such a scenario may be two people in front of a product in a shopping center or in front of an exhibit in a museum. They may pay "social attention" to each other, or "object attention" to the exhibit or product. Capturing their attention-focus may enable better adaptation of services to their needs. "Social attention" and "object attention" are broader perspectives in comparison to knowing which specific object attracted the attention, or how exactly a person attracted the momentary attention of the other. This work shows how even a small number of simple sensors combined with a relevant model may enable the assessment of broader attention-focus perspective.

Author Keywords

Ubiquitous computing, context aware computing, social signal processing, user model, group model.

ACM Classification Keywords

H1.2. Models and principles: user/machine systems. H5.2. Information interfaces and presentation: user interfaces.

General Terms

Experimentation, Human Factors, Measurement.

INTRODUCTION

Imagine two people in a shopping center, standing in front of a product. At any given moment, they may be facing the product, or facing each other. In the first case, there is high probability that they are interested in the product. In the second case, there is high probability that they have a social interest in each other. Their general social context is "a group of two people", and their general location context is "being at a shopping center". However, their attention may shift from "socializing" to "interested in a product" or to both. These are broader attention-focus perspectives in comparison to "I listen to you now" or "I see this specific object now". Understanding the focus of attention may enable a system to better adapt its services to these users. For example it may suggest product information when appropriate; avoid interference when people are involved in social attention; or suggest amusement to a person who is not paying attention to the product, enabling the other person to focus on the product. A second example is robotics, where a robot serving a person should adapt to the person's needs, and identify if the person is addressing the robot, or whether it is the right time for the robot to attract the person's attention. A third example is virtual presence, where a real world situation is represented in the virtual world. It would help if the avatars in the virtual world, representing the real world people, would be able to represent cues that reflect the people's attention. (e.g. if a person is facing another person then we, humans, assume that the probability that they are focused on each other increases).

James [1890] presented attention as follows: “Everyone knows what attention is. It is the taking possession by the mind, in a clear and vivid form of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others....”. Technologies to measure attention-focus, or cues for attention-focus, vary in their validity, accuracy, precision and maturity. Examples of such technologies are brain computer interfaces using EEG [Hamadicharef et al. 2009] and methods presented for example by Stiefelhagen et al. [2002], such as: eye-gaze detection, eyeball movement detection; face recognition, head pose and head orientation detection, audio cues and speech recognition. Most of the methods require devices, which either require effort for setup and calibration or are inconvenient to carry or non-portable at all. Therefore, a lot of research is done in a stationary environment (e.g. meeting room, standing in front of a product shelf, sitting in front of a computer). The proposed paradigm is to assess a *broader* attention-focus perspective (e.g. a *narrow* attention focus perspective would be knowing which specific object one is looking at, and a *broad* perspective would be understanding that the focus of attention is a product or a group mate). The broader perspective may enable using simpler measurement sensors, available on nowadays devices such as mobile smart phones. Implementation of such broader perspective requires modeling of the user, the group and the context. Low-level measurements as well as the inferred knowledge may be included in a User and Group Model (UM). The UM data is used to adapt to the user/s needs, as surveyed by Kray & Baus [2001]. The UM uses detailed values stored into its properties. These values are measured and inferred, starting with low-level signals, and continuing with the process of fusion and abstraction of these low-level signals. The UM is usually a general model, yet it is based on details. Organization of such details may enable to use them within low-level building blocks that can serve as the bricks of the UM building. This bottom up approach requires investigation of such building blocks as if they were mini-models. This study presents a mini-model of the attention of two people near an object in a museum. Analyzing such a scenario would require to know: (i) Who is in the group (ii) Where the group is (i.e. a museum) (iii) Are the two people in close proximity to each other? (iv) Is each one of them in close proximity to a specific exhibit? (v) Are they facing the exhibit or each other? (vii) Are they talking? The answer to each of the above questions may involve specific sensor data, fusion of data sources and inference. This work proposes a mini-model, of broad attention-focus, which may be combined with a UM while collecting and fusing data, measured by simple sensors.

BACKGROUND AND RELATED WORK

Llinas [2010] described the standard process of data (or information) fusion from several sources. This process prepares the data by common referencing; associates it by generating hypothesis, assessing them, and selecting the preferred hypothesis; estimates the current state and predicts the future states; and finally, exports the fused data for the use of another user or process. Context awareness is one implementation that requires information and data fusion. Dey and Abowd [2000] defined context as: "any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and application themselves". They further defined context-aware system: "A system is context aware if it uses context to provide relevant information and / or services to the user, where relevancy depends on the user's task". Dey [2010] presented tools for building context-aware applications. He described two main concepts: (i) Distributed information - the application is responsible for sensor data fusion and usage, while the sensor's data is accessed through a Widget, that prepares the sensor's data for the application. (ii) Centralized repository of data – the "blackboard" approach. The blackboard method may also be used by having several distributed blackboards [Corkill, 1991].

Ubiquitous computing (pervasive computing, ambient intelligence) paradigm [Weiser, 1991] contributes another point of view. It is aimed at utilization of nonintrusive networks of sensors and machine learning techniques for context detection and adaptation [Bettini, 2010] as information is distributed through the networks.

Data gathered following the above-mentioned approaches may be used for user and group modeling. A Group Model represents both the specific model of each individual, as well as the group as a whole. It refers to group dimensions that may be based on sociological theories such as communication, conflicts handling, controversy and more [Pizzutilo et al., 2005]. As a result, a system used by a group can adapt its services to the group using a Group Model [Dim and Kuflik, 2010].

A MINI-MODEL CASE STUDY

Our case study involves two people and an exhibit (object) in a museum environment (applicable also in a shopping center). The people's focus of attention may be used for adaptation and personalization. For example, if a person is interested in socializing with a group-mate it wouldn't make sense offering this person a personal device, which may cause isolation (e.g. earphones). There may be several *Attention-categories*: (i) attention to an object (such as product, label, presentation, commercial or exhibit); (ii) social attention - attention to a group member; (iii) integrated attention - a combination of the first two categories, such as in the case of conversation about an object (museum exhibit) that is being observed; (iv) navigation attention ('watch your step'); and (v) intrinsic attention – attention to internal thoughts, mind wandering, or to external distractions (e.g. noise, color, other people, etc.). The question that rises is how a system can distinguish between these categories of attention by using inference and sensor measurements. The identification of categories (i) through (iii) may be supported by fusion and reasoning about low-level sensor data, such as proximity, orientation, sensor detection sector, and voice detection. Attention-category (iv) may be supported by the change in orientation towards the direction of walk. Attention-category (v) may require additional data from sensors such as brain activity detectors. (iv) and (v) are beyond the scope of this paper. The following discussion focuses on the first three Attention-categories. It starts with a description of low level sensors in our experimental environment; continues with a theoretical analysis of the "two people and an object" scenario; and shows how basic measurements such as proximity and orientation can be integrated to assess an increase or a decrease in focus on the Attention-categories (i) through (iii). Finally, we discuss the ability to measure and reason about users attention in the situation of "two people and an object" scenario. The model may be easily expanded to more than two group members, if it is used to represent subgroups of two people from the group.

The PIL Project Sensor-suite

In the framework of the PIL[1] project [Kuflik et al., 2011] a Radio Frequency (RF) based positioning system was installed at the Hecht Museum[2]. It utilizes a wireless sensor network designed and produced by Tretec[3], composed of small (matchbox size) mobile RF tags called Blinds (Figure

---

[1] http://www.cri.haifa.ac.il/connections/pil/, accessed Dec 14th, 2011

[2] http://mushecht.haifa.ac.il/info_eng.aspx, accessed Dec 14th, 2011

[3] http://www.3tec.it/, accessed Dec 14th, 2011

1, left) and RF to TCP Gateways (Figure 1 right) that transfer the data reported by the Blinds over a local area network to the PIL server. Each Blind and gateway has a unique identifier. The Blind can be carried by a person (Figure 1 middle), or located near an exhibit (as a stationary beacon).

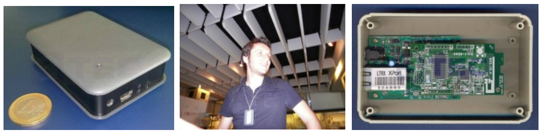

Figure 1. Mobile positioning device and gateway

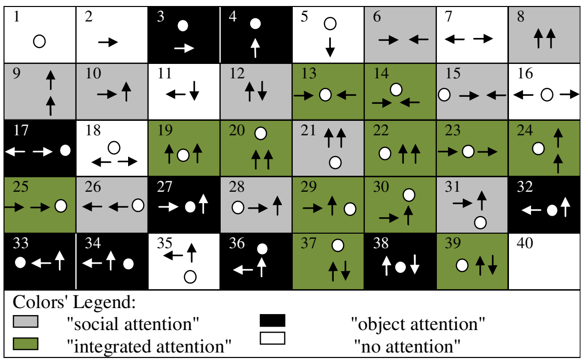

Table 1. Combinations and attention categories of "two people and an object" based on proximity and orientation

The Blind sensor has several important features, including: (i) Measuring proximity among Blinds, which allows to reason about the proximity among visitors. (ii) Detecting voice level and voice activity (due to privacy considerations voice is not recorded), a feature that can be used to assess the level of conversation among visitors as well as their proximity (in this scenario people may have conversation only if they are close to each other). (iii) Detecting orientation of visitors, using embedded magnetometers, enabling assessments such as whether visitors are facing each other, the exhibit, or standing back to back. Finally, (iv) detecting motion by using embedded accelerometers.

Mini-model of the "Two People and an Object" Scenario

The scenario we analyze is a common one: two people coming to a museum or a shopping center. They may be together or apart. They may also be next to an exhibit or a product. The property in focus is the *Attention-categories*: (i) "social attention" that represents high probability that the two people pay attention to each other (one person to the other, or mutual attention); (ii) "object attention" that represents high probability of attention of at least one of the people to the object; (iii) "integrated attention" that represents a combination of (i) and (ii); and finally, (iv) "no attention" where the probability is low that there is attention. These define the mini-model output.

We performed a theoretical analysis of the above scenario, taking into account the proximity and the orientation of the entities (person1, person2 and the object), where orientation resolution was limited to 90 degrees. The results are presented in Table 1. Each person is represented by an arrow symbol pointing at one of the four main orientations (left "←", forward "↑", right "→", and backward "↓"). The object, "○", may be positioned in front, behind, on each side, and between the two people. The analysis yielded 40 unique combinations based on proximity and orientation. Cell 1of table 1 represents an object alone, cell 2 represents a person alone, cells 3 through 5 represent a single person and an object, cells 6 through 12 represent two people (no object), cells 13 through 39 represent two people and an object, and cell 40 represent the case where there is neither a person nor an object. The above theoretical cases are further refined into the Attention-categories, as shown by the color-coded cells in Table 1. White cells represent "no attention" cases, when people are not in close proximity with each other (or with the object) or face away. Gray cells represent "social attention" when the two people are in close proximity and at least one person is facing the other (or standing side by side with the other). Black cells represent "object attention" when at least one person is in close proximity to the object and facing it. Green cells represent "integrated attention", where conditions for both "social attention" and "object attention" exist.

Measuring Attention-categories

The mini-model measurements are collected by three Blind sensors: one located next to the object (exhibit), and the other two are carried by the two people. The mini-model processing starts with the "identification" stage that associates an entity ID (person or object) with a Blind sensor ID. It continues with "proximity assessment" stage that uses proximity between Blinds and associated IDs to assess the entities' proximity. In parallel, the "orientation assessment" stage computes mutual orientation between Blinds. Finally, the "attention refinement" stage uses the entities' proximity and the mutual orientation to generate the Attention-category as the output of the mini-model.

There are some insights in regards to the measurements: (i) Although the orientation sensor within the Blind is accurate (better than 10 degrees), some of the results would remain ambiguous. For example, in combinations 8 and 9 both Blinds will have the same orientation while proximity is detected. Another example relates to 7 and 16 that cannot be differentiated because they have the same orientations, while proximity is not detected in both cases (because of turning away from the other person and / or to the object). (ii) As for proximity, if there is no proximity report, it does not mean that there is no proximity. Occlusions and interferences may cause false negative proximity reports. In this case the reasoning process assumes "no information", until enough messages are gathered. (iii) There are also false positive proximity reports that result from reflections and multi-path of the RF signal. Therefore, a low threshold of 10% of recent proximity messages was set for accepting proximity status. (iv) It is interesting to note that there are cases where the proximity detection limitation becomes an advantage. For example, in combination 7, although in the real world the two people are close to each other, their Blinds do not detect proximity (because the signal is blocked by their bodies, standing back to back), and they are considered having "no attention", which is true in reality because they stand back to back.

Evaluation

We evaluated the above theoretically analyzed model, in order to demonstrate how it can be used in practice. The evaluation was conducted in a natural museum environment with phenomena of RF interferences, occlusion and reflections. Two people were wearing the Blinds on their chest, having the Blind main detection sector in front of them. The object's Blind was positioned on a table oriented at a specific predefined direction. The object's Blind had free 360 degrees transmission and detection; while due to body shielding, detection of the Blinds carried by visitors degraded when facing away from each other or from the object. 36 of the 39 meaningful combinations in Table 1 were tested (in three cases we failed to collect data, and case 40 is meaningless). The tests took 1840 seconds, collecting 4671 Blind messages. Each case lasted for at least 30 seconds, with an average of 51 seconds per case, and an average of 130 Blinds' messages per case. Table 2 presents the results of the Attention-categories detection. Each of the categories (lines in the table) is detected at high recall and precision percentages as presented in table 3. As expected, as a result of the real world environment characteristics, that include shielding, interferences, occlusions and reflections, there are phenomena of false negative (where there are no measurements, even though they are expected) as well as false positive (where there are erroneous measurements), hence the attention assessment becomes probabilistic. Future work may use additional sensors such as voice detection to determine if there is an additional increase in the social attention probability.

Attention	Recall (%)	Precision (%)
No	86	75
Social	86	100
Object	79	83
Integrated	67	67

Table 2. Two people and an object mini-model results

CONCLUSIONS

The study presented the concept of measurement of broader attention-focus perspectives, and demonstrated it with the "two people and an object" scenario. It showed that this mini-model can encapsulate measurements and reasoning process, while exporting only the required output: Attention-category. The mini-model provides a generic framework to bridge the gap between sensors' low-level measurements and the abstract values needed for supporting personalization and adaptation to the momentary need of the small group: socialization, paying attention to the objects in front of them (e.g. a product in a shop or an exhibit in a museum), or both. In general, we showed a general approach where simple measurements from a small number of sensors may enable the inference of abstract UM attributes, given the right process and model.

REFERENCES

1. Bettini, C., Brdiczkab, O., Henricksenc, K., Indulskad, J., Nicklase, D., Ranganathanf, A., Riboni, D. A survey of Context Modeling and Reasoning Techniques. Pervasive and Mobile Computing, 6(2), 161-180 (2010).
2. Corkill, D. D. Blackboard Systems. In AI Expert, Vol.. 6(9), 40-47, September, (1991).
3. Dim, E., and Kuflik, T. Social Signal Processing: Detecting Small Groups Interaction in Leisure Activity. In The Proceeding of the 14th International Conference on Intelligent User Interfaces, IUI'2010, Hong Kong, China, 309-312, (2010).
4. Dey, A. K., and Abowd G. D. Towards a better understanding of context and context-awareness. In CHI'2000 Workshop on the What, Who, Where, When, and How of Context-Awareness, (2000).
5. Dey, A. K. Context-aware Computing. In Krumm, J. (edt.), Ubiquitous Computing Fundamentals, Taylor and Francis Group LLC, (2010).
6. Hamadicharef, B, Zhang , H., Guan, C., Keng, W., Phua, KS., Tee, KP., and Ang, K. Learning EEG-based Spectral-Spatial Patterns for Attention Level Measurement. In: Proceedings of ISCAS2009, Taipei, (2009).
7. James, W. The Principles of Psychology. Harvard UP, Cambridge, MA, (1890/1981).
8. Kray, C., and Baus, Y. A Survey of Mobile Guides. In conjunction with The Fifth International Symposium on Human Computer Interaction with Mobile Devices and Services, (2001).
9. Kuflik T., Lanir J., Dim E., Wecker A. J., Corra' M., Zancanaro M.and Stock O. Indoor Positioning: Challenges and Solutions for Indoor Cultural Heritage Sites. In (IUI 2011), 13-16 February 2011, Palo Alto, CA, USA (2011).
10. Llinas, J. A Survey and Analysis of Frameworks and Framework Issues for Information Fusion Applications. In Hybrid Artificial Intelligence Systems, LNCS, Vol. (6076/2010), 14-23, (2010).
11. Pizzutilo, S., De Carolis, B., Cozzolongo, G., and Ambruoso, F. Group Modeling in a Public Space: Methods Techniques and Experiences. In Proceedings of WSEAS AIC 05, Malta, Spain, 175-180, (2005).
12. Stiefelhagen, R., Yang, J., Waibel, A. Modeling Focus of Attention for Meeting Indexing Based on Multiple Cues. In IEEE Transactions on Neural Networks, Vol. 13, No. 4, (2002).
13. Weiser, M. The computer for the 21st century. In Scientific American, 265(3), 94–104 (1991).

Event Broadcasting Service
An Event-Based Communication Infrastructure

Gerrit Kahl, Christian Bürckert, Lübomira Spassova
DFKI
Affiliation
{*forename.surname*}@dfki.de

Tim Schwartz
Saarland University
Affiliation
schwartz@cs.uni-saarland.de

ABSTRACT
In this paper, we present the Event Broadcasting Service (EBS), which can be used in instrumented environments to exchange data between different devices and services. The EBS can be used with almost all operating systems and programming languages. It enables real-time exchange of large data sets and provides effective debugging tools. Example applications for smart spaces, such as a visualization dashboard and location observation, are presented.

Author Keywords
event broadcasting, platform-independent communication

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: Miscellaneous

General Terms
Design, Management, Measurement

INTRODUCTION
Nowadays, a variety of sensors have been installed in more and more public areas, which therefore can be regarded as instrumented environments. For example, retailers are increasingly embedding technology into their supermarkets in order to improve the shopping experience of their customers and support them in their shopping process. Moreover, customers can also act as human sensors and send current locations detected by their smartphones to the supermarket's server. This collected data could be useful for the shop manager, e.g. to send more employees to the cash points if needed. The collected data could also be deployed in other systems, for example, to personalize the content of assistant systems. To realize these goals, all information should be transmitted and received by different systems in real time. In order to manage the huge data exchange, an appropriate service has to be implemented in the instrumented environment. Due to the fact that the sensor systems in such an instrumented environment are often implemented in different programming languages, this service should provide an easy to use generic interface.

IUI Workshop on Location Awareness for Mixed and Dual Reality LAMDa12, 14, February 2012, Lisbon, Portugal.

In this paper, we propose an event-based communication structure, which is originally implemented in Java, but also offers a web interface such that events can be generated by simply calling a web page. All events are serialized to an XML representation of the corresponding class and then transmitted to a communication server. Other systems can register at this server for different types of events and are notified whenever an event is sent to the server. Afterwards, the events can be deserialized by a native Java deserializer, which can also be used on Android. In order to receive messages in other programming languages, a corresponding deserializer has to be implemented.

RELATED WORK
Johanson et al. developed a system called *iROS* (*interactive Room Operating System*), which can be used to transmit data between processes [1]. This system has been designed as middleware and used in an interactive workspace (*iRoom*) where multiple ubiquitous computing devices are connected to each other in order to help people coming together for collaborations. One part of iROS is the *EventHeap* [3], a communication server which is used by the computing devices in the iRoom to communicate in a client-server-architecture manner. Therefore, the data is packed into *events* which are serialized and transmitted through the server. A client registers at the EventHeap for the events it wants to receive. Furthermore, a TTL (Time-To-Live) can be set, so that clients who will register later can request older events from the heap. According to [3], there are Java, C++, and web service implementations, so that client processes in different languages can connect to the server implemented in Java. The authors mention that iROS is "friendly to existing languages and environments, and straightforward to support a wide range of devices and leverage their existing application bases." Unfortunately, we could not find any newer documentation, which leads to the conclusion that the development has stopped after the publication of the paper. Also the current documentation and sources are not available any longer, hence we could not validate this system using Android devices for transmitting and receiving events.

Metaglue from the Stanford University is a distributed multi-agent system for intelligent environments implemented in Java. It shall meet the requirement to handle large numbers of hardware and software components' interconnection [2]. It is possible to establish communication channels between single agents and maintain their state. It is also created to introduce and modify agents in a running system and to manage shared resources. Furthermore, an event broadcasting mechanism is part of its capabilities. To maintain the configuration and work on the running system, Metaglue has an SQL database and a web-based interface for modifying the agent's configuration-attributes at runtime. The agents make requests in an ad-hoc direct way and can also use event broadcasting to notify groups of agents about context-shifts in room applications. For debugging, a so called "Catalog monitor" is presented which displays all running agents and their reliance interconnections. Still, the authors admit that debugging is difficult, and they hope for the Java community to make more steps into the direction of distributed agent debugging [2].

In [6], two frameworks are presented: *Prism-SF*, an architectural style framework, and *Prism-MW*, an architectural middleware framework. These frameworks can be used to describe network architectures and then work on top of this to let different components communicate with each other. According to the authors, the frameworks can be used in many types of distributed systems like client-server or peer-to-peer networks. Prism-SF provides design guidelines for composing large distributed, decentralized, mobile systems and Prism-MW is a lightweight architectural middleware supporting the implementation of these guidelines [6]. The Prism-MW framework is a composition of a large amount of classes that can be extended and connected to create large distributed systems in which all components have the ability to communicate. They make use of Java's dynamic class loading and DLLs under Windows to add and remove communication ports at runtime. The authors claim that more than a dozen applications have been designed using various instances of Prism-SF, implemented on top of Prism-MW. The system supports PalmOS, WindowsCE and desktop platforms and even digital cameras and motion sensors. However, there is no support for Android devices.

Since all of these system are either very complex to install and configure (Prism and Metaglue) or are not available for modern devices like Android (iROS, Prism), we had to develop a new approach. This Event Broadcasting Service (EBS) aims at being easy to configure, to debug, and to extend.

EVENT BROADCASTING SERVICE

In order to enable an interconnection between different services in an instrumented environment, we developed an event-based communication infrastructure. For the design of this service, we devised four criteria. The developed service should enable users to easily install it and adapt their applications to the infrastructure within a relatively short timeframe. Since there is a variety of different services and sensors in instrumented environments, the communication architecture should offer interfaces for different programming languages and operating systems, especially for embedded and portable ones. In order to cope with a wide range of different applications, the communication interfaces have to be generic. Furthermore, the event broadcasting between several services should be guaranteed to work in nearly real time. Due to the huge amount of data that will be sent by several systems, a simple and easy to use debugging mechanism should be provided natively.

These requirements can be summed up by the following keywords:

1. Simplicity
2. Portability
3. Flexibility
4. Simple Debugging

In order to fulfill our predefined criteria, we decided to implement our infrastructure as an *Event Broadcasting Service* (EBS). Hence, only one server is needed, to which all clients can connect using web sockets and to which they can send events. These events are broadcast to every client connected to the server and filtered on the client side. The corresponding architecture is illustrated in Figure 1.

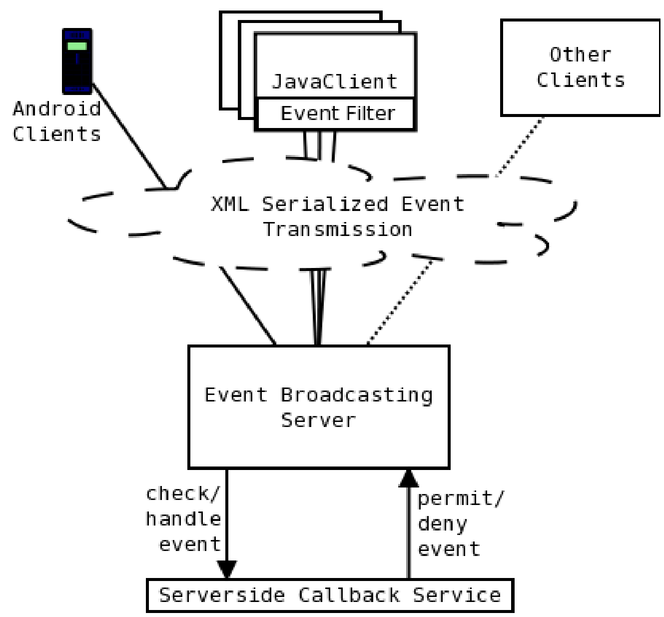

Figure 1. Architecture of the EBS

The EBS server is implemented in Java in order to be independent from the operating system. For the events, we decided to use an XML serialization before transmitting them as simple strings. Due to this serialization, events can be generated and parsed in any programming language supporting XML parsers. For example, Java already implements such a serializer and deserializer that can also be used with Android, namely the *xstream* library[1]. To also enable sensors to send events, the EBS additionally offers a web interface for generating and sending events with primitive datatypes. For this purpose, a URL containing the event name and parameters as key-value pairs has to be called. In contrast to the iROS EventHeap, we obtain the corresponding Java instance after deserialization, where all parameters and functions can be called.

[1] `http://xstream.codehaus.org/`

In order to simplify the debugging of the system, precise runtime exceptions are thrown including stack traces. For instance if the method for a *HelloWorldEvent*, which a client has registered for, is missing, the system outputs the error message "*You forgot to implement onEvent(HelloWorldEvent event) in your callback class HelloWorld*". To make runtime debugging more comfortable and efficient, a *DebugEventClient* can be connected, which simply outputs the XML string of every broadcast event. Every event contains an ID defined by the programmer, which makes it possible to infer its origin. All clients automatically send *HelloEvents* at regular intervals, containing the origin ID and a list of all events they are currently listening to, which makes it possible to recognize if clients are available or blocked. Development and debugging processes are further simplified by the opportunity to specify an auto reconnect. This means if the server is lost, a client blocks its send operations until the server is available again, while informing regularly about its own connection state. This makes it possible to restart the server and the clients in a running environment.

Implementation Details

In the currently available Java and Android implementations, all events are extensions of the class *Event*. Events can theoretically be very complex data structures composed of various other classes. To simplify the usage of the EBS, we decided to limit the client's functionality to a minimum, implementing only the following methods: *connecting*, *event transmitting*, and *event receiving*. In order to establish a connection, a client has to be created by defining its event receiving handler (callback) and a set of events it listens to. Afterwards, events can be transmitted by calling the *send* method, which takes an instance of an event class that is to be broadcasted. The event received on the server side will be automatically passed to the callback instances of the connected clients. The latter must implement a special method for every event they have registered for, which defines how a specific event is to be handled. These methods must be named *onEvent* and must take one argument of the corresponding event's class type. For example, a client listening to a *HelloWorldEvent* must provide the method *public void onEvent(HelloWorldEvent event)* in its own callback class.

To make the system even more flexible, the server can also be extended by a callback method, which is guaranteed to be executed before broadcasting a detected event and thus allows to permit or deny the broadcasting of specific events (filter). Some advantages of this extension are discussed in the following section.

Deployment in an Instrumented Shopping Environment

The Innovative Retail Laboratory (IRL) [8] is a small experimental instrumented retail environment, in which modern shopping assistance systems are developed and tested. In this environment, we have to face the problem of many different systems (mobile and embedded systems, servers, etc.) running all kinds of operating systems, e.g. Windows, Android, Mac OS and Linux. Since all components should be interconnected, a middleware had to be developed that allows the different systems to interact. Furthermore, all products of the instrumented environment operate on a centralized database. One of the challenges to be handled in this context was the need to inform all systems in the environment of possible changes, e.g. the re-location of an object or person.

Apart from the transmission of data from sensors and systems to other systems, the current global state of the instrumented environment needs to be observed. For this purpose, we decided to set up a database containing all information concerning the current state of the environment, e.g. the positions of all objects. In order to keep this data coherent, we use the aforementioned events to update the database. Since the changes of the database are mainly simple transactions, we outsourced them to a centralized service and included it into the core EBS server. All events sent to the EBS server are directed to the *Synchronization Service* (*SyncService*) before being broadcast. An event received by the SyncService triggers an update of the database and is then forwarded to the broadcasting algorithm or denied depending on the current filter options. This guarantees that the clients are informed about a change only after it has been captured in the database. The SyncService is the only component which is capable of editing the database. Since the events are broadcast, every service in our environment will be informed about changes and can get the information out of the event or the database, which is updated before the event has been broadcast.

In our opinion, the EBS infrastructure provides a suitable approach to decoupling services from their corresponding user interfaces (UI). The UIs listen to events and display changes while the backend services receive sensory data and produce appropriate events. Relevant user interactions with sensors or UI elements also result in events, which offers all listening clients the opportunity to react to them. For example, one of our systems reacts to the presence of certain objects and displays relevant information on a screen as soon as an object has been detected at a specified location and clears this information when the object is removed again. Whenever sensors detect the absence or presence of an item, a corresponding event is sent by the sensor client and received by the user interface component, which then reacts appropriately. Using this approach, graphical user interfaces can also easily be decoupled from the sensors, which facilitates the development and comparison of different UIs on different operating systems using the same sensor data. Additionally, the components can be tested beforehand, without having real sensor data, by just sending the corresponding events and hence simulating certain state changes. With this architecture, systems can react to events sent by sensors or by simulators in the same way, which offers a comfortable way to debug programs.

APPLICATION USING THE DUAL REALITY PARADIGM

Using this architecture, a great number of events can be sent by the sensors and systems of the instrumented environment. These events comprise different pieces of information including data measured by sensors or information provided by other systems. In order to provide a representation of the detected state changes, we are developing a component aiming at visualizing the current state of the instrumented environment using a dashboard-like metaphor. Each change of the environment detected by sensors will be transmitted via events to the EBS server, which will broadcast them to this dashboard component, which itself is registered as a client. The dashboard component aims at monitoring and controlling the services I/O behavior, which can be detected in the instrumented environment. Apart from the visual representation itself, interaction in this visual representation should have an influence on the real world, e.g., if the user changes a parameter in the virtual model, the corresponding change is reflected in the real environment. This bi-directional communication from the real world to a virtual model and vice versa is referred to as *Dual Reality* [5]. The dashboard component should also be implemented as a generic interface to enable the inclusion of simulators that can influence the virtual representation but also the real world [4].

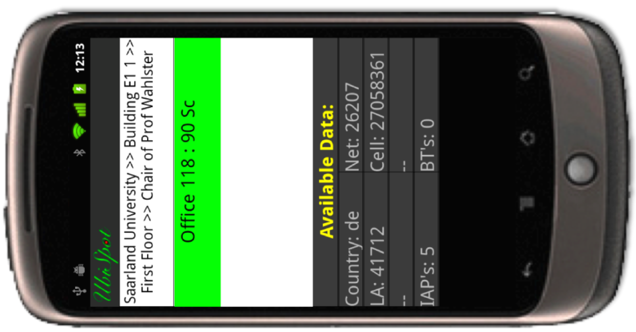

Figure 2. Smart phone running UbiSpot

In smart spaces, it is important to keep track of the locations of people in this environment. In [7], an *Always Best Positioned* system called UBISPOT is described, which uses cell towers, WiFi and Bluetooth information to estimate the current position of an object or person. Originally, this system has been implemented for SymbianOS smart phones. We reimplemented the algorithms for Android devices (a screenshot of UBISPOT for Android devices is shown in Figure 2). Using EBS, users can decide to share their anonymized locations with the infrastructure of a smart space. In this context, "anonymized" means that the identity of the person is not revealed. Still, the provided information can be helpful, for example in airports or large malls, to enable a manager to assign workers appropriately. These locations can, for example, be visualized in the previously introduced dashboard.

CONCLUSION AND FUTURE WORK

The Event Broadcasting Service provides an easy and efficient way to interconnect different services running on different devices by exchanging events. The server itself is implemented in Java, the clients, however, can run on any operation system due to the XML representation of the events. While interfaces for receiving events have just been implemented in Java, client services can be written in any programming language to send events using the web interface for event generation and transmission. Early tests of the EBS showed fast performance. In these test, clients ran on Mac OS X, Linux, Windows XP, Windows 7, and Android.

For future work, we plan to run intensive tests of the server and to further extend the EBS. One such extension will be a protocol for question answering, which enables to trace possible communication problems and further improves the stability by monitoring if queries have already been answered and optionally resending the appropriate events. Furthermore, the dashboard system will be extended to visualize further events delivered by the intelligent environment as well as usage statistics.

ACKNOWLEDGEMENT

This research was funded in part by the German Federal Ministry of Education and Research and developed in the projects EMERGENT (grant number 01 IC 10S01H) and SWINNG (grant number 01 IC 10S05A). The responsibility for this publication lies with the authors.

REFERENCES

1. Borchers, J., Ringel, M., Tyler, J., and Fox, A. Stanford Interactive Workspaces: A Framework for Physical and Graphical User Interface Prototyping. *Wireless Communications 9*, 6 (2002), 64–69.

2. Coen, M. H., Phillips, B., Warshawsky, N., Weisman, L., Peters, S., and Finin, P. Meeting the computational needs of intelligent environments: The metaglue system. In *Proceedings of MANSE'99* (1999), 201–212.

3. Johanson, B., and Fox, A. The event heap: A coordination infrastructure for interactive workspaces. In *Proceedings of the Fourth Workshop on Mobile Computing Systems and Applications*, IEEE (2002), 83–93.

4. Kahl, G., Warwas, S., PascalLiedtke, Spassova, L., and Brandherm, B. Management dashboard in a retail scenario. In *Workshop on Location Awareness in Dual and Mixed Reality. International Conference on Intelligent User Interfaces (IUI-11), February 13, Palo Alto,, California, United States*, G. Kahl, T. Schwartz, P. Nurmi, E. Dim, and A. Forsblom, Eds., Online-Proceedings (2011), 22–25.

5. Lifton, J., and Paradiso, J. A. Dual Reality: Merging the Real and Virtual. In *Proceedings of the First International ICST Conference on Facets of Virtual Environments (FaVE)* (July 2009).

6. Medvidovic, N., Mikic-Rakic, M., Mehta, N. R., and Malek, S. Software architectural support for handheld computing. *Computer 36* (September 2003), 66–73.

7. Schwartz, T., Stahl, C., Müller, C., Dimitrov, V., and Ji, H. UbiSpot – A User-Trained Always Best Positioned Engine for Mobile Phones. In *Proceedings of Ubiquitous Positioning Indoor Navigation and Location Based Services (2010)*, IEEE XPlore (2010). Online Publication, no pagenumbers.

8. Spassova, L., Schöning, J., Kahl, G., and Krüger, A. Innovative retail laboratory. In *Roots for the Future of Ambient Intelligence. European Conference on Ambient Intelligence (AmI-09), in Conjunction with 3rd, November 18-21, Salzburg, Austria*, o.A. (2009).

PolySocial Reality: Prospects for Extending User Capabilities Beyond Mixed, Dual and Blended Reality

Sally A. Applin
Centre for Social Anthropology and Computing
School of Anthropology and Conservation
University of Kent Canterbury, CT2 7NR
sally@sally.com

Michael Fischer
Centre for Social Anthropology and Computing
School of Anthropology and Conservation
University of Kent Canterbury, CT2 7NR
m.d.fischer@kent.ac.uk

ABSTRACT
The technology industry has evolved over the years with a development lens increasingly focused on end users and usage cases. Indeed, for the past decade or more, personas (the designer-created profiles of end users) have become stand-ins for various usage cases and user models. With regard to location aware software and mobile applications, the usage of Dual Reality and Mixed Reality as metaphors have functioned in a similar vein. Just as personas are not people, Mixed and Dual Reality do not fully represent or address the complex usage cases developing as more people do more things, with more software at more times and in more spaces than ever before. This new complex application ecosystem presents greater opportunities and challenges for application design. We discuss ways that developers can use PolySocial Reality (PoSR) to represent a more complete complex structural model of individuals interacting within multiple environments.

Author Keywords
Time, Space, Asynchronous, Ubiquitous, Pervasive, Dual Reality, Mixed Reality, Blended Reality, PolySocial Reality (PoSR), User Experience Design, Interaction Design

ACM Classification Keywords
H.1.2 [Human Factors]: Human information processing; J.4 [Social and Behavioral Sciences] Anthropology; B.4.3 [Interconnections (Subsystems)]: Asynchronous/synchronous operations; K.4.1 [Public Policy Issues]: Transborder data flow; K.4.3: [Organizational Impacts] Computer-supported collaborative work

INTRODUCTION
To fully exploit location awareness, future interaction design development for the User Experience (UX) will be increasingly directed by users as they create new capabilities situated in social, physical, and network space [1][2]. A conceptual understanding of the global interaction context within which people experience the social mobile web is needed, one that emerges from the aggregate of multiplexed data pathways connecting interacting individuals[3].

At the moment, interaction models tend to be based on fixed navigational pathways and single narratives. Future UX development for location awareness must provide environments for sociability and shared experiences within a multiplexed environment.

IUI Workshop on Location Awareness for Mixed and Dual Reality
LAMDa'12, 14, February 2012, Lisbon, Portugal.

COMPLEXITY AND THE SOCIAL MOBILE WEB
People must be social to survive. We are dependent on each other and the systems that we build with each other to exist. Edward T. Hall wrote, "Man and his extensions constitute one interrelated system," [4]. As much as we'd like to separate that which is 'social' from that which is in the environment, we cannot, for these are interdependent.

Originally, Social Media apps provided a network 'place' for people to simply connect with others and share media or text [5]. Now, the addition of mobile data devices, and/or smart phones, with location-aware apps (the social mobile web), has enabled people within this system to create and utilize new capabilities. These include being able to publish, broadcast and share their locations, earn badges, points or discounts for disclosing this information, and track the locations of others who are also members of these applications, and who share or contribute to their various schemes. Many people find that they are able to increase their social time by finding friends gathered at specific locations.

Thus, social mobile web apps based on connections that might be distant in time, space (locational framework, coordinate space, etc...), and place (location, local context) offer overwhelming opportunity and choice for people to communicate, collaborate and connect with each other. Furthermore, when using social mobile web apps, people are only partially engaging in shared common networks at any given time [2][3].

For the developer, there is much more to support. Because the people using these apps are innovating their own usage cases with these new capabilities there is a need for support for these people as they move through each new usage case. When the developer doesn't consider the multiple ways people are connecting, opportunities may be missed and more importantly, people may be impaired by not being able to utilize more capabilities. This in turn could impact those interrelated systems that humans need to exist. Since the offering of opportunities and their associated capabilities is multiplexed, what can app developers contribute towards supporting this model?

One way for developers to connect to the multiplexed social mobile web user is to support the complexity of usage cases. It may seem orthogonal to do so, as most developers and User Experience professionals are instructed to create a more simple system. In this case, however, the system may need to remain complex in order to fulfill user expectations.

Science and scholarship have been driven for the past few thousand years by the need to simplify phenomena to the point that reasonable descriptions and explanations could be achieved, and impressive results have emerged. But by the mid 20th century, it was clear that the amount of detail that contributed to physical phenomena was even greater than realized before, that mathematics did not have the capacity to provide perfect descriptions of phenomena, that uncertainty was a feature of reality (not a bug), that matter, energy and information were in principle interchangeable, that observers were a part of an observed phenomena, that there were limits to the universe, and that scale mattered.

Over four decades ago these understandings were leading to a new way of creating knowledge manifest in two basic forms, both of which have slowly eroded the focus on simplifying structuralism that dominated the 20th century. One was the rise of complexity theory, possibly first popularly manifest in the work of Réne Thom on Catastrophe Theory[6][7], who in addition to the main tonic of his work, laid out a mathematical framework for describing dynamical phenomena, real and imagined. In the humanities post-structuralism (and then postmodernism) came to the fore, and turned out to be a relatively good tool for exposing the shortcomings of structuralism, but provided no means to reconcile or replace structural approaches. This approach exposed complexity, but provided no tools to address complexity, although Actor Network Theory, as popularized and elaborated by Bruno Latour shows some promise [8].

POLYSOCIAL REALITY

Applin and Fischer [2][9] have suggested PolySocial Reality as a term for the conceptual model of the global interaction context within which people experience the social mobile web.

PoSR is based upon the core concept that dynamic relational structures emerge from the aggregate of multiplexed asynchronous or synchronous data creations of all individuals within the domain of networked individuated networked or local experiences. In other words, PoSR describes the aggregate of all the experienced 'locations' and 'communications' of all individual people in multiple networks at the same or different times.

For example, a PoSR context emerges when a person is walking down the street and talking on the phone and texting and another person is doing the same thing with them while both parties may be communicating through different channels to other people as well. Or when a person enters an environment and checks into foursquare which delivers a tweet and a Facebook update notating their location while another person responds to that in real time. The transmissions between people are fragmented, and PoSR describes the relationship emerging from these fragmented transmissions. PoSR describes the network transaction space that humans are inhabiting themselves and with others in order to maintain their relationships and engage in new activities with collective dependencies via the social mobile web. Thus, multiple-channeled network interactions lead to complex relationships with others.

If a person is processing multiplexed data creations and another person is processing others, and both people come together, how is commonality determined by and between the parties? If a third person joins in, how are they able to sort out where there is common ground? PoSR space can get very, very complex, pretty quickly.

As a interaction context, PoSR has positive and negative outcomes. A potentially positive outcome may be an expanded social network, a negative outcome may be that those expanded social networks are connected by small, single dimension attributes. Another may be that the fragmentation of PoSR encourages individuation, which makes it more difficult for humans to be social (and cooperative) with one another, even as they effectively have a larger social network. While implementations continue to focus on individuated orientations, this can further compound that problem.

To the extent that people share common sources of information while interacting with each other, the greater their capacity to collaborate. If they share too few channels relevant to a common goal, there may be too little mutual information about a transaction to interact and communicate well collaboratively. Poor collaborative interaction can lead to further relational fragmentation with the potential to promote individuation on a broad scale. By changing the means that humans use to manage space and time during their daily routines, developers can shift our experience from individuated, user experiences to enhanced sociability within a multi-user, multiple application, multiplexed messaging, PoSR environment.

We are not arguing against individualism, but promoting people's ability to control and augment their individual context through leveraging the elaborated collective capacities that defines humanity and enables individuals to create productive innovations.

COMPLEXITY AND MEDIATED INTERACTION

Our main purpose is to try to make more concrete how developers might leverage PoSR contexts to to create more dynamic complexity aware applications. As a stage towards a typology, Table 1 is a set of reference case types representing levels of Agent/Technology Interaction, with an indication of user applications and considerations for developers in supporting and extending these.

The table is organized around cases as: a) the mix of people and technology involved in a technology mediated activity; b) the basic user context with respect to problems and solutions; and c) the concepts and technologies a designer or developer might bring to the problem.

While not explicitly represented, the table is shaped by designer/developer approaches with respect to structured, object-based and agent-based technologies.

Case 0 is meant to set a baseline, and might include the archetype hacker who uses and composes a set of technologies with only functional contributions by other designers/developers. But it could also represent more ordinary people operating simple independent appliances like an alarm clock or VCR remote control to achieve results. Cases 1 and 2 represent most conventional applications where a UX is essential to enable more people to engage in the aggregate functionalities of Case 0, where an important part of the application is the metaphor or framework developed by the designer and implemented by the developer. While Case 1 might be based on structured development methodologies, Case 2 would tend to depend on object-based design for more than abstraction to effectively represent the required complexity in the agent/technology interaction.

Case 3 represents where most location aware applications are focused, where object-based approaches with inclusion of some agent-based technologies is needed to represent

Mediated Interaction	Interaction Layer	Developer/Designer (D/D) Layer
CASE 0	**Individual using a Technology**	**Single User/No UX**
Single person interaction with technology Context Free	• Analog or Digital things	• No load on D/D to explain or instruct beyond original implementation. • Functions and Operations. • No directions for usage • Minimal or no UX (e.g. Libraries or simple shell)
CASE 1	**Passive Integration of Technology with Local Environment**	**Single User/Addition of UX**
Single person interaction with technology In static environment Context Sensitive	• Designer access to environment is passive. • Users create contextual information, such as setting system clock, providing ID information, etc... • User interprets/ conforms to system metaphor to interact with program. • D/D sets up environment by forcing user compliance and through inferring contexts of use from user supplied information.	• Passive approach to local environment based on user input. • Make inferences that certain information about environment will be available from user. • D/D sets up environment by forcing user compliance to a system design, usually through some model metaphor (e.g. Form, GUI or Externally provided UX). • No assumptions for many possible contexts of use. • Object-based for abstraction.
CASE 2	**Active Integration of Technology with Local Environment**	**Single User/Non-Interactive Services/ UX Active**
Single person interaction with technology In dynamic environment	• View reports or observe technology outcomes based on dynamic data gathering relating to local environment • Limited interaction with external context-free network services.	• Program behavior can be modified based on information gathered. • Recognition of changing locations and other circumstances. • Modification of environment by the technology • Dynamic objects for emergent results.
CASE 3	**Interaction Between Environments**	**Single User/ Interactive Services/ UX Active**
Single person interaction with technology In local and remote environments [3] Dual Reality model can work	• Interactive services • Interactive exchange of information between contexts • Relationship between actions and outcomes in different environments. • Limited asymmetrical communication	• Service integration • Several environments can be modified as a result of the technology. • Aware of multiple use environments • Facilitate limited communication • Programming is a hybrid, mostly object based with some agents.
CASE 4	**Social Apps: Interaction Between Individuals within their Environments**	**2+ Users/Interactive Services UX Active/ Multi Place/ Homogenous People**
2+ familiar or similar people interacting synchronously or asynchronously. PoSR begins as a design consideration (PoSR less disruptive)	• Social interaction direct or indirect between two or more people. • Relatively homogenous individuals or individuals within a limited consistent set of roles. • Similar, but different, environments between interactants.	• Aware of multiple individual participants. • Support cooperation, groupware • Awareness that the technology partially defines the amount of detail available to each interactant [3]. • Awareness: users must infer missing information about others' contexts • Users as agents
CASE 5	**Social Apps: Interaction Between Differentiated Individuals within their Environments**	**2+ Users/Interactive Services UX Active/ Multi Place/ Heterogenous People**
2+ diverse people interacting synchronously or asynchronously. PoSR is fully functioning here PoSR is disruptive here	• Social interaction direct or indirect between two people • Differentiated individuals on language, culture, status, etc. or individuals within a diverse set of roles • Highly differentiated environments • Differential knowledge	• Details about the context of others are missing and may be difficult for individual users to infer or details that cannot be inferred. • Highly complex elements of differentiated environments are combined into structures that appear different from each users' POV. • Users as distributed dynamic unique agents.

Table 1. Actor/Technology Interaction Cases

interactions between multiple environments and their manipulation, while still limiting consideration of how different users might contribute to the success or failure of the application. Dual, mixed and blended reality are probably sufficient for this level of interaction.

Cases 4 and 5 represent where we think location aware applications are going based on present user generated applications. Effectively, in addition to all technologies in use, all users themselves become, at least in part, agents in the application and thus fall, in part, within the designer/developer responsibility [10]. The more complex model associated with PoSR is required here to coordinate this information. The main difference is in assumptions of homogeneity of agents and their activities assumed by the designer/developer. Case 4 assumes they are similar or very structured in differences. Case 5 is considerably more ambitious as it requires much more information be available to both the application and the users since it cannot easily be inferred or imposed by either users or designer/developer, but will ultimately be required to extend advanced user created applications to a broader group of end users, an ubiquity necessary for an application to be truly social. Elements of Case 4 are probably reflected informally in location aware applications, but will benefit from formal inclusion.

One possible approach consistent with our proposal is to adapt Auber's [11] anoptic representation. Auber argues that the collective intelligence of a group develops only to the extent each individual has access to at least one form of representation of the group's activity that includes all members in some form, varies with members activities and is considered legitimate by all. It need not look the same to everyone: each may have their own projection.

CONCLUSION

Concepts such as Dual Reality and Mixed Reality, and their subsequent technologies were sufficient for most location aware applications, but not sufficient to meet the opportunities and capabilities that location aware applications are opening up. PoSR is capable of representing these relationships, while including multiple users' relative viewpoints.

One approach to implementing applications corresponding to Cases 4 and 5 is to develop anoptic representations based on additional model agents that partially direct and coordinate activities of the user agents by establishing model 'best practice,' mediate communications and explicitly gather, seek and communicate information required by all parties in collaboration with users or their agent representations. In this manner the connections represented by PoSR can be managed with respect to positive and negative outcomes.

Simple examples include plotting people and their attributes and activities on maps. But even with a dynamic legend and the capacity to navigate through the history, this is limited. Bluebrains [12] constructs a soundtrack from fixed pieces associated with specific locations that people can interact with in a non-linear way. The agency of the individual creates the resulting soundtrack, which is tied to the locations visited. Again this has limits.

An appropriate means of representing PoSR-based descriptions should include creating some form of dynamic commentary constructed from any combination of visual, aural or language-based elements that can be modified, rescaled and browsed by end users to find information they require from the present or past about others they are interacting with directly or indirectly in a compact form.

PoSR descriptions offer location aware applications a tractable means of traversing the complexity of single and multiple user experiences while maintaining the complexity required by users to construct further applications of the technologies they employ.

ACKNOWLEDGMENTS

The Natural Environment Research Council (UK) for grant NE/I004149/1 and the National Science Foundation (NSF)

REFERENCES

1. Michael D. Fischer. “Cultural Agents: A Community of Minds.” *Engineering Societies in the Agents World VI*. Oguz Dikenelli, Marie-Pierre Gleizes and Alessandro Ricci (eds). Lecture Notes in Computer Science 3963, pp 259-274. Springer-Verlag, 2006
2. Sally A. Applin, and Michael Fischer. A Cultural Perspective on Mixed, Dual and Blended Reality. *IUI '10: Proceedings of the 14th international conference on Intelligent user interfaces*. ACM, 2011.
3. Joshua H. Lifton. *Dual Reality: An Emerging Medium*. Doctoral Dissertation. MIT 2007.
4. Edward T. Hall. *Beyond Culture*. New York. Basic Books. 1976.
5. Michael D. Fischer, S. Lyon, D. Zeitlyn. The Internet and the Future of Social Science Research. In: N. Fielding and R. Lee (eds.) *Online Research*, pp 517-534. Sage, London. 2008.
6. Thom, Réne. *Structural Stability and Morphogenesis*, W. A. Benjam, 1972, ISBN 0-201-40685-3.
7. David Aubin, "Forms of Explanations in the Catastrophe Theory of René Thom: Topology, Morphogenesis, and Structuralism," in Growing Explanations: Historical Perspective on the Sciences of Complexity, ed. M. N. Wise, Durham: Duke University Press, 2004, 95-130.
8. Bruno Latour. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University Press, 2005.
9. Sally A. Applin and Michael Fischer. Pervasive Computing in Time and Space: The Culture and Context of 'Place' Integration. 7th International Conference on Intelligent Environments. IE'2011. ACM 2011. pp 285-293.
10. Michael D. Fischer. Cultural dynamics: formal descriptions of cultural processes. *Structure and Dynamics.* v3:2 2008. http://www.escholarship.org/uc/item/557126nz
11. Olivier Auber (translation: Florence Meichel). "Concept of Anopticism Anoptisme" (Français). 2008. http://tinyurl.com/7e3o4hj (translation: http://p2pfoundation.net/Anoptism).
12. G. Geere. The National Mall: a location-aware app-album. Wired UK. May 27, 2011. [Online]. http://tinyurl.com/3o3g3fm

Mixed Environment Adaptive System for Point of Interest Awareness

Paulo Pombinho, Ana Paula Afonso, Maria Beatriz Carmo
Faculdade de Ciências da Universidade de Lisboa
Campo Grande, Lisboa, Portugal
ppombinho@lasige.di.fc.ul.pt, {apa, bc}@di.fc.ul.pt

ABSTRACT
Location based services (LBS) are now very common, but are mainly developed for outdoor areas. Despite the existence of some indoor LBS, they either require physical building infrastructures or employ a complex and expensive conjunction of positioning systems to achieve good accuracy. Furthermore, they are not commonly used for both indoor and outdoor environments. In this paper we propose a point of interest LBS for both indoor and outdoor environments that automatically adapts the interface according to the type of environment. We use an affordable and low-cost smartphone equipped with GPS, compass and accelerometer to provide these functionalities. The positioning algorithm proposed uses information obtained when the user is outdoors to improve the positioning accuracy while indoors.

Author Keywords
Indoor Location, Mixed Environments, Context Aware, Mobile Devices

ACM Classification Keywords
H5.2 Information interfaces and presentation: User Interfaces - *Prototyping*.

General Terms
Algorithms, Experimentation.

INTRODUCTION
Location based services are becoming very common in new mobile devices and, for this reason, the real time information about the location of users has become widely used in an extensive range of location based applications.

When used inside a building, location based applications can be used to show relevant information concerning the location of the user and aid the navigation in an unknown building, for example a museum or a university campus. Indoor location has also been suggested to allow emergency services to explore unknown areas in an easier and more efficient way [1, 2].

Despite being reliable and sufficiently precise when in an open field, GPS devices need to "see" a large portion on the sky to be able to correctly calculate the position of the user. Thus, when indoors, GPS devices become useless and the requirements of measurement error change. Alternative positioning systems, like GSM based algorithms, do not have sufficient precision or need an expensive physical framework. This limitation hinders the development of indoor location based applications [3].

An alternative positioning method is to use new mobile devices that have integrated accelerometer and digital compass to detect when the user is moving and the direction of the movement, and use this information to calculate the user location. However, the results obtained when using this type of approach are very dependent on the way each particular person moves, causing a potentially large error when used for long periods of time or distances [4].

Our goal is to develop an affordable system that does not need a physical infrastructure, and uses only the sensors that are commonly integrated in the new mobile devices. In this paper we present our work in progress on a system that aims to allow mixed (indoor and outdoor) positioning, automatically adapting the interface based on the user's location and how the user holds the mobile device. Additionally, this system collects, automatically and transparently, information when the user is outdoors, and uses it to improve the indoor positioning by adjusting the algorithm parameters for the type of movement being done by a particular user.

In the next section we will describe the most relevant related work. Afterwards we will explain the indoor positioning algorithm used and, in the next section, how this positioning can be improved using data gathered when the user is outdoors. We then describe the user interface of the developed mobile point of interest application to assist users both in outdoor and indoor environments and finally we present conclusions and future work.

LAMDa'12, February 14th 2012, Lisbon, Portugal.

RELATED WORK

There are some works that explore indoor positioning mechanisms. There are several diverse approaches that use transmitters of some kind, installed on the buildings, and corresponding receivers, carried by the user. Some systems use infrared transmitters [5], RFID tags [6], VHF radio [7], or Bluetooth beacons [8].

Several systems have explored the use of Wi-Fi network access points, and operate by identifying and processing the signal strength information of multiple base stations to triangulate the position of the user (see for instance [9]).

Regarding infrastructure free positioning, Kourogi et al. [10] use sensors placed on the waist of the user, to detect walking stance and velocity. Some approaches use shoe mounted sensors to detect the displacement made by the foot in each footstep and consequently the displacement made by the user [11]. Finally, Glanzer et al. [12] present a pedestrian navigation system that uses a set of diverse sensors to estimate changes in position and attitude, and obtain the final position of the user.

Although they focus mainly on how to obtain the indoor position, there are also some works that research different types of indoor location based services. As an example, in Jensen et al. [2] a review of several indoor location based services and systems is done.

Despite providing solutions for the problem, the works presented are either based on the existence of an infrastructure in each building, or the need of external sensors placed, for example, on the user's shoes or waist. These sensors are a potential limitation to the natural movement of the user or the practicability of the system. Furthermore, some of the systems require expensive equipments and others, although using cheap beacons, need to install a large number of these to obtain good accuracy.

Very recently, some preliminary works have started to appear that focus only on the use of sensors integrated in the mobile devices, like [13, 14]. However, both these systems and those referred above are designed to be used only indoors, do not take into account the movement made while the user is outdoors. There are also no works focused on the dynamic adaptation to these environment changes. Since we aim to develop a mixed environment (indoor and outdoor) application, our goal is to develop an approach that will use information obtained while outdoors to improve the users indoors positioning.

INDOOR POSITIONING ALGORITHM

To obtain the position of the user inside a building we have previously proposed an approach based on a step detection algorithm [4] that allows the detection of the user's motion and its direction. In the next section we will, briefly, describe this approach.

Step Detection Algorithm

The algorithm uses an accelerometer integrated in the mobile device to capture, in real time, the accelerations it is being subjected to. When the user is walking he will apply, not only, a forward acceleration, but also, with a greater magnitude, a vertical upward acceleration followed by a vertical downward one, in a consecutive way. Figure 1 shows an example of the shifts in acceleration during a three step movement.

There are four step detection parameters used to detect each step (Figure 1): A peak amplitude λ_p that represents the minimum positive shift in acceleration, caused by a step; a negative amplitude λ_t, that represents the minimum negative shift in acceleration; a Δt_{min} minimum time interval, that needs to pass for step to be detected, and a Δt_{max} maximum time that cannot be exceeded.

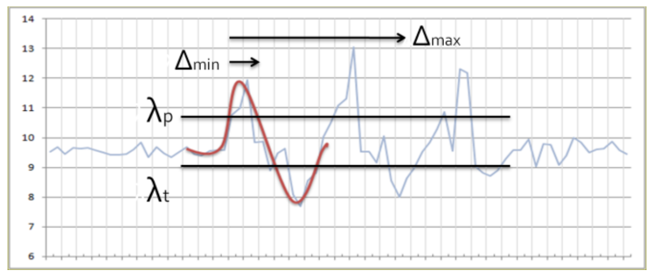

Figure 1. Shifts in the acceleration vector while the user is walking and step detection parameters. The thick red line shows the pattern for a single step.

To be able to identify the location of the user inside a building, the last known position is used, obtained with the GPS while still outside, as the initial position of the user. Next, as each step is detected, the orientation, obtained from the digital compass, and the medium step size is used to calculate de movement done by the user. By adding all these displacements it is possible to infer the trajectory of the user inside the building and calculate his current position.

Algorithm Accuracy Errors

Although the described algorithm gives an accurate positioning when used for short periods of time, it can, on the long run, accumulate a large accuracy error.

There are four types of errors present on this type of algorithm. Steps not being detected (false negatives); steps being incorrectly detected (false positives); errors in the size of the step used; and compass orientation errors.

For all of these types of errors, only the compass orientation errors are user independent, being caused either by poor compass accuracy or by disturbances in the magnetic field caused by metal in the vicinity or other devices interference. The remaining three types of errors are mainly due to the use of parameters that are not suited for the way a particular user moves or even the type of movement the user is doing at a specific time.

Due to this, if these parameters can automatically adapt themselves to each user and their type of movement, the errors can be minimized giving the algorithm a much better accuracy.

OUTDOOR ASSISTED INDOOR POSITIONING

To be able to improve the positioning of the user while inside a building, we use information obtained when the user is outdoors, where we can use the GPS to verify the accuracy of the step detection parameters. However, the type of movement done when a person is outdoors is different from the type of movement when he/she enters a building [4].

Indoor / Outdoor Comparison

In [4] we have performed a user study that identified the correlation between the movements performed in outdoor areas and those done in indoor areas. In these experiments seven users (with heights and weights ranging from 160 cm to 180 cm and 65 kg to 81 kg, respectively) walked at different paces, both outdoors and indoors. The data collected during the experiments allowed the analysis of the walking pattern of each user and determine, for each step, the optimal detection patterns.

By comparing the optimal detection patterns for each user, in each environment, and for different speeds, we were able to calculate the ratio, for each user, between the indoor and outdoor experiments at the same speed. The average ratios obtained were 0.96 for the λ_p, 1.05 for λ_t, and 1.1 for the Δt.

It is also essential to choose the right step size since, in the long run, it can originate a high amount of error. The average step size obtained in [4] is 65 cm. However, this study also shows significant variations depending on the speed of the users and also if the movement is indoors or outdoors. Despite capturing very diverse step lengths, the ratio obtained between each pair of indoor / outdoor experiments is fairly constant (0.9).

Indoor Step Detection Parameter Adjustment

To be able to adapt the indoor positioning parameters we first need to find the optimal parameters for the user while outdoors. To achieve this we capture the positioning information, obtained from the GPS, and calculate at certain intervals the displacement and speed of the user, and also create a log of the values returned by the device's accelerometer.

Using this information, we can automatically adjust the step detection parameters in order to be coherent with the user's movement. Furthermore, using the distance traveled and the number of steps, we are able to calculate the average step size of the user, while outdoors.

Having obtained the optimal step detection parameters for the outdoor environment, we adapt these values for the indoor environment by applying the indoor / outdoor ratios described in the previous section. Thus, it is possible, to improve automatically, and in a transparent way to the user, the accuracy of the system.

USER INTERFACE

The developed mobile application assists the user in searching nearby points of interest in mixed environments and consists of three main interfaces, two for outdoor use and another for indoor use. The application was developed using the RUBI open source augmented reality platform [15].

Outdoor Interface

By using the device's accelerometer, we are able to detect the device's pitch, yaw, and roll angles, and consequently how the user is holding the mobile device. When the user is outdoors, there are two types of interface that can be changed automatically.

Figure 2. Outdoor augmented reality view.

If the user is holding the mobile device in front of him, perpendicular to the ground, we use an augmented reality view that depicts nearby points of interest as circles drawn over a real time video feed from the device's camera (Figure 2). On the lower left part of the screen, a compass radar is shown to enable the user to become aware of other points of interest that exist in the surrounding area.

Figure 3. Outdoor map view.

When the user lowers the mobile device, parallel to the ground, we switch to a map view, where the nearby points of interest are depicted as icons drawn over the map (Figure 3).

Indoor Interface

When the user enters an indoor area, the application automatically changes the view to a 2D floor plan of the building, showing the points of interest that exist inside the building, depicted as icons drawn over the floor plan (Figure 4). The trajectory walked by the user is, optionally, drawn over the map as a green dotted line.

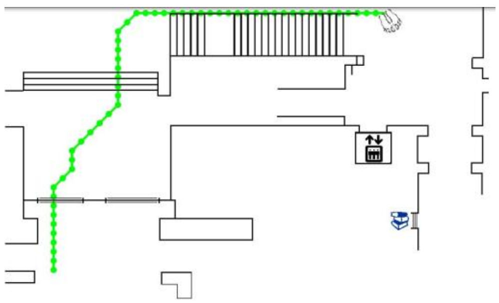

Figure 4. Indoor floor plan view.

CONCLUSIONS AND FUTURE WORK

In this paper we have proposed an affordable and low-cost location based system that works simultaneously in indoor and outdoor environments. The proposed system adapts automatically the interface depending on the user environment, and also depending on how the mobile device is being held.

To be able to obtain an accurate positioning while indoors, the proposed application uses information obtained while the user is outdoors, to adapt the positioning parameters specific to that user and also the type of movement the user is currently doing. These parameters are also adjusted when the user enters an indoor area, to compensate differences between indoor and outdoor movement.

Regarding future work, we intend to integrate a filtering and searching module, which would allow the user to search for a specific point of interest, and also to develop a mobile navigation assistant that directs the user to the requested point of interest.

Finally, we intend to perform an extensive user experiment, to be able to understand how much more accurate the system is while using the proposed automatic adaptation, when compared to the use of fixed positioning parameters. We also want to evaluate the proposed interface and use the NASA TLX [16] to understand the workload of the proposed approach on the users.

ACKNOWLEDGMENTS

The work presented here is based on research funded by the FCT - Fundação para a Ciência e Tecnologia (Portuguese Science and Technology Foundation) through the SFRH/BD/46546/2008 scholarship.

REFERENCES

1. Pahlavan, K., Li, X., and Makela, J-P. Indoor Geolocation Science and Technology. *IEEE Communications Magazine (Feb. 2002).* 112-118.
2. Jensen, B., Kruse, R., and Wendholt, B. Application of Indoor Navigation Technologies under Practical Conditions. In *Proc. of the WPNC'09*. (2009), 267-273.
3. Beauregard, S., and Haas, H. Pedestrian Dead Reckoning: A Basis for Personal Positioning. In *Proc. of the WPNC'06.*(2006), 27-36.
4. Pombinho, P., Afonso, A. P., and Carmo, M. B. Point if Interest Awareness Using Indoor Positioning with a Mobile Phone. In *Proc. of the PECCS 2011.* (2011), 5-14.
5. Hiyama, A., *et al*. Position Tracking Using Infra-Red Signals for Museum Guiding System. In *Proc. of the UCS'05.* (2005), 49-61.
6. Ghiani, G., *et al.* A Location-Aware Guide Based on Active RFID's in Multi-Device Environments. *Computer-Aided Design of User Interfaces VI*, Springer London (2009), 59-70.
7. Ikeda, T., *et al.* ComPass System: An Low Power Wireless Sensor Network System and its Application to Indoor Positioning. In *Proc. of the CSTST 2008.* (2008), 656-662.
8. Bruns, E., *et al*. Enabling Mobile Phones to Support Large-Scale Museum Guidance. *IEEE Multimedia, April 2007.* (2007), 16-25.
9. Ekahau Wi-Fi Tracking. http://www.ekahau.com.
10. Kourogi, M., *et al.* Pedestrian Dead Reckoning and its Applications. In *Proc. of the "Let's Go Out" Workshop in conjunction with ISMAR'09.* (2009).
11. Jiménez, A. R., *et al.* Indoor Pedestrian Navigation using an INS/EKF Framework for Yaw Drift Reduction and a Foot-Mounted IMU. In *Proc. of the WPNC'10.* (2010).
12. Glanzer, G., et al. Semi-autonomous Indoor Positioning Using MEMS-based Inertial Measurement Units and Building Information. In *Proc. of the WPNC'09.* (2009).
13. Dekel, A., and Schiller, E. DRec: Exploring Indoor Navigation with an Un-Augmented Smart Phone. In *Proc. of the MobileHCI'10.* (2010), 393-394.
14. Serra, A., Carboni, D., and Marotto, V., Indoor Pedestrian Navigation System Using a Modern Smartphone. In *Proc. of the MobileHCI'10.* (2010).
15. Silva, P., *et al.* Rubi: An Open Source Android Platform for Mobile Augmented Reality Application. In *MAR Workshop, MobileHCI'11.* (2011).
16. NASA TLX. http://human-factors.arc.nasa.gov/groups/TLX/

Ubiquitous personalization of a smartphone, used as a universal controller

Bojan Blazica
Xlab Research
Pot za Brdom 100, 1000 Ljubljana, Slovenia
bojan.blazica@xlab.si

Daniel Vladusic
Xlab Research
Pot za Brdom 100, 1000 Ljubljana, Slovenia
daniel.vladusic@xlab.si

Dunja Mladenic
AILab, Jozef Stefan Institute
Jamova 39, 1000 Ljubljana
dunja.mladenic@ijs.si

ABSTRACT

Universal, natural, ubiquitous and collaborative - these four words best describe our vision of a smartphone, used as a universal controller for human-computer interaction. Thus, we named the proposed framework UNUC. The idea is to enable people to use their smartphones to control and interact with different devices like computers, domestic appliances, public information spaces etc.. In other words, UNUC represents a 'communication hub' for interacting with a vast range of devices. Because of this, UNUC not only unifies interaction, but also opens up many possibilities for the personalization of interaction, which we discuss in this paper.

Author Keywords

Ubiquitous, collaborative, smart phone, personalized interaction, communication hub

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

The amount of interaction between humans and computers (machines) is increasing daily; a fact supported by the mere existence of terms as ubiquitous computing [1], pervasive computing [2] and ambient intelligence [3]. The vocabulary of interaction differs greatly depending on whether we are interacting with a personal computer, a public information panel or a coffee machine. Moreover, the device or machine we are communicating with, with its purpose and its technical characteristics, actually defines the interaction as a whole. While it is logical that the purpose of a machine defines the interaction the object is capable of handling, this is not the case for the machine's technical capabilities (which can be limited); first, these capabilities can represent limitations to interaction and second, for each object the user has to learn a new interaction language. To overcome this, a smartphone could be used as a universal means of communication for various objects. Besides providing a unified interaction language, the use of a smartphone also opens up vast design spaces in terms of adaptive personal interfaces.

In this paper we give an introduction to UNUC followed by a description of the options it offers for ubiquitous personalization.

UNUC DESCRIPTION

We are currently working on UNUC, a framework for a Universal Natural Ubiquitous and Collaborative controller (similar to the one presented in [4]). These properties stem from and are tightly connected to the construction of UNUC, which consists of three parts: a mobile application that gathers the user's intents, a 'server side' application in the device that implements the reaction of the device to the user's intents, and a protocol for communication between them. The controller is *universal* because it allows the user to communicate with different devices in a unified way, *natural* because it takes advantage of gestural interaction supported by multitouch displays and the plethora of sensors present in most smartphones, *ubiquitous* because it is be possible to use it with virtually any device capable of some sort of wireless connection (e.g. Bluetooth, WiFi) and *collaborative* because there is no constraint that would forbid multiple controllers to connect to the same device, thus allowing collaborative interaction.

IUI Workshop on Location Awareness for Mixed and Dual Reality LAMDa'12,

14, February 2012, Lisbon, Portugal.

Figure 1: UNUC used to manipulate an object in 3D space: the object's orientation is controlled by the phone's orientation obtained from the phone's sensors.

Generally, UNUC supports basic interaction tokens as pointing a cursor in 2D, selecting objects, choosing from a list etc., implemented by simply sending basic UI events (e.g. touch events) over a communication channel to the target device. The current state of implementation of UNUC is aimed at testing it as a 3D user interface, where the orientation of an object or of the camera (Roll-Pitch-Yaw) is bound to the relative changes in the phone's orientation (from the phone's gyroscope), while the speed of movement along the XYZ axes is controlled through the touchscreen as can be seen in Figure 1.

Personalization possibilities of UNUC

UNUC as a platform is designed around a smartphone, which means that it is able to connect to the internet, has substantial computing capabilities, knows the location of the user, has access to the user's social networks, can be used to make payments etc.. All these features offer immense possibilities for user modeling and personalization. We divide them in two categories: general (deductive) and specific (inductive) personalization.

General/deductive personalization

Similarly to deductive reasoning, deductive personalization tries to apply some general knowledge (theory) about the user to a specific situation. For example, suppose we know that the user is a fan of a certain football team (UNUC could infer that from the user's past behavior or his communications on social networks). When he opens the television, UNUC could automatically switch to a channel that broadcasts the game of his favorite team.

Specific/inductive personalization

On the other hand, there are specific situations that repeat themselves over and over again. By using machine learning methods UNUC can learn a model of these situations in order to handle them proactively. For example, the volume of the television could be automatically adjusted to the user's preferences. When using UNUC to control a coffee machine, the preferred drink could be placed at the top of the list of available drinks and the amount of sugar could be set automatically.

FUTURE WORK

Our work on UNUC will continue along two main lines: further development of the infrastructure that allows a smartphone to be used as a universal controller and the exploitation of UNUC for ubiquitous personalization as discussed in this paper. As the latter is still in an early stage, we hope to get some valuable feedback from the workshop on how to continue and where to direct our efforts.

ACKNOWLEDGEMENTS

This research was partly funded by the European Union, European Social Fund, Operational Program for Human Resources, Development for the Period 2007-2013 and by the Slovenian Research Agency.

REFERENCES

[1] Brian Epstein home page. http://www.epstein.org/brian/ambient_intelligence.htm

[2] Hansmann, U., Merk, L., Nicklous, M., and Stober, T. Pervasive Computing. Springer, 2003.

[3] Weiser, M. The computer for the 21st century. Scientific American 3, 3 (1991), 3–11.

[4] Vanderhulst, G., Luyten, K. Ubiquitous access to the internet of things. In *Proceedings of InterSO'11* (2011), 13–16