

Ubiquitous Collaborative Activity Virtual Environments

Aryabrata Basu
University of Georgia
basuarya@uga.edu

Andrew Raij
University of South Florida
rai@usf.edu

Kyle Johnsen
University of Georgia
kjohnsen@uga.edu

ABSTRACT

We introduce a new paradigm of collaborative computing called the Ubiquitous Collaborative Activity Virtual Environment (UCAVE). UCAVEs are portable immersive virtual environments that leverage mobile communication platforms, motion trackers and displays to facilitate ad-hoc virtual collaboration. We discuss design criteria and research challenges for UCAVEs, as well as a prototype hardware configuration that enables UCAVE interactions using modern smart phones and head mounted displays.

Author Keywords

Virtual environment, collaborative, mobile, ubiquitous, virtual reality, display

ACM Classification Keywords

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

General Terms

Human Factors; Design; Measurement; Experimentation.

INTRODUCTION

Collaborative virtual environments (CVEs) enable spatially distributed users to work together on shared tasks in a real-time artificial reality. While a variety of interfaces exist for CVEs, researchers have noted that immersive interfaces offer unique interaction affordances and potential benefits to collaborative work [9]. However, immersion is difficult to achieve in a widespread, practical manner. The “work to make it work” has been substantial; sufficient equipment has only been available in dedicated research and development laboratories, and these conditions limit widespread deployment [5, 9]. In this paper, we present the ubiquitous collaborative activity virtual environment (UCAVE) concept, which enables immersive virtual collaborations with minimal, portable infrastructure.

An immersive interface to a CVE most commonly includes technology for natural viewing of a virtual environment and control over an avatar embodiment. Examples include the immersive interfaces provided by the DiVE and Massive

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

CSCW'12, February 11–15, 2012, Seattle, Washington, USA.
Copyright 2012 ACM 978-1-4503-1086-4/12/02...\$10.00.



Figure 1 A user operating a UCAVE implementation inside and outside featuring a lightweight head-mounted display, smart phones and immersive 3D rendering.

CVEs [4], such as head-mounted and spatially immersive displays with user body tracking for avatar control. This is in contrast to non-immersive interfaces, which typically employ fixed planar displays and indirect control schemes such as a keyboard, mouse and on-screen widgets.

While non-immersive interfaces are highly accessible to end-users and support a wide range of interactions, the viewing and control schemes are noted to be frustrating and perhaps misleading in some collaborative work situations [9]. For example, where an avatar is gazing is not necessarily where a user is gazing. Immersive interfaces are often suggested to counter such effects, where the user is directly controlling the avatar and seeing what others perceive the avatar to see [9, 10]. However, the practical reality of immersive interfaces includes high cost, visual quality tradeoffs, extensive setup, user encumbrance, and limited deployment possibilities.

The UCAVE concept presented here offers a vision that this reality is changing; that immersive interfaces to CVEs are possible and can be as useable, portable and inexpensive as non-immersive interfaces. The UCAVE aims to make an immersive interface a choice that is dictated by application requirements and not by logistical limitations.

UCAVE Applications

UCAVEs are targeted at different applications than those of traditional immersive CVEs. The targeted applications are more in-line with collaborative mobile applications, where an immersive CVE being quickly and easily available to users may be enticing.

There is a growing trend towards video conferencing on mobile devices. We believe that UCAVEs can add a new dimension to this application, particularly in the case of multi-party conferencing, where conversational gaze cues are misrepresented in video. When necessary, such as when it is important for conversational participants to share the same space, participants could switch into an immersive viewpoint via the UCAVE hardware. There would be a tradeoff of the loss of some facial expression, but gesture and gaze cues could be maintained. In addition, it is then possible to support concepts such as Transformed Social Interaction [2]. Studying such ad-hoc immersive encounters will also be of interest to social psychology researchers, where the UCAVE could scale the scope of experiments to dozens or even hundreds of users.

The mobility and convenience of the UCAVE also offers new opportunities for CVEs to be applied in locations outside of dedicated laboratories. For example, consider an architectural firm designing a new building. To use a traditional immersive CVE, each architect and client would need to have dedicated laboratory space. With the UCAVE, multiple users can participate in an immersive design review from the convenience of their own space, without the expensive, encumbering, and large equipment commonly found in immersive CVEs.

BACKGROUND

A set of important, interrelated events has occurred that has made the UCAVE vision possible, where it would have been unlikely before. First, there is a rather sudden ubiquity of mobile computational power and high-speed mobile communication. The Gartner group estimates that 468 million smart phones will be purchased in 2011 [8]. These ultra-portable computers are interconnected on networks capable of supporting CVEs, and have the computational power to render complex 3D environments, largely owing to the popularity of mobile 3D gaming. Second, head-mounted displays (*e.g.* Vuzix Wrap 920) have emerged to support large-screen viewing and privacy for consumers using such mobile systems. Such head-mounted displays have substantially lower field-of-view, but are lighter, oriented towards casual use, and are a small fraction of the cost of professional quality head-mounted displays. Finally, motion-control gaming devices have made motion-tracking devices inexpensive and ubiquitous. While largely based on inertial tracking, there are also consumer-level mechanical, electromagnetic and optical tracking systems available. These, coupled with the ubiquity of motion control sensors already in smart phones *form the hardware basis for portable immersive virtual reality systems and the UCAVE concept.*

That is not to say that the UCAVE is the first concept of a portable immersion system. Mobile augmented reality (AR) systems have existed for over a decade, using backpack-worn computers, head-mounted displays and tracked controllers [11]. The field is also moving towards smart phone platforms, which support augmented reality applications using built-in inertial sensors, global positioning systems, cameras, and displays [12]. However, collaborative AR systems have fundamentally different application domains and technical challenges than CVEs.

Collaborative AR applications are local collaborations, as the point of AR is to see virtual information fused with the real environment, while CVEs are targeted towards applications for geographically distributed users. Since users are geographically distributed, CVEs usually incorporate avatars for a number of reasons, including network bandwidth control and for a 3D user representation [3]. Thus, a fundamental challenge to CVEs is providing users with appropriate interfaces to manipulate his or her avatar and virtual objects. This can be accomplished through a variety of means, both direct and indirect.

In AR, the fundamental challenge is tracking the pose of the display to ensure consistent overlay of virtual information on the real world. However, this is not a requirement in a CVE. In fact, the lack of real world constraints is often a significant reason for using CVEs. In CVEs, physical space constraints do not limit the number of people that can take part in an interaction to the number that can comfortably fit in the real environment. CVEs also allow interactions not possible in the real world, such as two people simultaneously manipulating the same part of virtual space or viewing an object from the same location.

UCAVE BENEFITS

Hindmarsh *et al* raised a number of concerns pertaining to what would typically be called a non-immersive CVE interface (an ordinary keyboard, mouse, monitor configuration) [9]. These were:

- The horizontal field of view of the display was limited, causing difficulty observing spatial references of others while simultaneously observing the referenced object (source and target).
- Actions were not always reflected by user embodiments, which can also be extended to include both virtual actions and real world actions (*i.e.* a user disengaging from the CVE interface)
- Navigation of the CVE was clumsy, owing to, amongst other things, the lack of an intuitive interface for navigation
- Parallel actions were not supported, such as moving an object and changing viewpoint simultaneously.

Hindmarsh *et al* suggested that an immersive interface (a tracked HMD with multiple tracked position sensors) could address these issues, but argued against an immersive interface because of logistical issues of cost, robustness, and setup difficulties [9].

The UCAVE aims to make the benefits of immersion accessible by reducing these logistical issues. *It should support immersive interaction, but not at the expense of user mobility or ease of use.* In other words, it should provide the expected affordances of a conventional immersive CVE interface: natural interactive viewpoint and avatar control, but be able to be carried around by the user and immediately deployed. While on its surface the principle is simple, immersion, mobility and ease of use are often conflicting, and thus design tradeoffs must be made.

UCAVE ARCHITECTURES AND PROTOTYPE

A general UCAVE architecture is shown in Figure 2. The smart phone (or similar handheld computing device) is the central core of the UCAVE architecture. Modern smart phones have a high-resolution display, a powerful system-on-chip (for interface, radio communication, graphics, and audio processing), and an array of sensors (*e.g.* touchscreen, accelerometers, gyroscopes, and magnetometers). The smart phone is connected to a head mounted audio-visual display. Additional body-worn sensors are incorporated as needed through personal area network technologies. When connected over a wide area network (*i.e.* the Internet), remote users can join other UCAVEs, creating collaborative workspaces. Finally, at certain locations in the environment, external sensors may be available that provide additional capabilities to the user (*e.g.* a motion capture system).

A prototype of the UCAVE hardware has been constructed using only consumer level components. The hardware consists of a jailbroken Apple iPhone 4G (~\$599 USD), and a Vuzix Wrap920 Stereoscopic 640x480 (~\$299 USD) head-mounted display. The Vuzix Wrap920 has an iPhone connector that enables video to be transmitted from the iPhone and displayed on the HMD. Stereoscopic images are transmitted via side-by-side signaling. The Vuzix Wrap920 is powered by a small battery pack. To complete the UCAVE hardware, a stereo headset is employed for audio input and output.

As seen in Figure 1, the iPhone and the Vuzix HMD's display and battery components were assembled together into a single wireless head-mounted device, mounted by means of an elastic band, Velcro, and an iPhone case. This design has unique logistical properties. First, the entire device can be folded and placed in the user's pocket. The case allows the iPhone to be easily detached for other uses. Furthermore, mounting the iPhone removes the requirement that the user hold the iPhone. This relieves the user's hands for other activities and removes a wire from the configuration, making the design more compact. If desired, a second iPhone (or any device that has WiFi or Bluetooth communications, such as a Nintendo Wii Remote) can be employed as a hand-held interaction device.

Currently, we have demonstrated an immersive VE featuring avatars rendered with the open source Object-oriented Graphics Rendering Engine (OGRE). Immersive

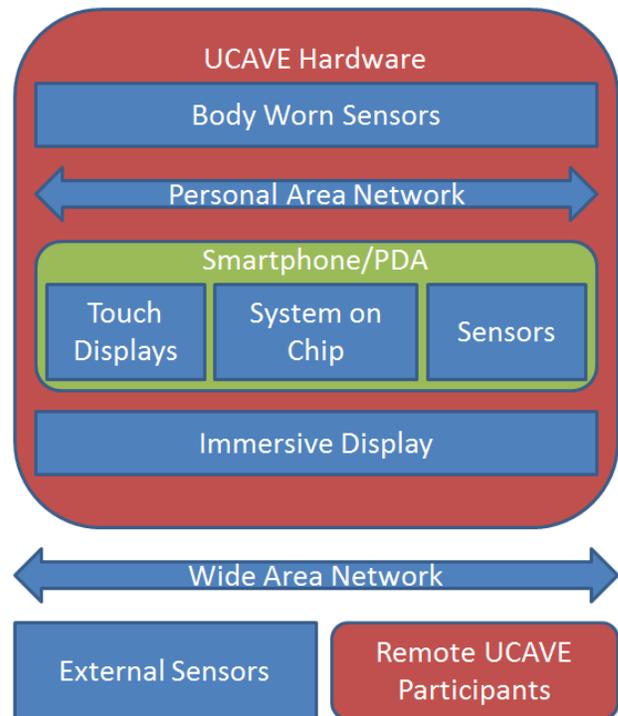


Figure 2 The UCAVE Architecture

rendering is enabled through the built-in iPhone magnetoinertial orientation sensor. The simulation runs in real time, at 60 frames-per-second on the iPhone, with two 7000 polygon count avatars (with a skeletal and facial bone rig), and a 14000 polygon count virtual room.

User Immersion

For the prototype, we focused on providing a natural immersive viewpoint and natural avatar head and hand control. This is because gaze and hand gestures account for a substantial portion of non-verbal communication [1]. In a virtual reality laboratory, it would be possible to use motion capture techniques and tracking systems with fixed reference frames (optical, mechanical, electromagnetic, etc). While not strictly impossible with the UCAVE, it is useful to consider less encumbering and more practical approaches that would work in arbitrary environments and afford the user a greater degree of mobility.

To provide an immersive viewpoint, a source less head orientation estimation is conveniently obtained from the inertial sensors built into the head-mounted iPhone. To compensate for the iPhone being potentially tilted with respect to the face plane, calculation of an orientation offset is necessary at the start of an interaction. This calibration step can be completed by the user looking in a known direction and signaling the software. The head orientation computed for immersive rendering is directly used for the head orientation of the user's avatar as seen by other UCAVE users.

Avatar hand control has proven more challenging to support in accordance with the UCAVE design goals. The current

design features two lightweight, head-mounted, mechanical position sensors taken from the MadCatz GameTrak motion control device. These sensors are inexpensive (\$75 USD) and offer excellent position tracking characteristics (1 cm accuracy, 1mm resolution, no jitter). From the tracked hand position, which is in the frame of reference of the user's head, the motion of the avatar's hand is derived through an inverse kinematics approach. The current limitation of this approach is that it does not provide as much range of motion as collaborative activities may demand (limited to approximately the field-of-view of the Vuzix display).

CHALLENGES AND LIMITATIONS

In contrast to spatially immersive displays, *e.g.* [6], head-mounted displays, such as the one shown in Figure 1, can be battery powered, lightweight, fit in a small pocket when not in use, and provide consistent visual quality from all viewing locations. While this makes head-mounted displays well suited to the UCAVE, the primary tradeoff is a relatively narrow horizontal field of view (~25-40 degrees, comparable to sitting three meters away from a two meter width television) provided by all but the most expensive head-mounted displays. A narrow field-of-view has many negative implications for CVEs, including lack of spatial awareness of other participants and lack of intuitive understanding of what is perceived by others [7]. However, users may compensate for the narrow field of view through sweeping head-rotations, and the UCAVE facilitates such actions by being lightweight and wireless. It should also be noted that aural immersion in the UCAVE can be complete, and this may be used to create spatial awareness of surroundings and others. Furthermore, without support for eye-tracking (currently difficult in head-mounted displays), a narrow field of view encourages the use of head motions to indicate attention [2].

The impromptu deployment possibilities of the UCAVE in the real world (*e.g.* in a crowded public space or in a natural outdoor setting) also raise important issues. First, supporting navigation about the VE is challenging. Unlike outdoor augmented reality systems, natural walking navigation is difficult, because users may bump into objects or people that are occluded by the head mounted display. As an alternative, techniques such as walking-in-place or using a smart phone touch screen as a joystick or pointing device could be used. Finally, as the technology matures, it will be possible to obtain an accurate 3D map of the real environment using vision sensors embedded in the smart phone. This 3D map could be used to implement collision avoidance algorithms or as the virtual environment itself.

CONCLUSIONS

The potential application space of widespread deployment of immersive collaboration technology is vast. Perhaps any CVE application that currently requires laboratory spaces could be extended to millions of people, and thousands of existing mobile applications could begin to leverage immersion. To enable this vision, we are currently preparing a practical prototype for study in the field. Our

focus is on redesigning the tracking system to improve range of motion. Inertial and optical tracking approaches could improve range of motion and may not require custom modifications to the hardware. It is our belief that collaborative virtual reality "in the wild" offers exciting new challenges to the VR and CSCW communities and opportunities to develop novel applications free from the confinements of controlled laboratory spaces.

REFERENCES

1. Argyle, M. and M. Cook. *Gaze and Mutual Gaze*. (1976), xi, 210 pp.
2. Bailenson, J.N. and N. Yee. A longitudinal study of task performance, head movements, subjective report, simulator sickness, and transformed social interaction in collaborative virtual environments. *Presence: Teleoperators & Virtual Environments*, 15,6 (2006), pp. 699-716.
3. Benford, S., J. Bowers, L.E. Fahlén, C. Greenhalgh, and D. Snowdon. User embodiment in collaborative virtual environments. In *Proc. ACM Conference on Computer Human Interaction (SIGCHI)*. (1995), pp. 242-249.
4. Benford, S., C. Greenhalgh, T. Rodden, and J. Pycoc. Collaborative Virtual Environments. *Communications of the ACM*, 44,7 (2001), pp. 79-85.
5. Bowers, J., J. O'Brien, and J. Pycoc. *Practically accomplishing immersion: cooperation in and for virtual environments*. in *ACM Conference on Computer Supported Cooperative Work*. (1996), pp. 380-389.
6. Cruz-Neira, C., D.J. Sandin, and T. DeFanti. The design and implementation of the CAVE. In *Proc. ACM SIGGRAPH*. (1993), pp. 135-142.
7. Fraser, M., S. Benford, J. Hindmarsh, and C. Heath. *Supporting awareness and interaction through collaborative virtual interfaces*. in *ACM Symposium on User Interface Software and Technology*. (1999), pp. 27-36.
8. Gartner Group. *Forecast Analysis: Mobile Devices, Worldwide, 2008-2015, 1Q11 Update* (2011).
9. Hindmarsh, J., M. Fraser, C. Heath, S. Benford, and C. Greenhalgh. Object-focused interaction in collaborative virtual environments. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7,4 (2000), pp. 477-509.
10. Otto, O., D. Roberts, and R. Wolff. A review on effective closely-coupled collaboration using immersive CVE's. In *Proc. ACM International Conference on Virtual Reality Continuum and its Applications* (2006), pp. 145-154.
11. Piekarski, W. and B.H. Thomas. *Tinmith-Mobile Outdoor Augmented Reality Modelling Demonstration*. in *IEEE/ACM International Symposium on Mixed and Augmented Reality*. (2003), p. 317.
12. Reitmayr, G. and T. Drummond. Going out: robust model-based tracking for outdoor augmented reality. In *Proc. IEEE/ACM International Symposium on Mixed and Augmented Reality*. (2006), pp. 109-118.