A Tele Immersive System for Collaborative Artistic Creation

Pierre Boulanger and Matha Benitez
Department of Computing Science
University of Alberta
pierreb@cs.ualberta.ca
Tel: (780) 492-3031

The main idea behind tele-immersive environment is to create an immersive virtual environment that connects artist across networks and enables them to interact not only with each other, but also with various other forms of shared digital data (video, 3D models, images, text, etc.). Tele-immersive environments may eventually replace current video and telephone conferencing, and enable a better and more intuitive way to communicate between people and computer systems. In order to accomplish this, participants to a meeting has to be represented digitally with a high degree of accuracy in order to keep a sense of immersion. For example, in a normal conference setting people talk to each other around a table with ease and without having to think which window to look at and which sound stream should be turned on or off. Tele-immersive environments should have the same ”feel” as a real meeting. Interactions among people should be natural.

To facilitate natural communications among users in a tele-immersive environment, participants need to be represented in a realistic way. There are several techniques in which this is can be done. To name a few, methods such as articulated 3D model (avatars), image based rendering using live real-time video, or some combination of the previous two methods e.g. 3D model textured from a live video is currently being investigated. Articulated 3D models of people are computationally expensive to animate and for high-resolution models, rendering time becomes a concern. At the same time, low-resolution models are not very convincing and interactions no longer become natural. Image- based rendering approach uses live video of a user from multiple cameras, producing a so-called video avatar. Using this technique the user can be rendered photo-realistically. In other to create such a system, we need to solve the key problem of how to create in real-time new images from a fixed network of cameras that will correspond to new viewpoints. We also need to do this for two virtual cameras corresponding to the inter-ocular distance of each participant. If we represented users as an articulated 3D model this is an easy question to answer, simply render the 3D model from the different perspectives. Using video avatars, this question becomes more complicated because we need to generate views of the user at angles in which there may be no camera available.

The proposed system described in this paper use data from a position tracking sub-system to locate the participants (102a, 102b), and then determines which two cameras (208a and 210a in Figure 1) that most closely approximate a view of participant (102a) from the perspective of participant (102b). The system then selects those two cameras to supply video images of participant (102a) to participant (102b). The system similarly selects cameras 208b and 210b to supply video images of participant 102b to participant 102a. The system then separates the image of each participant 102a, 102b from the background that appears in the respective video images. This is done using an improved version of the PFINDER algorithm [Wren et al. 97] that perform foreground /background segmentation using a background mask and some stereo information supplied by the two closest video cameras.

The system then transforms these respective video image pairs to create a stereo pair of video images separated by a nominal inter-ocular spacing of participant 102b. A view interpolation algorithm similar to the one described in [Baba et al. 00] [Seitz97] is used and improved for this context. Each transformed video pair is then transmitted to the participants (102a), (102b) and incorporated into the respective participant's view of the virtual meeting on a polygon located in the virtual world and display using a stereo texture technique as illustrated in Figure 3.

As participants move around at their respective locations, the system tracks their position and select appropriate camera pairs. In the example, shown in Figure 2, when participant 102b moves forward and to the right of the position showed in Figure 1, the system select cameras 210b and 212b to capture views of participant 102b. The system likewise select cameras 208a and 206a for providing the most appropriate perspective for capturing views of participant 102a to be supplied to participant 102b. The position information related to each participant is also used to process the captured audio of each participant's voice, in order to reproduce the sound of each participant's voice in the 3D space of the virtual meeting room and to locate the rendered polygon representing the participant.

In this paper, we will describe a new binocular view interpolation algorithm based on a re-projection technique using calibrated cameras and an estimate of the position and orientation of the users using two cameras. We will discuss the various aspects of this new algorithm and of the hardware systems necessary to perform these operations in real-time.

The system under development is focusing on a *Collaborative Artistic Creation* application. One of the objectives of this work is to simulate body presence in virtual worlds. An earlier work that explores this concept was embodied in an art piece entitled "Spectral Bodies," 1991 [Hayles95]. This exploratory work was based on the experiments of Dr. James Lackner (Brandieis University) in proprioception (a sense of being present in one's body) [Lackner92]. Lackner's experiments, which created illusions in the subject of where parts of the subject 's body might be, were combined with interactive images of the body in the virtual world. Users of such systems will interact with a visual representation of their bodies in virtual worlds.

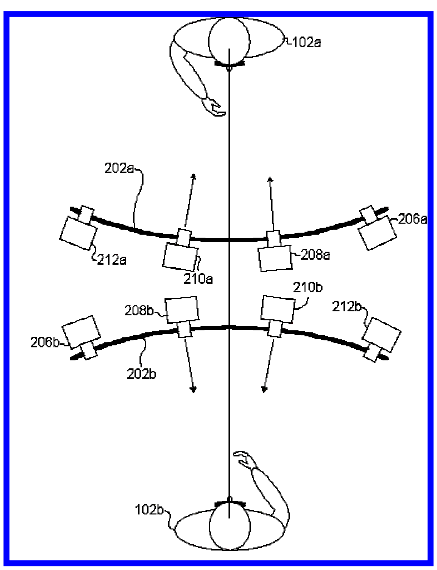

Figure 1: Binocular view selection and interpolation for a front view selection

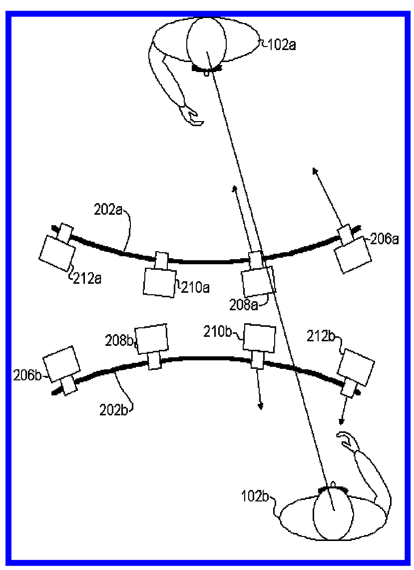

Figure 2: Binocular view selection and interpolation for a side view selection

Figure 3: Inserted stereo video texture in the virtual world

References

[Baba et al. 00] Baba S., Saito H., Vedula S., Cheung K.M., and Kanade T. *Appearance-Based Virtual-View Generation for Fly Through in a Real Dynamic Scene,* In VisSym '00 (Joint Eurographics - IEEE TCVG Symposium on Visualization), May, 2000.

[Hayles95] Hayles, N. K. *How to Put Bodies Back in the Picture*, Immersed in Technology: Art and Virtual Environments, Cambridge: MIT Press, 1995.

[Lackner92] Lackner, J.R. *Spatial orientation in weightless environments*. Perception, 21, 803-812, 1992.

[Seitz97] Seitz S.M., Image-Based Transformation of Viewpoint and Scene Appearance, Ph.D. Dissertation, Computer Sciences Department Technical Report 1354, University of Wisconsin - Madison, October 1997.

[Wren et al.95] Wren, C. R., Azarbaye jani, A., Darrell, T., and Pentland, *A. Pfinder: Real-Time Tracking of the Human Body*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 780-785, July 1997.