

**Telepresence with no Strings Attached:
An Architecture for a Shared Reality Environment**

Christian Côté, Shawn Arseneau, and Jeremy R. Cooperstock*

Keywords: augmented reality, telepresence, person tracking, gesture recognition, image processing

Poster Submission

***Contact author:**

Jeremy R. Cooperstock
Centre for Intelligent Machines, McGill University
3480 University Street
Montreal QC H3A 2A7
Canada

Tel: +1-514-398-5992
Fax: +1-514-398-7348
Email: jer@cim.mcgill.ca

Telepresence with no Strings Attached: An Architecture for a Shared Reality Environment

Abstract

The Shared Reality Environment is an ongoing project that explores the use of virtual reality technologies to achieve realistic computer-mediated human-human interaction. The project integrates immersive displays, spatialized audio, haptics, and gesture recognition, through a minimal latency network architecture. As our primary goal is to provide distributed participants with a convincing sense of co-presence without inhibiting natural, spontaneous interaction, the environment must employ unobtrusive technology wherever possible.

Keywords: augmented reality, telepresence, person tracking, gesture recognition, image processing

1 Background

The Shared Reality Environment is one of many telepresence research projects conducted around the world [3, 4, 6]. Among these, the “Office of the Future” [6] at the University of North Carolina is particularly interesting, as it imaginatively addresses many of the same issues as our own: seamlessness of the technology, realism of the experience, and quality of the interaction. However, the constraints imposed in adapting the technology to a real office environment are somewhat restrictive in terms of the scope of applications we would like to accommodate. Indeed, distributed musical rehearsals or distance education are typically ill-suited to an office environment.

Our objective with the SRE is to overcome the limitations of conventional telepresence tools using novel technologies and practices. For example, high-end virtual reality immersive displays, such as the CAVE [2], are more compelling and visually engaging than desktop computer monitors for human-computer interaction. Similarly, we believe that large screen displays, in which participants are projected at “life-size” allow more effective human-human interaction [4, 6]. Likewise, high-resolution spatialized audio can support such demanding applications as musical rehearsals or performances [3] as well as multiple simultaneous conversations. Haptic feedback can be introduced to help bridge the physical separation of remote individuals. Such feedback could be in the form of reproducing the floor vibrations in response to a user walking around to the room. Finally, gesture recognition could allow virtual shared objects, such as

CAD models, to be manipulated by distributed design teams, enabling a new range of computer-supported collaborative tasks. By incorporating all these technologies into a single system, we believe it is possible to establish a *shared reality* in which distributed users are able to interact freely, unhindered by the constraints of conventional “state-of-the-art” videoconference systems.

The SRE is composed of multiple rooms, each of which contains an enclosure of three screens of rear-projected video, a multi-channel sound system for the generation of spatialized audio, various haptic transducers and gesture recognition mechanisms, all interconnected via a high-speed network. An important element of our research is the concept that interaction with the technology must be transparent to the users. This implies that the computer should recognize what the user is attempting to do, and not the other way around. Furthermore, we avoid any form of body-worn trackers or other special clothing restrictions that one might be tempted to use for gesture recognition or person tracking.

1.1 Participant Representation

One of the first challenges to establishing communication between several remote participants is representing each of them realistically in a single virtual space. These representations, or avatars, may not be perfect, but should be sufficiently convincing to allow seamless interaction.

A naive solution would be to simply have an unmodified video projection, where the camera output directly feeds the video wall display. While this would certainly be easy to implement, it only offers mediocre immersive realism for our applications.

Our alternative is to perform background removal on the camera input to obtain an image of the user isolated from the scene. The image is then inserted into the virtual environment (see Figure 1). This approach has the advantage of offering an acceptable level of realism, especially pertaining to occlusion. The main shortcoming of our approach is the absence of volume of our avatar. One possible solution would be to use stereo cameras to collect depth information and construct a relief model of the participant [4]. While this method holds promise, experimental results clearly indicate that the technology needs to mature before we can adopt it for practical applications.

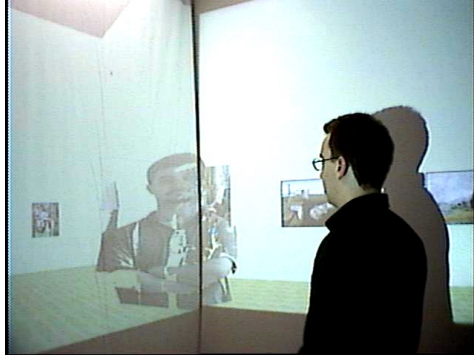


Figure 1 Remote participant avatar with background removed

1.2 User tracking

Typical trackers for the CAVE include electromagnetic devices such as Ascension's Flock of Birds™, wearable transmitters, or optical markers [1]. While the performance of such devices is impressive, their reliability comes at the expense of reduced user freedom, requiring either that the user wears one or more sensors or is tethered by a cable.

Our approach makes use of passive input (ie. digital video) to free the user of wearable devices. The tracking algorithm uses the processed video image with the background removed, a step that has already been performed for the purpose of representation in the virtual space, to locate the user. With the a priori knowledge of the camera positions, it is possible, with the use of three cameras, to track any feature we choose.

While the algorithm is fairly primitive and cannot compete with the accuracy of commercial trackers, the initial performance results are highly encouraging.

1.3 Gesture Recognition

Gesture recognition and collaborative manipulation are typically achieved through the use of wearable data gloves or trackers, trackball or joystick-like devices, markers or optical algorithms [1]. These devices offer attractive features including high precision and haptic feedback, however the requirement of wearing a special device in order to access certain functionality of the environment risks reducing the spontaneity of expression. We would consider it absurd if, for example, in order to read our email, we first had to put on special "email-reading" glasses.

Video-based methods, on the other hand, offer a direct coupling between users' actions and the corresponding effect. A generic video-based gesture recognition algorithm can be constructed as follows: First the region of interest, i.e. the user, is isolated from the rest of the scene. Next, the cut-out user image is segmented into its constituent components, such as head, arms and legs. Finally, analysis is performed to establish the relative

motion of each body part. Tracking the pointing gesture of a user is shown in figure 2.



Figure 2 Simulation of CAD model manipulation in the SRE

Various techniques exist for the segmentation and analysis phases, including the relatively low-cost blob approach employed by Penny's "Traces" tracker [5].

2 References

1. Buxton, B. A directory of sources for input technologies, online resource: www.dgp.toronto.edu/people/BillBuxton/InputSources.html
2. Cruz-Neira, C., Sandin, D. J. and DeFanti, T. A. "Surround-Screen Projection-Based Virtual reality: The Design and Implementation of the CAVE" Computer Graphics, SIGGRAPH Annual Conference Proceedings, pp.135-142, 1993
3. Konstantas, D., Orlarey, Y., Carbonel, O. and Gibbs, S. "The Distributed Musical Rehearsal Environment." IEEE Multimedia, Volume 6, Issue 3. pp. 54-66. July-Sept. 1999.
4. Ogi, T., Yamada, T., Tamagawa, K. and Hirose, M. "Video Avatar Communication in Networked Virtual Environment." The 10th Annual Internet Society Conference. Yokohama, Japan. 2000.
5. Penny, S., Smith, J. and Bernhardt, A. "Traces: Wireless Full Body Tracking in the CAVE." ICAT Virtual Reality Conference. Japan. 1999.
6. Welch, G., Fuchs, H., Raskar, R., Towles, H. and Brown, M. "Projected Imagery in your 'Office of the Future'." IEEE Computer Graphics and Applications. Volume 20, Issue 4. pp.62-67. July-Aug. 2000.