Rapid prototyping and evaluation of intelligent environments using immersive video

Pushpendra Singh, Hai Nam Ha, Patrick Olivier, Christian Kray, Zhiwen Kuang
Amy Weihong Guo, Phil Blythe, Phil James
Newcastle University
Newcastle upon Tyne, United Kingdom
p.l.olivier@ncl.ac.uk

ABSTRACT

A major problem in developing ubiquitous computing applications is the simulation of the required infrastructure and the environment in which they are to be deployed. Developers have to rely on either low-fidelity techniques (such as paper prototypes and mental walk-throughs) or simply wait for a full scale deployment. Similarly, even where an existing infrastructure exists preliminary evaluations of such applications is problematic due both to the logistics involved in coordinating subjects and experimenters on site, and the uniqueness of each interaction (it is impossible to reliably repeat specific visual, spatial and sensory contexts in an actual deployment). We propose an approach using immersive video with surround sound and a simulated infrastructure to create a realistic simulation of a ubiquitous environment in the software design and development office. We describe the motivations and design of our immersive video system and discuss our initial experiences of how it can be used to aid developers in the requirements capture, design, development and deployment of ubiquitous computing applications and intelligent environments.

General Terms

Human Factors, Experimentation

Keywords

Immersive video, rapid prototyping, evaluation, modeling.

1. BACKGROUND AND MOTIVATION

Development and implementation of ubiquitous computing applications require rapid prototyping and early evaluation. However, the infrastructure to support such application is often not available at the time of application development and developers have to wait for a full-scale deployment of sensors and other components before their prototypes can be deployed and evaluated. Ubiquitous computing environments are also more complex than other limited environments and require a direct interaction between users and system components which makes it hard to predict how a user is going to behave. While this makes implementation of an application difficult, it also means that the full potential of an ubiquitous computing application can not be assessed without a full implementation.

Research has shown that prototyping systems for studies is important to understand how a system is fit for purpose and will in practice be utilized. The degree to which an ubiquitous computing application can be useful depends on the ability of designer to understand and anticipate a user's needs. Prototyping helps in understanding these needs. The take-up of ubiquitous computing applications in everyday life is therefore dependent on the development of prototyping techniques that can provide useful insights into likely usage patterns.

Many ubiquitous computing applications depend upon the nature of a spatial interaction between a user and a physical environment, both in terms of a user's interest in the elements of the physical environment (e.g. landmarks in a mobile navigation application) and interactions between the user and sensors deployed either on the user or on objects in the environment (e.g. components of a positional tracking systems). To capture such spatial issues virtual reality models of a scenario have been proposed. Such systems utilize existing game engines in which a replica of a physical environment is constructed and augment this environment with virtual sensors.

The few reported deployments of virtual reality for prototyping ubiquitous applications have used game-like desktop simulations as high fidelity immersive visualisation of such models require large stereoscopic displays, such as in a CAVE, which are not widely available. However, although such virtual reality models are capable of capturing the spatial characteristics of a physical environment they inevitably lack the visual quality of real world. Furthermore, the process of creating a 3D model of a real scenario is well known to be a very time-consuming and costly process, and even the most sophisticated 3D designer is incapable of recreating real world experiences and unexpected events of environments such a shopping centers and airports.

Even where there is an underlying sensors and wireless network infrastructure many problems arise in relation to the evaluation of ubiquitous computing applications at the site of an actual deployment. Such processes are time-consuming and particular events and activities might only be intermittently available (e.g. such as the arrival of a particular plane or bus in intelligent transport systems). Evaluations of actual deployments typically also limit the experimenter's ability to elicit continuous feedback from users during the evaluation and preclude the possibility of recreating the same situation repeatedly to test problems that arise as a results of particular transient configurations of the user, world and infrastructure.

Thus, there is a pressing need for a tool which can facilitate both rapid prototyping and evaluation of computing ubiquitous applications within the software development office. Existing techniques, such as paper prototyping [4], suffer from their inherently low fidelity (in relation to both visual, physical and sensory phenomena) when applied to mobile and ubiquitous systems. These contextual factors (such as crowded environments, visual and aural impressions of a location) are hard to recreate or convey using standard methods. One possible solution is to use a Wizard of Oz methodology whereby a user's personal computer (typically a mobile phone or other handheld computer) is controlled externally by an experimenter (the wizard) to simulate its planned behavior [3]. A number of problems remain for such WOz-based approaches: they still requires participants/experimenters to physically visit the place where the system is to be used; subsequent trials will not take place under exactly the same circumstances (e.g. it might be raining on one occasion, while it might be sunny another time, crowd densities will vary, etc.).

2. THE IMMERSIVE VIDEO APPROACH

We propose the use of *immersive video*, to overcome limitations of applying traditional prototyping and early evaluation approaches to mobile systems, to capture the sensory experiences that we expect users to be exposed to at locations of deployment of location-based services. By capturing video (imagery and sound) at the site of the intended deployment of a location-based service and simulating the sensor infrastructure, system developers can have ready access (i.e. in their development office) to a high fidelity recreation of a user's experience of using a prototype mobile system on location.

2.1 Simulating sensory experience

Immersive video systems have been proposed on numerous occasions as an alternative form of entertainment multimedia (e. g. [5]), with the goal of extending the traditional constraints of film to provide the user with a sense that they are in some way truly immersed in an environment. Our system harnesses these techniques using one or more video cameras to capture synchronized footage at key locations within an intended usage scenario. Depending on the display facilities available to the designers, a wide field-of-view of the scene is captured (typically between 140-360°) and then replayed on multiple screens in front of (or within) which the user stands.

Footage captured at each location should be edited to create shots that can be looped to create a sense of being at a point in space and time, without the sensory disturbance that can result when observing a jump from the end of a shot back to the beginning (both in terms of the image and the sound). A *base shot* at each location is first created and a number of *event shots*, relevant to the location, may be added depending on the nature of the application under development. Depending on the field-of-view that has been captured, multiple base shots for a location may be created corresponding to a user standing in different orientations. Event shots correspond to activity that is typically salient to the application under development, for example, the arrival or departure of a bus in an intelligent transport system, or a change in the state of a pedestrian crossing signal in a mobile guide application.

2.2 Simulating Events

Base shot and event shots are organized into a simple state-based model of the environment. Two-way links are created between the base shots at a particular location, as transitions between these states corresponds to users reorienting themselves. One way links are formed between one or more of the base shots corresponding to spatially adjacent locations (according to the direction in which a user is allowed to move in the scenario that has been recreated). Finally, two way links are added between the base shots at a location and their associated event shots. A user's location and orientation in space, and their location in time (i.e. in relation to the events that may occur) thus correspond to their location in the state space of shots.

Physical movement (either reorientation or translation) is simulated by transitions between base states, and the passage of time at a location (either real or otherwise) by transitions from base shots to event shots (and back). These transitions, which correspond to the passage of a user in space or time, are not performed by the user but by a separate Wizard of Oz-style controller who drives that immersive video system in response to the user's verbal requests. Figure 2 shows state diagram for our application.

In our current system we have deliberately chosen this configuration in preference to augmenting the mobile device carried by the user with additional controls (which would do more to undermine the integrity of the users interaction with the device), or any hands free interaction technique, which would most likely be hard to deploy in a typical environment where software development is undertaken. Transitions between shots are realized using standard techniques from cinematography, such as cross-fading the image and sound of the source and destination shots, a visual style that is instantly recognizable, and readable, for users.

2.3 Simulating the technology infrastructure

Capturing the user's sensory experience of an environment is one component of our mobile services prototyping and evaluation system. The remaining component aims to simulate the sensor data corresponding to the base and event shots. During the creation of the state space of shots, XML-based sensor files are created for each shot specifying the values of different sensor readings that will occur at the corresponding time and location. This sensor abstraction layer allows the specification and simulation of GPS, Bluetooth, RFID and infrared inputs for each shot. The abstraction layer renders the origin of sensor data completely transparent to an application thereby allowing users to stand within the immersive video system and experience the behavior of an application exactly as if they were in such a context in a real sensor infrastructure. Figure 1 shows a fraction of a sample sensor file, with readings for GPS and related states that can be reached.


```xml
<?xml version="1.0" standalone="no" ?>
<ivideo>
<sensor stype="gps" value="" irid=""
   longitude="54.97750833333333"
   latitude="1.61345555555556"
   startframe="0" endframe="3000" />
<state name="station2_4" astartframe="0"
   aendframe="3000"
   centersrcfile="../AVI/c24z1.avi"
   leftsrcfile="../AVI/l24z1.avi"
   rightsrcfile="../AVI/r24z1.avi"
   xmlfile="../XML/s24.xml"
   istartframe="" iendframe="" />
<state name="station1_1" astartframe="0"
   ...
   ...
</ivideo>
```


Figure 1: XML file for one state showing sensor values and other possible states reachable from this state.

Crucially, from the user perspective, an application, and its user interface, will behave exactly as if they were in the actual location in a real-world deployment. The video and sound provide a rich impression of the actual space, and the large field of view heightens their sense of immersion as they interact with a device. As discussed above, control of a user's location in the state space is maintained by a wizard, a task that is facilitated by only having the allowable states available. We are currently using this approach to ease the evaluation and prototyping of mobile and ambient applications while enabling the discovery of context-related issues. Our experience has been that much of the fidelity of a user's real-world experience can be captured using the immersive video system and interface and application developers can very quickly generate prototypes and evaluate them using the immersive video.

2.4 Immersive video & requirements capture

An immersive video simulation of a mobile application scenario contributes to the development process in a number of ways [5]. Our primary goal was to create a mechanism by which the evaluation of such a system can be conducted both as early as possible in the development life-cycle and within a sensory context which the user will typically encounter when using the application. The use of immersive video addresses a number of additional problems in the development of mobile and ambient systems. Firstly, requirements capture for such systems is known to be significantly more difficult than for traditional static applications. By definition mobile systems will be used in a wide range of environments and the users of such systems will often be involved in multiple tasks simultaneously.

Capturing this richness of deployment is problematic using standard text or graphical formalisms and notations for characterizing use cases. Even seemingly rich approaches such as scenarios [1] and personas [2] fail to capture much of the richness of a real world deployment. In the light of this shortfall in the description of the requirements of mobile systems designers and developers struggle to understand the full implications of their design choices, and it is almost inevitable that major (and costly) interface design revisions will occur late in the process.

The process of developing an immersive video simulation goes some way to mediating this shortfall in the expressiveness of traditional approaches to requirements capture. On one hand the selection of the usage scenario, and the specific locations at which the filming is undertaken comprises an informal statement as to the anticipated context in which the system will be used, the range of relevant and incidental events that occur at different locations, and the data (i.e. sensor data) available to the mobile device. On the other hand the footage itself explicitly captures features of the environment that are impossible to include in a traditional scenario or use case description, such a the level and nature of ambient sound, the degree of clutter and motion in the user's visual field, and the precise spatial characteristics of important features in an environment.

3. INITIAL EXPERIENCES

We have developed the immersive video as a prototyping tool as part of continuing ubiquitous computing research at our institute. Presently we have three projects which demand a robust prototyping tool – each project has additional requirements that makes the tool particularly appropriate. In this section we will demonstrate the system by illustration of its application to ASK-IT, a European Union Framework 6 project that aims to enhance inclusion through the application of ambient intelligence. One goal of ASK-IT is to provide mobility impaired users with seamless location sensitive guidance and information as they arrive at a metro station, exit and navigate to a cultural location such as a museum, and enter the museum and view its exhibits. The user interface is primarily map-based but also provides textual information on exhibits in a museum. Having an evaluation tool with which views can be elicited without exposing users to a real environment is particularly useful for such a vulnerable user group.

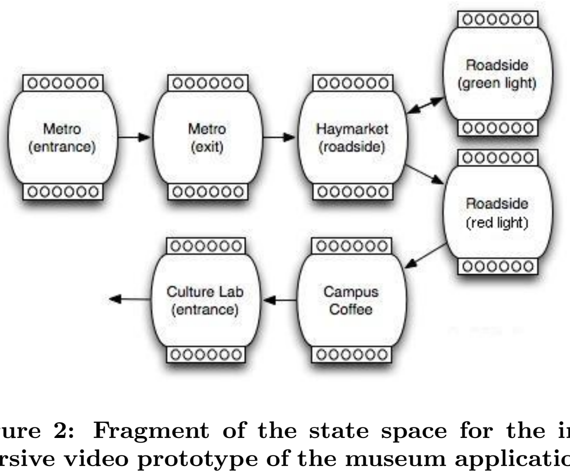

Figure 2: Fragment of the state space for the immersive video prototype of the museum application.

For ASK-IT we have prototyped a GPS-enabled PDA, Blue-

Figure 4: Immersive video in action: a user interacting with a mobile device while being surrounded by a video replay of the location, where it is intended to be used

tooth indoor positioning, RFID e-ticketing and infrared beacons for pointing gestures. The actual application works in following manner: when a user arrives at metro station the PDA detects the platform using Bluetooth positional sensing and displays a map of the platform with directions to the escalator and lift. As the user exits the metro station a GPS signal is detected and a map of surrounding area is displayed with direction information towards the museum. A panoramic photograph (covering a 360° view) at the top of the display is used to help the user identify the direction he or she needs to walk.

When user enters into the museum, the PDA detects the new location (using Bluetooth positional sensing at the entrance to the museum) and displays a map of museum. The user is also issued with an e-ticket for the museum (using RFID tag attached to PDA) and information pertaining to the e-ticker is displayed on PDA. Within the museum, when user points the PDA at an exhibit, a supplementary multimedia presentation relating to that exhibit is displayed on the device. The detection of which exhibit the user is pointing the device at is implemented using infrared beacons. Figure 3 shows an example of the actual user interface displayed on the PDA. Immersive video of the environment was used in both the development and initial evaluation of the applications. Figure 2 show a fragment of the shot state space used.

During development, it reduced considerably the need to visit the test site which would have been both time-consuming and also required the installation of infrastructure elements, such internal Bluetooth devices at the platform of the metro station and in the museum. This gave developers considerable freedom to develop their application without committing either to the class of infrastructure devices to be used or their detailed configuration on site. Possible users were also able to evaluate the system at an early stage and highlighted some useful interface design issues (such as interaction with the panoramic images). Finally, with the aid of the demo, participants were able to gain first hand experience of navigating the simulation of the space and the behavior of the interface. Figure 4 shows our local set-up using a three walled CAVE-based display. However, a similar impact can be achieved by projecting video on walls of a corner of office-room.

4. FUTURE WORK

In future, we plan to use immersive video as a standard prototyping tool other ubiquitous computing based projects. Two major projects, on which we are current working are (i) real-time 3D visualization of urban data for emergency response and (ii) the assessment of future traveler information systems. The emergency response scenarios project involves the development of a mobile device that provides real-time 3D visualizations of urban data for emergency response scenarios. Requirements capture for the application include working closely with police, fire and other emergency services. This means that application development needs to include input from different user groups. Also, access to information related to emergency facilities such as hospitals is

Figure 3: User interface on mobile device: interactive panoramic picture (top), map with route (middle), and menu-bar (bottom)

required and then evaluation of such an application will be difficult due to its nature of emergency response scenarios.

Here the immersive video offers significant benefits in overcoming these problems. Limited access to important buildings such as hospitals, police or fire headquarters can be overcome by videoing them once and annotating it with all the information that will be available. Immersive video will also allow easy evaluation of application by different user groups because they can evaluate it in their office-space and can provide input. Video of real environment will provide easy means of assessing a user's ability to visually register the 3D graphics with the environment without having to be on-site.

Our future traveler information systems project is a response to the observation that increased car use in many urban areas has led to growing traffic congestion, which not only threatens economic growth but also results in poor air quality, noise and global warming. A range of marketing and management measures have been piloted in various areas in the UK including travel awareness campaigns, workplace and school travel plans, personalized travel planning and public transport information. These measures usually seek to offer better information and opportunities to influence the choice of transport mode made by individuals, and reduce the level of car use. These measures can have a significant effect on individual travel choices when individuals or particular groups are targeted and tailored information is provided based on individual journey characteristics.

The key problem relating to the provision of better information is that the locations and contexts that travelers find themselves in are constantly changing, thus their information needs are both immediate and highly context sensitive. However, a large-scale implementation that collects a full set of travel-related information in real-time, and disseminates such information in a context aware manner, is not possible within the existing information systems infrastructure. A future traveler information system (FTIS) will have relevant environmental information, live transport and traveler information, and will compile them proactively with stored historical information. Such a system will therefore always offer customized and timely travel-related information based on traveler context awareness. Such a scenario raises two important questions:

- How will people perceive such proposals for future traveler information systems?
- Does the proposed future traveler information system have the potential to bring about a significant modal shift?

The ideal way to explore the behavioral change that results from FTIS use is to allow users to interact with such systems as part of their everyday travel routine, such as commuting, shopping and recreation. However, FTIS supported by pervasive computing is still very much an aspiration as large scale deployments of pervasive computing environments do not exist. Building a physical mock-up of FTIS is expensive and time-consuming, it will also cover a large geographical area (usually a city) which will make evaluation of such a system cumbersome.

Immersive video will be very helpful in creating a realistic sense of a user's experience of an application and allow participants to interact with a simulated system. It will help in development and evaluations of applications related with FTIS. It will also be able to stimulate the experience of interacting with an FTIS by providing participants with a sense that they are in some way truly immersed in an environment where the FTIS has been implemented.

The utility of immersive video can be clearly visualized with respect to the projects mentioned above. For the development of immersive video as a standard prototyping tool, we aim to develop a range of applications to make it easier to build an immersive video environment after taking the videos. Two major components of immersive video system are:

1. Wizard: the wizard controls the sequence of video clips based on the state diagram as explained in section 2.2. Arranging clips into right sequence order and then feeding them to wizard is quite time-consuming.
2. Sensor information: we are using XML files to store the sensor information about every clip. Though this approach is general it is cumbersome to create individual XML files and make changes in them.

Furthermore a number of interaction modes within the immersive video system need to be explored:

1. User interaction:
   - How to move backward or forward in a scene (without intervention from the Wizard)?
   - How to select objects available in a scene (for enhanced interactivity)?
   - Physical interaction techniques, such as opening a door?
2. System interaction: presently only the wizard injects events according to its observation of the user's behaviour. But events can be generated based on the level and nature of interaction that a user is engaged in. Events are also not only the result of a user triggering sensors but a result of external events too, such as plane arriving, receiving a call.

5. CONCLUSION

A key issue in mobile and ambient computing is the effort required to rapidly prototype and evaluate user interfaces and applications. Existing technologies for these tasks suffer either from low fidelity (e. g. paper prototypes, mental walkthroughs) or effectively require a near full-scale deployment. In this paper, we proposed our approach using immersive video and a simulated infrastructure to create a realistic environment in the office or lab. Immersive video provides a low-cost and rapid means to prototype user interfaces and applications, and to evaluate them in a realistic simulation of the context, in which they are intended to be used. Our approach also allows for user studies in the lab, where participants can be easily exposed to specific environment conditions in a realistic and repeatable manner. We described our inital experiences and future plans of using immersive video for a range of problems and at the same time developing tools to make immersive video an easy-to-use rapid prototyping tool for ubiquitous computing based applications.

6. REFERENCES

[1] J. Carroll. *Making Use: Scenario-Based Design of Human-Computer Interactions*. MIT Press, 2000.

[2] A. Cooper. *The inmates are running the asylum.* SAMS/Macmillan, 1999.

[3] M. T. J. H. Kaj Makela, Esa-Pekka Salonen and R. Raisamo. Conducting a wizard of oz experiment on a ubiquitous computing system doorman. In *Proceedings of the International Workshop on Information Presentation and Natural Multimodal Dialogue*, pages 115–119, 2001.

[4] L. Liu and P. Khooshabeh. Paper or interactive?: a study of prototyping techniques for ubiquitous computing environments. In *CHI '03: CHI '03 extended abstracts on Human factors in computing systems*, pages 1030–1031, New York, NY, USA, 2003. ACM Press.

[5] S. Moezzi, A. Katkere, D. Y. Kuramura, and R. Jain. Immersive video. *vrais*, 00:17, 1996.