Christian Kray · Patrick Olivier · Amy Weihong Guo · Pushpendra Singh · Hai Nam Ha · Phil Blythe

Taming Context: A Key Challenge in Evaluating the Usability of Ubiquitous Systems

Abstract A key defining property of ubiquitous computing is the notion of context. Ubiquitous systems have to sense changes in their context and adapt to it in a meaningful way. While this feature can greatly benefit the user, it also poses a challenge in terms of evaluating and comparing systems that behave differently when exposed to changes in their context. How can experimenters reliably determine whether a specific effect or behavioural pattern is caused by system properties, a particular state of the context or a combination of both? How are they to compare two systems when much of the user experience is determined by the context? In this paper, we discuss the use of *immersive video* as a technique to address some of the problems in evaluating ubiquitous systems resulting from context sensing and adaptation. Immersive video creates a very realistic audio-visual impression of real-world locations that we combine with a simulation of the sensor data. As a result, this technique enables a high degree of control over the context while providing users with a realistic environment in which to interact with a ubiquitous system.

Keywords Context awareness · Immersive Video

1 Introduction

In his often cited paper "The Computer of the 21st Century", Weiser introduced the idea of ubiquitous computing, a new way to interact with a large number of computers that are embedded in the environment. A key property of the systems he envisioned is that they "weave themselves into the fabric of everyday life until they are indistinguishable from it" [24]. Consequently, a ubiquitous system has to be aware of its *context* and adapt to it in order to seamlessly provide the kind of service that is appropriate for a particular user at a particular location in a particular situation.

Context is an inherently vague concept (oftentimes described as "anything but the main object of concern that is of relevance for the object"), and the term "context" is used in a variety of disciplines (e. g. linguistics or theoretical computer science) to denote different concepts. A commonly used definition in the field of ubiquitous computing was introduced by Dey and Abowd. They define context as "any information that can be used to characterise the situation of an entity, where an entity can be a person, a place or physical or computational object" [5].

From the perspective of a ubiquitous application, contextual information can be derived from all current and previous readings returned by the available sensors (including recorded interactions with users). Oftentimes, there will be a context model that allows for deriving a more abstract description of context (e. g. "it's dark" instead of a number of readings from a set of light sensors). Naturally, such a model will always represent only a subset of the "full" context as experienced by users engaged in activities in the real world. From the perspective of human users, context is a much richer concept including, for example, the audio-visual impression the current environment, their emotional state, the presence of other people and their relationship with them.

The complexity and importance of context do not only pose a significant challenge for developers of such systems but also raise the bar for the evaluation ubiquitous applications as the context frequently is highly dynamic. For example, how can experimenters reliably determine whether a specific effect or behavioural pattern is caused by system properties, a particular state of the context or a combination of both? How are they to compare two systems when much of the user experience is determined by the context, which is highly dynamic?

In this paper, we discuss a novel evaluation technique, which uses *immersive video,* and which can be applied to evaluate and compare the usability of ubiquitous systems while providing both a rich context and tight control of

C. Kray, P. Olivier, A.W. Guo, P. Singh, H.N. Ha, and Phil Blythe
Newcastle University, Newcastle upon Tyne, United Kingdom
E-mail: c.kray@ncl.ac.uk

contextual factors. Before we introduce our approach in detail, we will briefly discuss related work in Section 2. Both the ideas underlying immersive video and the technical realisation will be presented in Section 3. We will also report on initial experience we have gathered so far from applying our approach (Section 4). The paper concludes on a summary of the main contributions and an outlook on future work.

2 Related work

Due to the central role context plays in ubiquitous computing, there is a vast body of work looking into the definition, acquisition, and modelling of as well as the adaptation to context. In addition to the definition mentioned in the previous section [5], others have made a point of highlighting that it is more than just location [19]. Nevertheless, the acquisition and modelling of location data has been a very prominent subject in ubiquitous computing in general and in research on context-awareness in particular [23, 9, 7]. The spectrum of ubiquitous applications that adapt to their context is very broad, ranging from tourist guides [3] over information displays [13] to generic toolkits [5, 11] that can be used for variety of purposes.

A large number of ubiquitous applications have been exposed to evaluation. Since the type and purpose of these vary widely, it comes as no surprise that this is mirrored in the choice of evaluation methods. At a very abstract level, we can distinguish methods that are more aimed at evaluating the technical performance of a system, e. g. its maximum throughput or average error rate, and methods that are aimed at evaluating the experience of actual users, e. g. whether people can complete a task faster or with fewer errors when they use a ubiquitous system.

Methods for evaluating technical aspects of a system include, for example, measuring system performance in categories such as number of errors, speed and accuracy of the context model (e. g. when compared to a known ground truth) or whether or not a system is able to detect a change in context (cf. e. g. [18, 7, 8]). While methods aimed at evaluating technical aspects oftentimes allow for direct comparison (e. g. how accurately or precisely it can sense the location of an entity), there are currently no widely accepted benchmarks or benchmark scenarios such as, for example, the RoboCup competition in robotics [22] (although there are some initial suggestions [16]).

In order to evaluate ubiquitous applications with human users, other methods need to be applied. In human-computer interaction, evaluation methods are oftentimes divided into "lab-based" studies and "in-the-field" studies, and there is a heated debate which one is more useful when analysing context-aware systems [10]. Traditional lab-based studies have the advantage of providing experimenters with a great degree of control over all aspects, e. g. the exact time available to complete a task. They also allow for very precise and detailed measurements such as tracking eye movement or reaction time, and consequently are well suited to repeat the same experiment under the exact same conditions. Usability studies in the field enable subjects to use the system in the actual environment, where it is meant to be used. Such studies are thus more likely to reveal the impact of contextual factors. Typical methods used for in-the-field evaluation include observational studies [1], long-term deployment [4] and cultural probes [15]. Due at least partially to the wide range of methods, there is also no set of widely accepted benchmarks or benchmark scenarios that would facilitate the comparison of different ubiquitous systems in terms of their usability.

Both lab-based and in-the-field studies suffer from a number of disadvantages that add to the difficulty of establishing common benchmarks that capture "real" user behaviour. In the case of lab-based evaluation, a key problem relates to the provision of context: as studies take place in a laboratory (and not at a location where a system is intended to be used), subjects have to either ignore context or are asked to imagine it. This obviously has implications regarding how well the influence of context can be captured using a lab-based approach. In-the-field studies suffer from the opposite problem: as they are conducted "in-the-wild", there is little control over the context. Hence, the context can and will change in unexpected ways so that even comparisons/evaluation within one experiment can be tricky (not to mention comparisons across experiments and systems). In addition, the logistics involved in running an in-field studies can be prohibitive, e. g. in meeting the requirement for legal permissions, having to fully deploy a system prior to the study (potentially), and having to instrument the environment in order to take measurements. In the following section, we will introduce an evaluation technique using immersive video that addresses many of these problems by combining aspects of lab-based and in-field studies.

3 Immersive Video

In order to overcome the shortcomings that emerge when applying traditional evaluation approaches to mobile and ubiquitous systems, we propose the use of *immersive video* [14] to capture the sensory experiences that we expect users to be exposed to at the intended site(s) of use. By capturing video (imagery and sound) at these locations and by simulating the sensor infrastructure, we are able to produce a very realistic recreation of a the actual environment in the real world. Figure 1 shows a prototypical system we have developed being used in our CAVE (an immersive 3D visualisation installation).

Fig. 1 A user interacting with a mobile device while being surrounded by an immersive video replay of the location, where it is intended to be used.

3.1 Basic Idea

Immersive video systems have been proposed on numerous occasions as an alternative form of multimedia entertainment (e. g. [14]). Our system [20] harnesses these techniques using one or more video cameras to capture synchronised footage at key locations within an intended usage scenario. Depending on the projection 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. In addition, the audio recorded at the location being displays is replayed. Within this realistic recreation of the actual location, users can then interact with a mobile or ubiquitous system.

3.2 Capture

Footage captured at each location must be edited carefully 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 created first 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. For the case studies we have run so far, we used a three-camera set-up (shown in Figure 2) that we then replayed in our CAVE suite using three screens in a roughly semicircular configuration (left and right screen at a 45°angle to central screen).

Fig. 2 Camera configuration used to capture footage for immersive video (used with permission).

3.3 Spatiotemporal Modelling

In order to link a set of scenes and to enable meaningful interaction, we organise base shots and event shots into a simple state-based model of the environment. Two-way links are created between the base shots at a particular location, since 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. The resulting model hence describes the relationship between locations and events both in space and time but also enables interaction by a user during an experiment (e. g. movement). Although we have used a static model for all studies we have conducted so far, it is a straightforward task to make it more dynamic, e. g. by introducing timed changes or by modifying the model in response to user interaction.

3.4 Simulating Sensor Data

Although mostly invisible to the user, the data that is gathered by sensors is an important ingredient defining the user experience in ubiquitous systems. In order to provide a "realistic" stream of data, we use a simple data structure that is linked to the state space and defines what each sensor can "see" at that location. In addition, we have developed a sensor abstraction toolkit application developers can use. It can transparently use either real sensor data recorded in the field or simulated data that is replayed in the immersive environment. From the application's point of view it is completely transparent whether it is being used in the field or in conjunction with the immersive video facility.

3.5 Application

Once the footage, the state space and (possibly) a sensor simulation have been created, it is possible to run experiments using a control software we developed. This software manages the playback of footage on the screens and records the movement and location of a subject within the state space. It also provides an interface for the experimenter to control the motion within the state space, allowing for Wizard of Oz style studies. For example, in response to a user action (such as pressing a button on a handheld device or a voice command) the experimenter could make the system change to a different state, which is linked to footage showing the system's reaction. In addition, the experimenter is free to inject events at any point, e. g. triggering the arrival of a bus in an application evaluating a mobile assistant for the public transport system.

Physical movement of the subject (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). Transitions are realised 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 recognisable, and readable, for users. Although it is possible to trigger transitions directly in response to actions of a user, we have mainly used the Wizard of Oz approach so far, i. e. the experimenter controls movement within the state space in response to verbal requests of the subject. Figure 3 shows a photograph taken during a recent study depicting the experimenter (on the left) and a subject.

3.6 Evaluation

Since studies using immersive video take place in the lab, it is in principal possible to apply any evaluation technique used in traditional settings, including eye-tracking or other physiological sensors. In addition, some methods usually employed for in-the-field studies (such as observational or ethnographic methods) can potentially be applied as well as a very realistic environment is provided. In the initial case studies we have conducted so far, we used a combination of questionnaires, observations and free-form feedback.

Fig. 3 Photograph taken during a recent study showing experimenter and a subject in the immersive video environment (used with permission).

4 Initial experience and discussion

We have used the immersive video approach to evaluate two different systems: a mobile assistant providing support for disabled people [21] and a prototype of a future travel information system for public transport [6]. In both cases, feedback from participants was very positive. Several commented that the realism was making them feel almost as if they were in the actual location. The transitions used to simulate physical movement were also received well, and the idea was readily understood by most participants. Although we did not conduct any studies yet focussing on the actual experience and perception of immersive video on its own, there is initial evidence that people find it a believable simulation of the actual location.

In addition, the studies brought to light some issues that would have been hard to uncover using traditional lab-based studies. These include the use of references to objects in the environment (and their audiovisual properties) and how they relate to the application/interface being evaluated. For example, the mobile assistant used in the first study used a panoramic picture of locations to help people in identifying the direction in which they had to go. Using the immersive video approach, we discovered that subjects had issues with mapping this panoramic photograph to the actual environment. Spotting this problem was greatly facilitated by the rich visual environment provided by the immersive video, which enable participants to actually perform the mapping task rather than just imagining it.

While the immersive video approach is still being developed and further research into a number of aspects is needed, one of its key application areas relates to the comparison of mobile and ubiquitous systems. By creating immersive video, a sensor simulation/model and the corresponding state-space, it is possible to repeat the exact same experiment with the same context. Hence, in order to compare two context-aware applications, they can be tested under the same conditions and thus be compared in a number of categories such as user performance, error rates or other user-related aspects.

For example, let us assume we wanted to compare different systems providing navigation support, e. g. one using dynamic signage [12], one using maps displayed on a mobile device [2], and one using cross-modal cues [17] to convey directions to their users. Once an immersive video environment has been created (as described in Section 3), it is possible to have participants perform the same navigation tasks (such as getting from A to B) while being exposed to the same context conditions but using different systems. Consequently, this method will yield results that can be compared across different systems, and thus enables, for example, the identification of the system that causes most navigation errors or the one that results in the fastest performance.

Immersive video environments can therefore serve as a basis upon which to establish standard benchmark tests for ubiquitous systems that involve users. Once a number of relevant scenarios have been identified (such as navigation support or information access), reusable immersive video environments can be created to compare the performance of different systems built to address tasks within those scenarios. In other areas such as robotics, the availability of standard benchmarks and competitions has led to great advances (cf. e. g. [22]). It will be interesting to see whether the same can be achieved in the area of usability evaluation of ubiquitous systems.

5 Conclusion

In this paper, we discussed the idea of using immersive video to evaluate mobile and ubiquitous systems. We have analysed some shortcomings of traditional lab-based and in-the-field studies, and pointed out how immersive video can overcome some of those. We described the approach we have taken to produce and apply immersive video in evaluating systems with users, and we provided some evidence of its benefits based on experience gained in two user studies. In addition, we outlined how this technique can be used to construct benchmarks for mobile and ubiquitous systems that allow for a reproducible and reliable comparison of different systems.

In the future, we will investigate further improvements to the realism of the techniques (such as enabling actual physical movement) as well as to the underlying model (such as making it more dynamic). We will also design and conduct user studies to directly compare this approach to in-the-field studies as well as to lab-based studies in order to identify which method is best suited to capture particular aspects of usability.

References

1. L. Arnstein, C.-Y. Hung, R. Franza, Q. H. Zhou, G. Borriello, S. Consolvo, and J. Su. Labscape: A Smart Environment for the Cell Biology Laboratory. *Pervasive Computing Magazine*, 1(3):13–21, July-September 2002.
2. J. Baus, C. Kray, and K. Cheverst. A survey of map-based mobile guides. In L. Meng, A. Zipf, and T. Reichenbacher, editors, *Map-based Mobile Services*, pages 197–216. Springer, Berlin, Heidelberg, New York, 2005.
3. K. Cheverst, N. Davies, A. Friday, and C. Efstratiou. Developing a context-aware electronic tourist guide: Some issues and experiences. In *Proceedings of CHI 2000*, pages 17–24, Netherlands, 2000.
4. K. Cheverst, A. Dix, D. Fitton, A. Friday, and M. Rouncefield. Exploring the utility of remote messaging and situated office door displays. In *Proceedings of MobileHCI 2003*, pages 336–341, Berlin, Heidelberg, New York, 2003. Springer.
5. A. K. Dey and G. D. Abowd. The context toolkit: Aiding the development of context-aware applications. In *Proceedings of the Workshop on Software Engineering for Wearable and Pervasive Computing*, Limerick, Ireland, June 6 2000.
6. A. W. Guo, P. Blythe, R. Fairchild, P. Olivier, P. Singh, and H. N. Ha. Using immersive video to evaluate future traveller information system. In *Proceedings of the 14th World Congress on Intelligent Transport Systems*, Beijing, China, 9-12 October 2007.
7. A. Harter, A. Hopper, P. Steggles, A. Ward, and P. Webster. The anatomy of a context-aware application. *Wireless Networks*, 8(2–3):187–197, 2002.
8. M. Hazas, C. Kray, H. Gellersen, H. Agbota, G. Kortuem, and A. Krohn. A relative positioning system for spatial awareness of co-located mobile devices and users. In K. G. Shin, D. Kotz, and B. D. Noble, editors, *Proceedings of the third international conference on mobile systems, applications, and services (MobiSys 2005)*, pages 177–190, New York, 2005. ACM Press.
9. J. Hightower, B. Brumitt, and G. Borriello. The Location Stack: A Layered Model for Location in Ubiquitous Computing. In *Proceedings of the 4th IEEE Workshop on Mobile Computing Systems & Applications (WMCSA 2002)*, pages 22–28, Callicoon, NY, USA, 2002.
10. J. Kjeldskov, M. B. Skov, B. S. Als, and R. T. Høegh. Is it worth the hassle? exploring the added value of evaluating the usability of context-aware mobile systems in the field. In S. A. Brewster and M. D. Dunlop, editors, *Mobile HCI*, pages 61–73, Berlin, Heidelberg, New York, 2004. Springer.
11. G. Kortuem, C. Kray, and H. Gellersen. Sensing and visualizing spatial relations of mobile devices. In P. Baudisch, M. Czerwinski, and D. Olsen, editors, *Proceedings of UIST 2005*, pages 93–102, New York, 2005. ACM Press.
12. C. Kray, G. Kortuem, and A. Krüger. Adaptive navigation support with public displays. In R. S. Amant, J. Riedl, and A. Jameson, editors, *Proceedings of IUI 2005*, pages 326–328, New York, 2005. ACM Press.
13. J. F. McCarthy, T. J. Costa, and E. S. Liongosari. Unicast, outcast and groupcast: Three steps toward ubiquitous, peripheral displays. In G. D. Abowd, B. Brumitt, and S. A. N. Shafer, editors, *Proceedings of Ubicomp 2001*, pages 332–345, Berlin, Heidelberg, New York, 2001. Springer.
14. S. Moezzi, A. Katkere, D. Y. Kuramura, and R. Jain. Immersive video. In *VRAIS '96: Proceedings of the 1996 Virtual Reality Annual International Symposium (VRAIS 96)*, page 17, Washington, DC, USA, 1996. IEEE Computer Society.
15. E. D. Mynatt, J. Rowan, S. Craighill, and A. Jacobs. Digital family portraits: supporting peace of mind for

extended family members. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 333–340, New York, NY, USA, 2001. ACM Press.

16. L. Nazhandali, M. Minuth, and T. Austin. Sensebench: toward an accurate evaluation of sensor network processors. In *Workload Characterization Symposium, 2005. Proceedings of the IEEE International,*, pages 197–203. IEEE Computer Society Press, 2005.
17. P. Olivier, H. Cao, S. W. Gilroy, and D. G. Jackson. Crossmodal ambient displays. In *British Computer Society HCI: Engage*, Queen Mary College, University of London, UK, 2006.
18. N. B. Priyantha, A. Chakraborty, and H. Balakrishnan. The Cricket location-support system. In *Proc. of the Sixth Annucal ACM International Conference on Mobile Computing and Networking*, pages 32–43, New York, 2000. ACM Press.
19. A. Schmidt, M. Beigl, and H.-W. Gellersen. There is more to context than location. *Computers and Graphics*, 23(6):893–901, 1999.
20. P. Singh, H. N. Ha, Z. Kuang, P. Olivier, C. Kray, P. Blythe, and P. James. Immersive video as a rapid prototyping and evaluation tool for mobile and ambient applications. In *MobileHCI '06: Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*, pages 264–264, New York, NY, USA, 2006. ACM Press.
21. P. Singh, H. N. Ha, P. Olivier, C. Kray, Z. Kuang, A. W. Guo, P. Blythe, and P. James. Rapid prototyping and evaluation of intelligent environments using immersive video. In *Proceedings of MODIE workshop at Mobile HCI'06*, Espoo, Finland, 12-15 September 2006.
22. The RobCup Federation. Robocup home page. http://www.robocup.org/ (last checked 27 May 2007).
23. R. Want, A. Hopper, V. Falcao, and J. Gibbons. The Active Badge location system. *Transactions on Information Systems*, 10(1):91–102, 1992.
24. M. Weiser. The computer of the 21st century. *Scientific American*, pages 94–100, September 1991.