Integrating Live Video for Immersive Environments

Michitaka Hirose, Tetsuro Ogi, and Toshio Yamada
University of Tokyo

Cabin, an immersive projection display, is a room-sized, five-screen system that can display both computer graphics and video images. Several Cabins connected via a broadband network form the Cabinet system. Cabinet includes video avatars—a key new technology for sharing virtual worlds. Using video avatars, we have experimentally evaluated the ability to express positional information between distant users.

Immersive projection technology (IPT) has become a popular virtual reality display system.[1] Many projection-based displays such as the Cave Automatic Virtual Environment (CAVE) and C2 are currently used at various research institutes worldwide.[2-5] This kind of display can provide a wide field of view and high-resolution stereo images to multiple persons.

In 1997, the Intelligent Modeling Laboratory (IML) at the University of Tokyo developed Cabin (Computer Augmented Booth for Image Navigation) as an extended immersive projection display system.[6] Cabin is a room-sized, five-screen display for both computer graphics images and video images.

In 1997, we extended this IPT to a networked environment by connecting Cabin to other immersive projection displays, such as CoCabin at Tsukuba University and Univers at the Communication Research Laboratory of the Ministry of Posts and Telecommunications. This research project, called Cabinet, aims to share distant virtual 3D spaces based on computer graphics images and video images by using networked immersive projection displays.

Overview of Cabin

We designed Cabin's physical structure and features to facilitate display of both computer graphics and video images, as follows.

Five-screen configuration

Cabin has five stereo screens—one at the front and one each on the left, right, ceiling, and floor. Figure 1 shows the external appearance, and Figure 2 shows the structure of Cabin. The display space is 2.5m × 2.5m × 2.5m. Its raised frame lets all screens be rear-projected. A magnetic sensor (Polhemus Ultratrak Pro) tracks the position and orientation of the viewer's head, and 3D images from the user's viewpoint are projected onto the five screens using Electrohome Marquee 8500 stereo projectors.

Because of the five-screen configuration, Cabin provides stereo images with an extremely wide field of view (FOV). Consequently, it effectively synthesizes a life-sized immersive VR world.

Tempered glass screen

To construct the five-screen configuration, we first had to address several problems. The floor screen must support the viewers' weight. In the case of CAVE, which has no ceiling screen, the floor screen image can be front-projected from the ceiling. However, with Cabin, since both the ceiling and floor screens are rear-projected, the floor structure must support users standing on the rear-projected screen. Also, the interface design between the floor screen and the side screens must prevent a wide border, which would cause visual inconsistency.

Therefore, the floor screen consists of 30-mm tempered glass, in two layers of 15-mm tempered glass plate. According to the theoretical calculation, this structure can support a distributed load weight of 2500 kg in a 1.0m × 1.0m square area. Although this could handle the weight of 25 people, we limit the number of participants to three.

Image generation system

Cabin has two kinds of image generation systems, one for computer graphics images and one for real video images (see Figure 3). To generate computer graphics images, it uses five graphics workstations (SGI's i-Station)—one assigned to each screen to generate stereo images. The five workstations connect via a shared memory network (Systran ScramNet), which permits synchronizing displayed data for each screen.

To generate a video image, 10 VCRs feed images to 3D converters, which create five stereo video streams—one for each screen. These video streams are synchronized frame by frame using the time codes of the video controllers (Sony V-Box), then projected onto each screen.

1070-986X/99/$10.00 © 1999 IEEE

Figure 1. External appearance of Cabin.

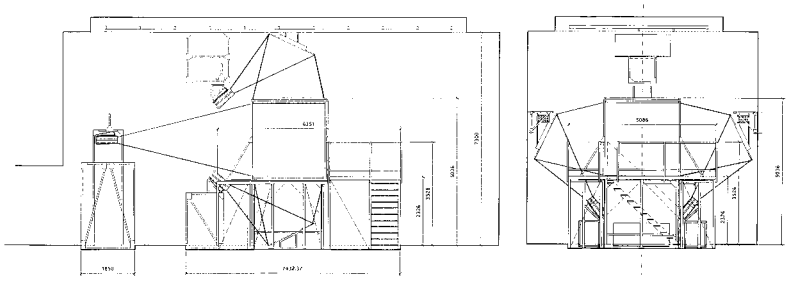

Figure 2. Five-screen configuration of Cabin.

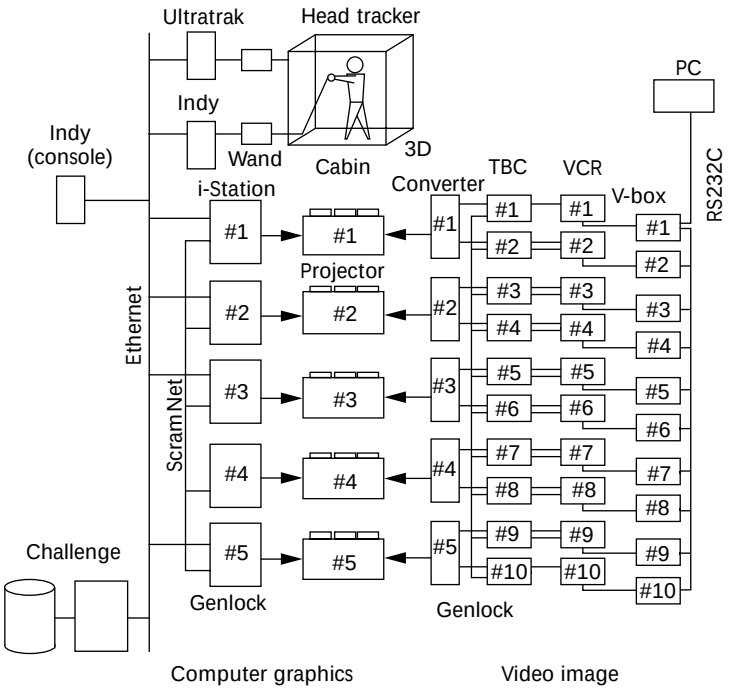

Figure 3. Image generation system for Cabin.

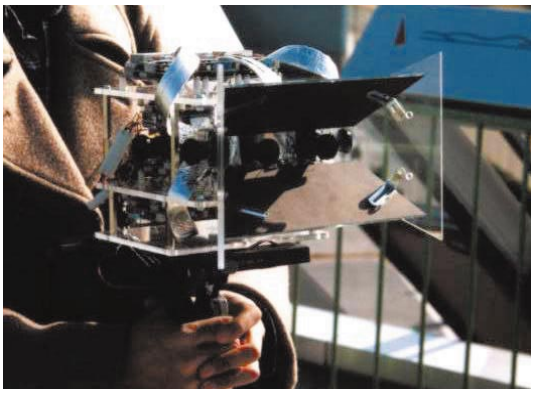

Figure 4. Multilens camera.

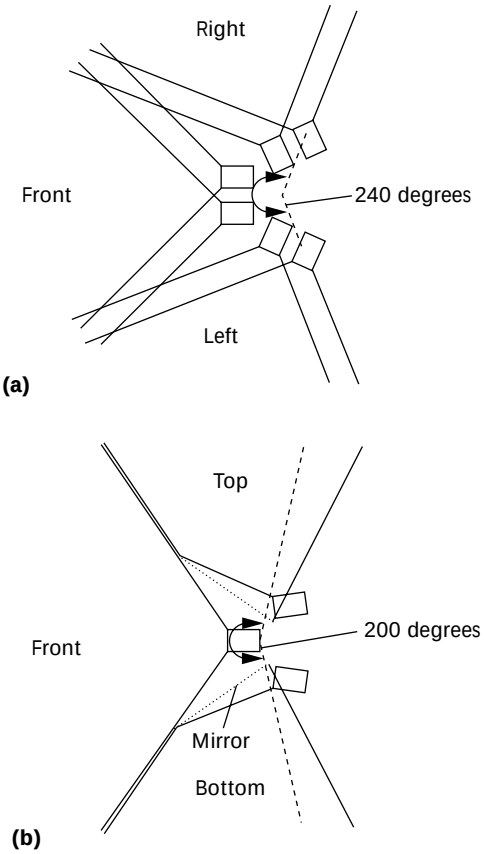

Figure 5. Configuration of lenses. (a) Top view. (b) Side view.

Multilens camera

To record and generate a virtual world based on real images requires a special multilens camera. Figure 4 shows the prototype system of the multilens camera for Cabin. This camera has 10 lenses to capture the five stereo video images for the five screens (see Figure 5). Pairs of lenses are mounted in each direction in front and to the left, right, top, and bottom. Each lens has an 89-degree horizontal viewing angle and a 125-degree vertical viewing angle. Mirrors serve for the top and bottom lenses. In total, this camera has a 240-degree horizontal viewing angle and a 200-degree vertical viewing angle. Although several mirrors are positioned to coincide with each camera's

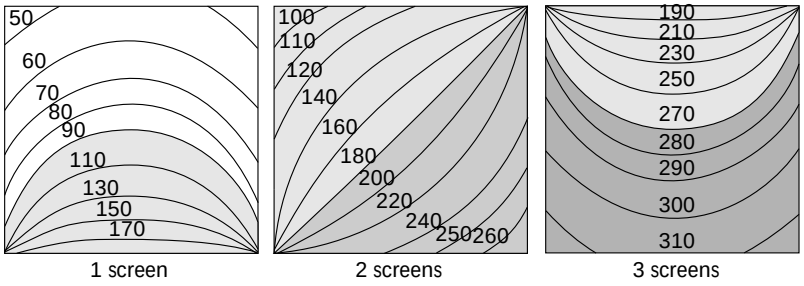

Figure 6. Fields of view for several screen configurations.

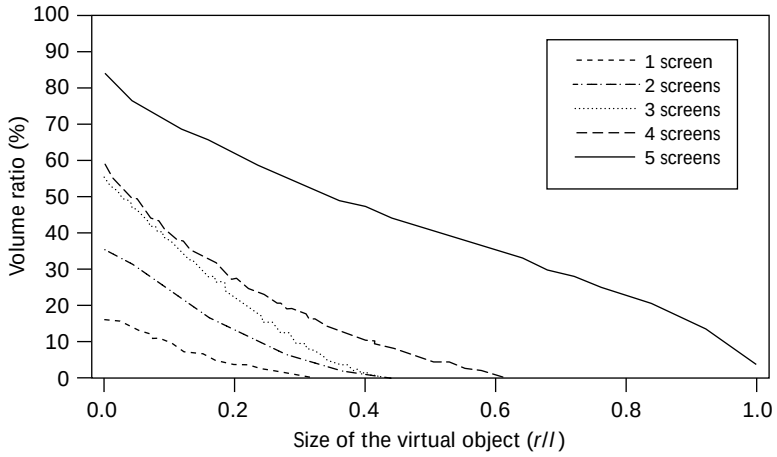

Figure 8. Volume ratio of viewpoints where a virtual object can be seen completely.

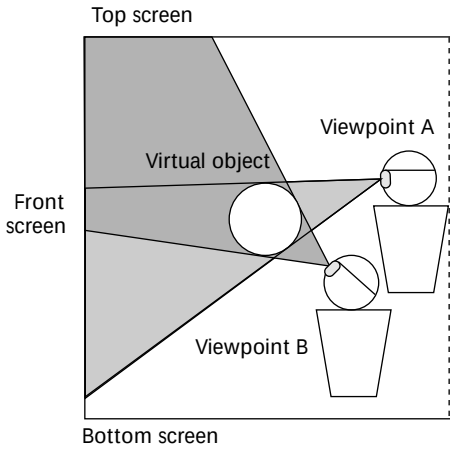

Figure 7. Viewpoint and necessary screens.

viewing center, dead angles for near areas and overlaps for far areas remain.

Cabin's Effectiveness

Cabin's space sharing features, enabled by its physical design, allow measurement of its applicability.

Wide field of view

Cabin's five-screen configuration provides quantitative effectiveness. First, using five screens enlarges the user's field of view. Figure 6 shows the FOV values—illustrated by contour lines—for several screen configurations in the square areas. These results show that Cabin expands the user's FOV relative to the number of screens. The user standing at the center of the display space has a 270-degree FOV—up and down and right to left—because of the three screens horizontally and vertically. This means that the image border goes unnoticed and the stereo image covers almost all the viewing field (180 degrees) even if the user looks around.

Freedom of viewpoint

The wide FOV feature effectively displays a wide area of the virtual world that exists outside of the cubic display space. However, the user needs a wider moving area to see a 3D object displayed inside the display space from various viewpoints (see Figure 7). For example, if the user's viewpoint is located at position A in Figure 7, only the front screen displays the object. However, when the user's viewpoint moves to position B in Figure 7, displaying the object completely requires both the front and top screens. Therefore, the more screens available in the display system, the more freedom the user has to move around. This capability proves essential for generating motion parallax in a virtual environment.

Figure 8 shows the fraction of the total volume from which the user can see virtual objects without obstruction by screen borders. We plotted this graph from numerical simulations for several screen configurations, that is, one screen (front), two screens (front and side), three screens (three walls), four screens (three walls and floor), and five screens (three walls, ceiling, and floor).

Assume the displayed object is a sphere of diameter r, normalized to the screen size l. Also, the vertical movement of the user's viewpoint is limited to 170 cm up from floor level (the horizontal movement is 2.5m × 2.5m). According to the simulation, when $r/l = 0.3$, for example, the user's viewpoint can move around 0.9 percent of the display space for one screen, 5.3 percent for two screens, 9.0 percent for three screens, 17.5 percent for four screens, and 53.5 percent for five screens. Based on this result, we conclude that the

Figure 9. Walkthrough application in Cabin.

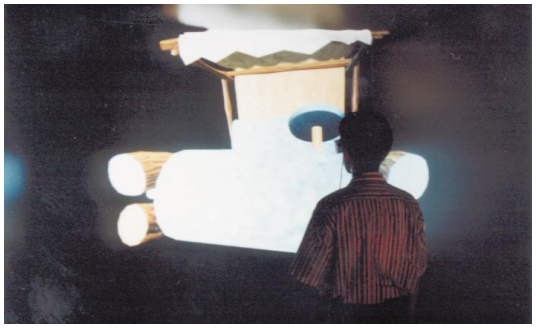

Figure 10. Virtual mockup visualized in Cabin.

freedom of the user's viewpoint exceeds the simple ratio of the screen numbers.

Fields of application for Cabin

The evaluation of the FOV and freedom of viewpoint shows Cabin's effectiveness for presenting both the virtual space that extends outside the display space—surrounded by the screens—and virtual objects located inside the display space. Therefore, Cabin could benefit various fields.

For example, Cabin could display a virtual trip, facilitate tele-existence, and support urban design to visualize a wide area of the virtual space. It could also display a virtual mockup or a scientific visualization to visualize a 3D virtual object. Figure 9 presents one example of an application, demonstrating a walkthrough in a virtual city. Figure 10 shows an application with a virtual mockup.

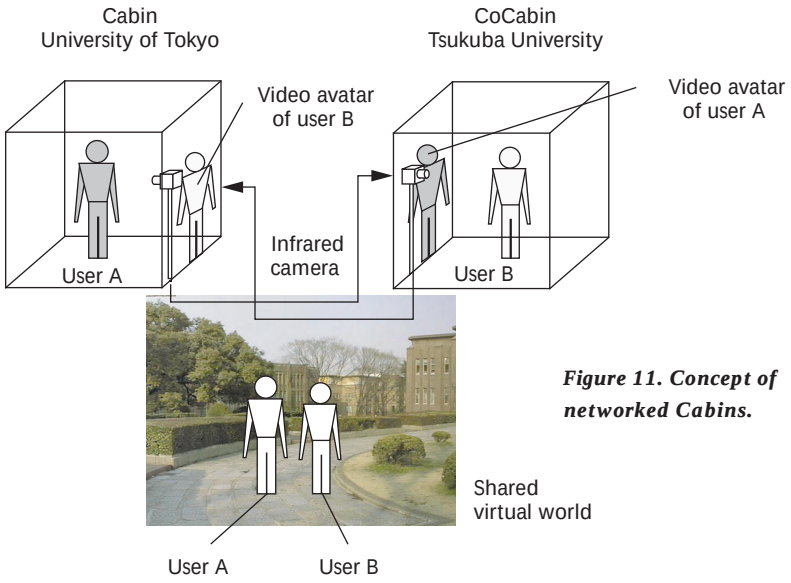

Figure 11. Concept of networked Cabins.

The Cabinet project

The next step in this research extended Cabin to a networked environment by connecting several Cabins via broadband communication lines (see Figures 11 and 12). In this immersive virtual environment, participants at remote locations experience natural communication. This kind of network differs from the simple connection of an ordinary video conferencing system because it transmits both image information and spatial information, such as the users' positional

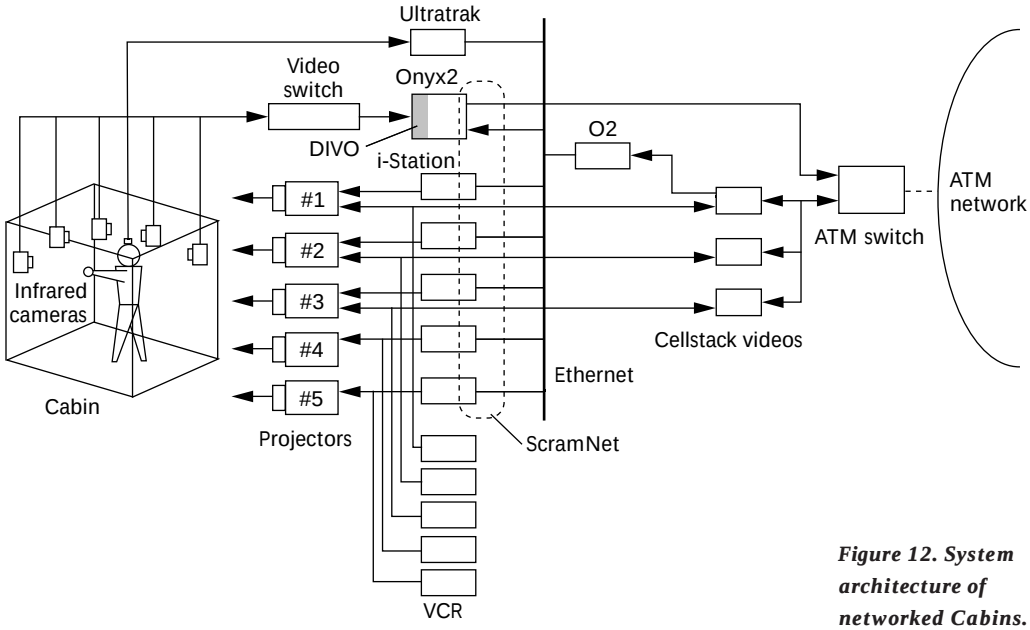

Figure 12. System architecture of networked Cabins.

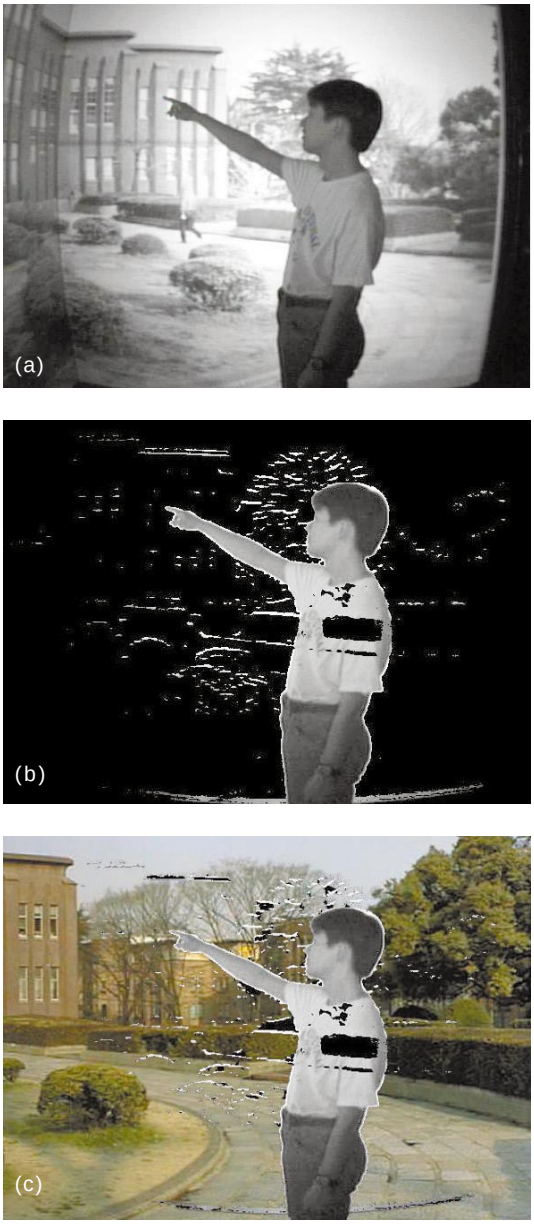

Figure 13. Process of making a video avatar. (a) Captured image. (b) Extracted user's figure. (c) Integrated image.

relationships. Consequently, in this environment users have the sense of being in the same space and sharing the same world—presence.

Therefore, networked multiscreen environments can perform as virtual offices or laboratories for collaborative work. In scientific visualization applications, for example, researchers at remote locations can share visualized numerical simulation data and hold discussions while looking at the same data. This environment helps a scientist analyze phenomena while collaborating with distant colleagues. With a shared video image world, participants can undergo a common experience by looking at the same scene captured by the multilens camera. This environment fosters collaborative work through tele-existence.

Cabinet and MVL

One of the preliminary projects using networked Cabins, Cabinet—a communication experiment—began in 1997. In one instance, CoCabin at Tsukuba University and Univers (Unified Virtual Environment and Space) at the Communications Research Laboratory were connected to Cabin.[7]

CoCabin is a small, cubic, multiscreen display with three 90-inch screens, one each at the front, the left, and the right. Univers is an open-type multiscreen display having three 100-inch screens placed at obtuse angles. A broadband network connects these multiscreen displays—necessary because of the large amount of data transmitted. Currently, CoCabin and Univers connect to Cabin via 75 megabits per second (Mbps) asynchronous transfer mode (ATM) networks.

The Cabinet communication experiments take place as part of the Multimedia Virtual Laboratory (MVL) project sponsored by the Ministry of Posts and Telecommunications. MVL offers a concept for creating a distributed virtual laboratory where researchers jointly engage in a project via a high-speed network and can feel as if they share the same space. This concept supports applications in science, engineering, education, medicine, and conferences, for example. The organizers expect Cabinet to be a key technology in the MVL project. In the future, Cabinet will extend to other sites, including the United States and Singapore.

Cabinet system architecture

As mentioned previously, virtual worlds displayed in Cabins are classified into computer graphics and video image worlds. Therefore, the system equipment for Cabinet should let users share both types of virtual worlds. Figure 12 shows the system architecture of networked Cabins. The system transmits data directly using Internet Protocol (IP) over ATM protocol via the ATM switch (Fore Systems ASX-200BX). National Television Standards Committee (NTSC) video images from VCRs are compressed into motion JPEG data by video codecs (K-Net Cellstack video) and transmitted over the ATM network.

At this stage, the system provides only three video data channels because of the incomplete equipment of the video codecs. One of the original graphics workstations (SGI's i-Station) gives way to an SGI Onyx2, which has two video input and output ports. One port captures the video

image of the user's figure, and the other inputs the video image of the shared world. The workstation superimposes the user's figure on the image of the shared virtual world.

When sharing a computer graphics world, the system transmits only the user viewpoint data and the changed model. The displayed image is generated and rendered on each site so that users can see the world from their perspective. In this case, the required bandwidth isn't large. Conversely, when sharing a video image world, the system transmits images for three screens from the VCR site to the linked site through the network, and all users see the same scene. This requires a broadband network. A bandwidth of 10 to 15 Mbps transmits each video image, and the rest of the band handles the computer data.

A more important technical problem exists: displaying each user. The video avatar offers one solution.

Video avatar

A networked environment must project human images on a mutual display to help users communicate smoothly. Although distributed virtual environments[8] often use computer graphics avatars, natural communication with a polygon-based avatar is sometimes difficult because it can't represent the user's facial expression. Therefore, we transmit a video image of the user's figure directly. In the case of an immersive projection display, all users must share positional relationships in the virtual world because of their immersion in the same virtual environment. They can't interact effectively without this knowledge. We can't say that "space" is "transmitted" or "shared" until this positional relationship is shared.

The process of making the video avatar follows. The camera captures the image of the user in Cabin and extracts the image of the user's figure from the background. This image is transmitted to another site and integrated with the shared virtual 3D world. Installing several cameras inside Cabin allows images from various viewpoints. In other words, motion parallax is synthesized in the shared world. Finally, the rendered image is projected as a video avatar in the immersive projection display.

Extracting the user's figure

To use a video avatar in Cabin required solving several problems. First, it's necessary to capture the user's image clearly in Cabin's dark display space. Since the brightest point is only 10 lux, we use an infrared camera to obtain a clear image. Consequently, the image of the current video avatar appears in black and white.

Next, we must extract an image of the user's figure from the camera image. A popular method of extraction uses a chromakey such as the blue screen.[9] However, Cabin can't use this method because the background image is an important component of the virtual world.

We developed an alternative method to extract the user's figure. Figure 13 shows the process of making a video avatar. Figure 13a shows an image taken by the infrared camera at one site. The computer captures this and, at the same time, simulates a reference background image. Subtracting this reference image from image (a) obtains image (b) as the user's figure. The resolution of the user's image is 720×486 pixels, and the data size is 350 Kbytes. This image data (in Figure 13b) is transmitted to the connected site. Finally, the image in Figure 13c is generated at this site by integrating the transmitted image (b) with the image of the shared virtual world.

Size of video avatar

The extracted user's figure must display in 3D space at its actual size and in the correct position, not just superimposed as a 2D image. To adjust the displayed size, we consider the camera conditions and the relationships among the users' positions in the virtual space. Figure 14 shows the method of superimposition taking the human size into consideration. A position sensor tracks the user's position (x, y). If the camera has a viewing angle of α (alpha) and the user stands at a distance d from the camera, calculate the size of the user's image W by

$$W = 2d \tan(\alpha/2)$$

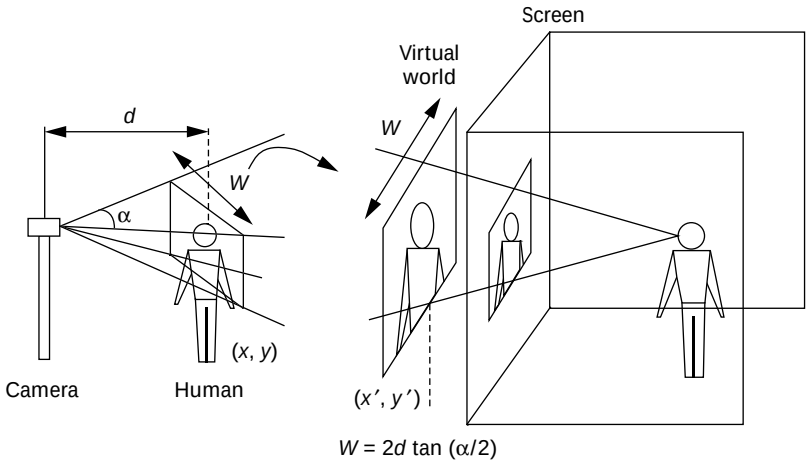

Figure 14. Adjustment of video avatar's size.

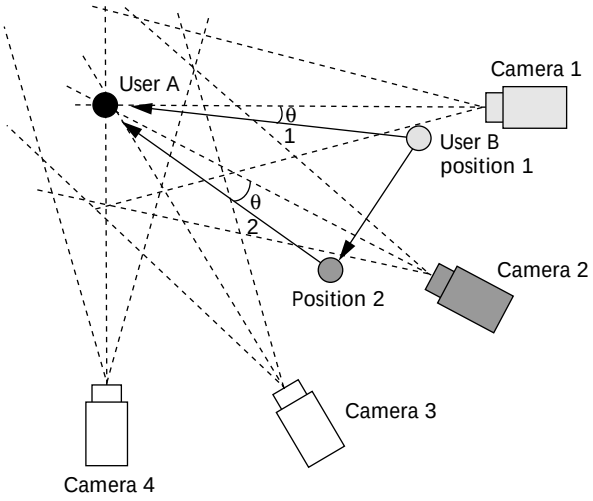

Figure 15. Users' positional relationship and camera selection.

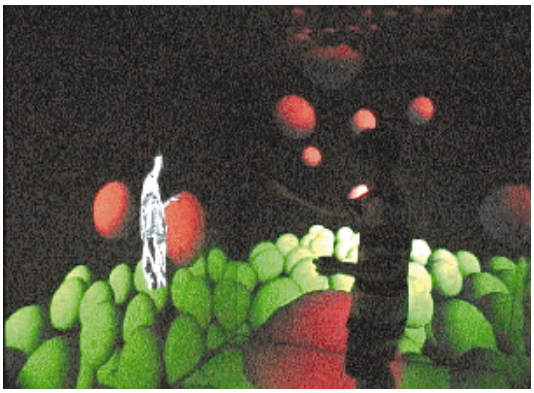

Figure 16. Example of sharing a computer graphics world.

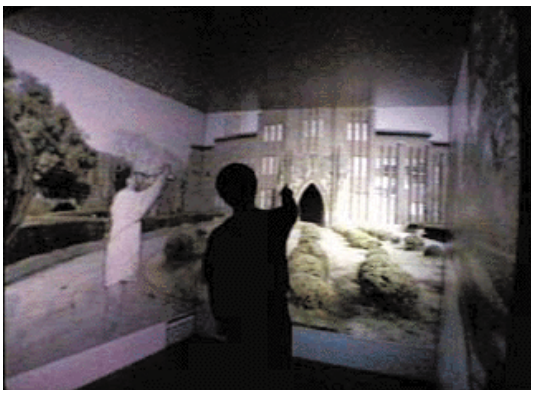

Figure 17. Example of sharing a video image world.

The system renders the image of the extracted user's figure as texture data on a transparent panel. The size of this panel is adjusted to W, placed at the user's position (x', y') in the virtual world and rendered by perspective projection. Finally, the video avatar on the panel appears at its actual human size.

Camera switching to generate motion parallax

The next problem is how to represent the video avatar as a 3D object. When the infrared camera takes an image of user A at one site, it should appear from user B's viewpoint at the connected site. However, moving the camera in the display space is difficult. Therefore, we employed an image-based rendering technology.

Of the several small cameras (17 mm in diameter) set up in the display space, the camera nearest user B's viewpoint is selected and used. The cameras are positioned at 30-degree intervals, as shown in Figure 15. When user B occupies position 1, camera 1 is selected, and when user B moves to position 2, camera 2 is selected. This method approximately reconstructs the relationship of the users' positions to the shared virtual world using motion parallax as a cue.

System performance

Remote users communicate with each other by transmitting their own video images using the video avatar method. Figures 16 and 17 show examples of sharing a virtual world in Cabinet. In Figure 16, users share a computer graphics world of scientific visualization data. In Figure 17, users share a video image world.

In these applications it's important to generate video avatars with as little lag time as possible, to facilitate smooth communication. The total lag time in generating a video avatar consists of capturing the video image, extracting the user's figure, transmitting the data, superimposing the user's figure, and so on. In the current system, the total time lag reaches 0.74 seconds. The refresh rate reaches 4.1 Hz for generating a video avatar in the shared virtual world.

Expression of positional information

We evaluated the communication quality of the shared virtual space by measuring users' pointing accuracy within the space.

Pointing experiment

In the networked Cabins, users talk to each other while looking at the same virtual object, such as a virtual mockup or visualized data. Effective communication requires transmitting

positional information about the object. Although the video avatar uses a 2D video image, it also has 3D information such as head tracking and motion parallax as a result of switching between several cameras. This supports expressing positional information in the shared virtual environment.

We evaluated whether one user could discern the position another user pointed at in the shared space when using the video avatar method. We conducted the experiment between Cabin and a 70-inch one-screen display instead of a Cabin to Cabin connection. Figure 18 shows the experimental setup. The shared virtual world included a number of balls arranged at the grid points of a square lattice. The user in front of the 70-inch screen (the pointer) pointed at one of the balls. Looking at the video avatar figure, the Cabin user (the subject) indicated orally the ball selected. The balls were spaced 10 cm, 20 cm, and 30 cm apart. If the subject selected the wrong ball, we calculated the position error from this spacing. Although the pointer stood in front of the screen to look at the balls, the subjects in Cabin could walk around the display space and look at the avatar from various directions. Figure 19 shows the experiment.

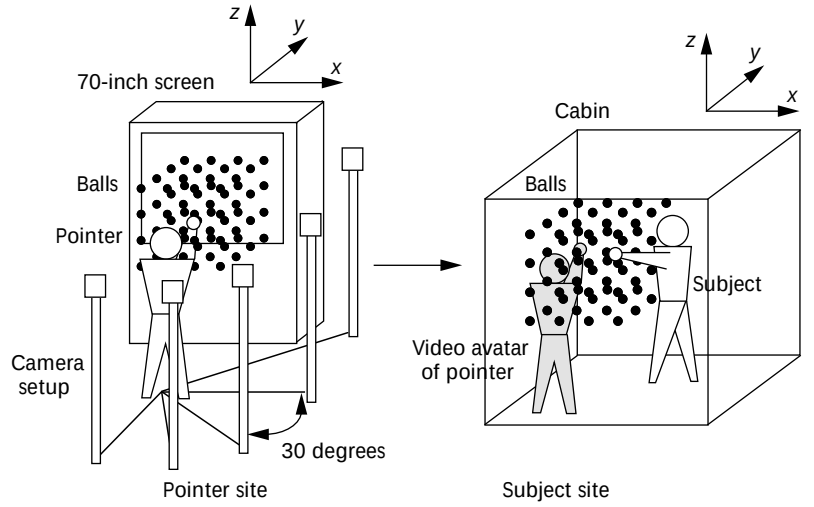

Figure 18. Setup for pointing experiment.

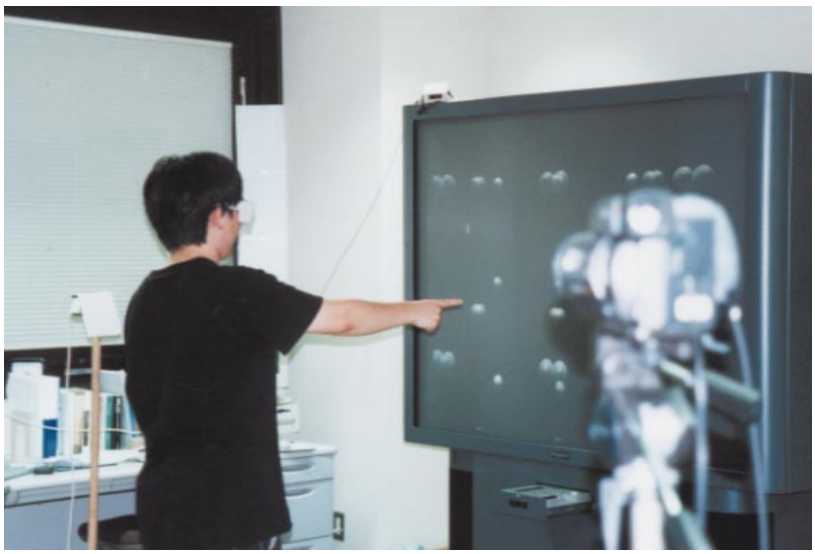

Figure 19. Pointing experiment (pointer site).

Experimental results

Table 1 shows the average and the standard deviation of the error for five subjects in this experiment. The average error in perceiving the indicated position was 18.8 cm. Comparing the errors along the *x*, *y*, and *z* axes, the *x* axis error was the largest, probably because of insufficient use of the motion parallax by the user. Since the cameras ranged from the right side around to the back of the pointer, the subjects often moved to the pointer's right side (*x*-axis) to judge the pointing position. Improving the camera arrangement reduces these errors because errors along the *y* and *z* axes were small. From these results, we can conclude that the method of switching cameras to employ the user's motion parallax effectively helps a user perceive pointing positions.

Table 1. Result of pointing experiment (in centimeters).*

Ball Spacing	*x* Error	*y* Error	*z* Error	Total Error
30 cm	15.0	3.0	0.0	17.6
	(16.3)	(9.1)	(0.0)	(16.4)
20 cm	18.0	4.4	1.6	21.1
	(14.1)	(8.4)	(5.5)	(14.0)
10 cm	13.8	5.4	3.8	17.5
	(12.3)	(6.8)	(5.7)	(12.5)
Total	15.6	4.3	1.8	18.8
	(14.4)	(8.1)	(4.8)	(14.4)

*Regular numbers represent averages, numbers in parentheses represent standard deviations.

Conclusion

To date, we can't transmit stereo video images for multiscreens completely because of insufficient network bandwidth and incomplete system equipment. However, Cabin and its extension into a networked environment, Cabinet, demonstrate the possibility of immersive communication via networked multiscreen displays. Future work will include sharing stereo video image worlds

using five screens and reducing the time lag of generating a video avatar using a broader communication bandwidth. MM

Acknowledgements

We thank Ken Tamagawa, Koji Hiratsuka, Tatsuhiro Tsuchida, and Kim Sungyoon for their assistance. This work was partly supported by the Telecommunications Advancement Organization.

References

1. H. Bullinger, O. Riedel, and R. Breining, "Immersive Projection Technology—Benefits for the Industry," *Proc. 1st Int'l Immersive Projection Technology Workshop*, Springer-Verlag, Stuttgart, Germany, 1997, pp.13-26.
2. C. Cruz-Neira, D.J. Sandin, and T.A. DeFanti, "Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE," *Proc. of Siggraph 93*, ACM Press, New York, 1993, pp.135-142.
3. M. Deering, "Making Virtual Reality More Real: Experience with the Virtual Portal," *Proc. of Graphics Interface 93*, Canadian Information Processing Society, Ontario, Canada, 1993, pp. 195-202.
4. C. Cruz-Neira, "Immersed in Science and Engineering: Projection Technology for High-Performance Virtual Reality Environments," *Proc. of ICAT 96*, Virtual Reality Soc. of Japan, Chiba, Japan, 1996, pp. 77-81.
5. S. Muller, "Experiences and Applications with a CAVE at Fraunhofer IGD," *Proc. 1st Int'l Immersive Projection Technology Workshop*, Springer-Verlag, Stuttgart, Germany, 1997, pp.97-108.
6. M. Hirose, "Development of an Immersive Multiscreen Display (Cabin) at the University of Tokyo," *Proc. 1st Int'l Immersive Projection Technology Workshop*, Springer-Verlag, Stuttgart, Germany, 1997, pp. 67-76.
7. M. Hirose et al., "Communication in Networked Immersive Virtual Environments," *Proc. 2nd International Immersive Projection Technology Workshop*, (Conf. organized by Iowa Center for Emerging Manufacturing Technology and the Fraunhofer Inst.), CD-ROM, 1998.
8. J. Leigh and A.E. Johnson, "Supporting Transcontinental Collaborative Work in Persistent Virtual Environments," *IEEE Computer Graphics and Applications*, Vol.16, No.4, 1996, pp. 47-51.
9. F. Hasenbrink and V. Lalioti, "Towards Immersive Telepresence Schlosstag 97," *Proc. 2nd Int'l Immersive Projection Technology Workshop*, (Conf. organized by Iowa Center for Emerging Manufacturing Technology and the Fraunhofer Inst.), CD-ROM, 1998.

Michitaka Hirose is an associate professor in the Department of Mechano-informatics, University of Tokyo. His main research interests are system engineering and human interface. He received a BEng in 1977 and Dr Eng in 1982 from the University of Tokyo.

Tetsuro Ogi is an associate professor at the Intelligent Modeling Laboratory, University of Tokyo. His main research interests are virtual reality and scientific visualization. He received a MEng in 1986 and Dr Eng. in 1994 from the University of Tokyo.

Toshio Yamada is a joint research worker between the University of Tokyo and Gifu Prefecture. His main research interest is immersive projection technology. He received a MEng in 1994 from the University of Fukui.

Readers may contact Ogi at the Intelligent Modeling Laboratory, The University of Toky, 2-11-16, Yayoi, Bunko-ku, Tokyo 113-8656, Japan, e-mail tetsu@iml.u-tokyo.ac.jp.