Immersive Telecommunication Using Stereo Video Avatar

Tetsuro Ogi, Toshio Yamada, Ken Tamagawa, Makoto Kano, Michitaka Hirose
Gifu MVL Research Center, TAO
IML, The University of Tokyo
{tetsu, yamada}@iml.u-tokyo.ac.jp, {tama, mkano, hirose}@cyber.rcast.u-tokyo.ac.jp

Abstract

Immersive projection displays such as CABIN and COSMOS have been connected through the broad band network. This kind of network environment is expected to be used as a multimedia virtual laboratory. In particular, video avatar technology has been developed in order to realize high presence communication in this multimedia virtual laboratory. A video avatar is a computer-synthesized three-dimensional image created using live video. This method has the characteristics of being a natural, accurate and convenient communication tool. In this study, communication capabilities of the video avatar were experimentally evaluated. In addition, the video avatar technology was applied to several communications applications, such as that of guiding colleagues and undertaking design work in networked immersive projection displays.

1. Introduction

Recently, multi-screen immersive projection displays such as CAVE have become very popular as virtual reality display systems[1][2]. Since this type of display surrounds the user with computer graphics stereo images projected onto screens, the user can feel a high sense of quality of immersion. On the other hand, advances in the network infrastructure have allowed us to transmit a large amount of data between remote places. Therefore, by connecting immersive projection displays using the broad band network, a high presence virtual world can be shared between remote places[3].

As a national project aimed at sharing a high presence virtual world using these networked immersive projection displays, the multimedia virtual laboratory (MVL) project has been promoted by the Ministry of Posts and Telecommunications in Japan. The multimedia virtual laboratory is a concept of linked network environments in which remote users can mutually communicate through the network as if they are in the same place. In order to realize the concept of the multimedia virtual laboratory, the construction of a high presence communication technology in the shared virtual world has been a key issue.

The computer graphics avatar is often used as a communication method in the networked virtual world[4]. Though this method represents the three-dimensional user's figure in the virtual world, it is difficult to represent natural facial expressions using the polygon model. A method of using a previously captured facial texture image has also been proposed[5]. However, it is difficult to represent instantaneous facial expression by deforming the original image of the face. In order to represent a high presence image of the user, the development of avatar technology using a video image is desired[6].

In the video conference system or associated communication systems such as ClearBoard[7], remote users can communicate face to face using their live video images. However, we cannot say that they are sharing a three-dimensional world, because these systems only transmit two-dimensional video images. In the InterSpace[8] or Virtual Human[9] systems, though a live video image is used to represent the avatar's face, its body is created using the polygon model. In the immersive projection displays, avatar technology using a full-length video image of the user would be effective.

Therefore, in this study, a stereo video avatar was developed. This technology realizes high presence communication within the networked immersive projection display by transmitting three-dimensional information about the user, such as position and gestures, by using a live video. This paper describes the video avatar technology developed in this study, and the communication capabilities of this method.

2. MVL network

The multimedia virtual laboratory is a high presence telecommunication environment constructed on the broad band network. This concept can be applied to various applications, such as collaborative design, remote conferencing, education and so on. For example, Figure 1 illustrates an example of a multimedia virtual laboratory being applied to a collaborative design task. In this figure, the computer scientist, the experimental engineer and the

Proceedings of the Virtual Reality 2001 Conference (VR'01)
0-7695-0948-7/01 $17.00 © 2001 IEEE

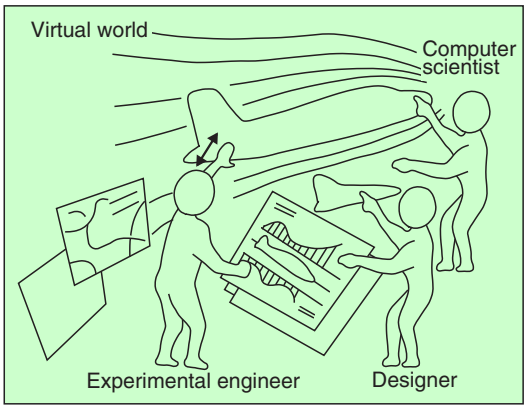

Figure 1. Concept of multimedia virtual laboratory

designer are jointly working in the shared virtual world to develop an airplane. Although these researchers are not usually in the same place, the multimedia virtual laboratory enables them to meet and hold discussions with each other through the network.

In this study, in order to realize the concept of the multimedia virtual laboratory, a network environment was constructed by connecting several immersive projection displays. Recently, the multi-screen immersive projection display has become popular, and several kinds of these displays have been developed. For example, CABIN was developed at the Intelligent Modeling Laboratory at the University of Tokyo[10]. CABIN is a CAVE-like cubic display that has five screens, situated at the front, on the left, right, on the ceiling and on the floor. In addition, COSMOS was developed at the Gifu Technoplaza and TEELeX was developed at the National Institute of Multimedia Education by extending the screen configuration of CAVE[11][12]. In COSMOS, a back screen was added to surround the user completely by six screens. Though TEELeX also has six screens, the ceiling screen is only half sized so that the floor image can be projected from the top. On the other hand, by simplifying the CAVE system, CoCABIN and UNIVERS were developed at the University of Tsukuba and at the Communications Research Laboratory respectively[13][14]. CoCABIN is a three-walled cubic display that surrounds the user, who sits at a table. UNIVERS is also a three-screen display, but the angles between the screens can be changed from a wall-type configuration to a cubic configuration.

Since the users in the multi-screen immersive projection displays are surrounded by computer graphics stereo images, they can experience a high presence virtual world. In this study, these immersive projection displays were connected through the Japan Gigabit Network (JGN) to construct the research environment of the multimedia virtual laboratory. JGN is a nationwide optical-fiber network developed by the Telecommunications Advancement Organization of Japan,

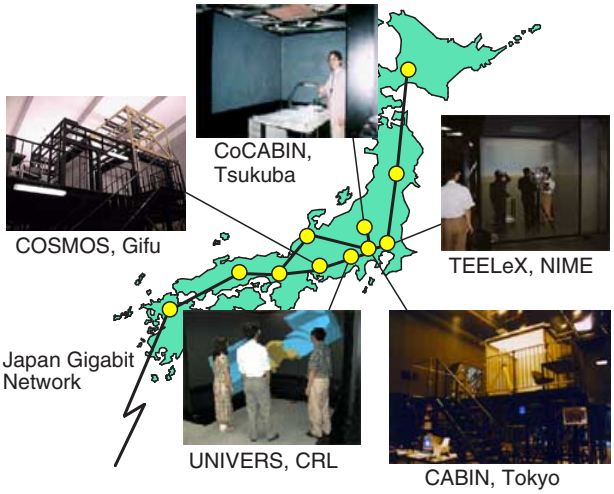

Figure 2. Network environment of multimedia virtual laboratory

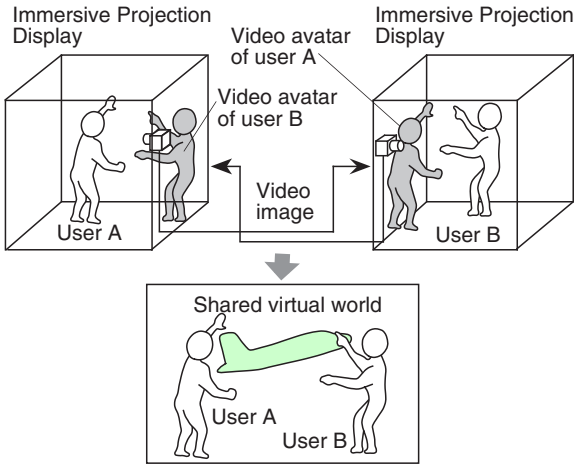

Figure 3. Video avatar communication in the networked immersive projection displays

and it has been used for research and development activities. In this system, CABIN, COSMOS, TEELeX, CoCABIN and UNIVERS are connected mutually by a 155Mbps ATM of the JGN. Figure 2 shows the network environment constructed in this study.

3. Video avatar

3.1 Concept of video avatar communication

In order to realize high presence communication in these networked immersive projection displays, video avatar technology was developed. The method that we propose can represent the user's figure using live video with a three-dimensional geometric model. In the networked environment, remote users can communicate with a high

Figure 4. Video avatar superimposed on the virtual world

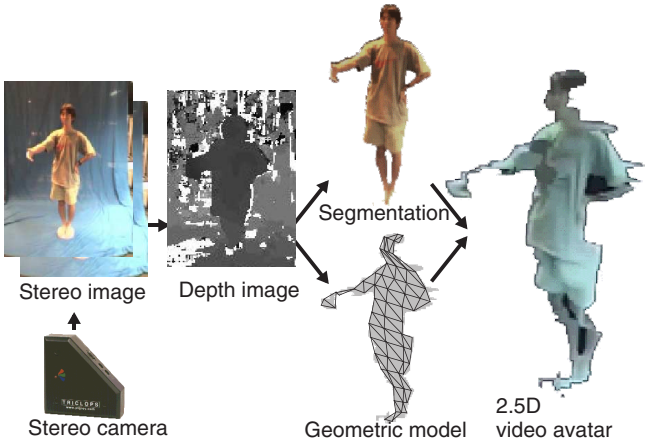

Figure 5. Basic process of making 2.5 dimensional video avatar

presence sensation by transmitting their video avatars mutually.

Figure 3 shows the concept of video avatar communication in the networked immersive projection displays[15]. In this method, the user's image is captured by a video camera placed within the immersive projection display, and only the user's figure is segmented from the background. This image is transmitted to the other site, and superimposed on the shared virtual world as a video avatar. In this case, by using a camera with wide field of view, a full-length video avatar can be created. Therefore, remote users can communicate face to face in the immersive virtual world using their video avatars.

Figure 4 shows an example of a video avatar superimposed on the shared virtual world.

3.2 Creation of a 2.5 dimensional video avatar

In order to generate a three-dimensional video avatar, a geometric model of the user must be created while the user's video image is being filmed. To achieve this, a stereo camera was placed within the immersive projection display, and it was used to capture the user's image. Figure 5 shows the basic process of making the stereo video avatar developed in this study. By using a stereo camera, depth data can be calculated for each pixel in the captured image using the stereo matching algorithm.

In this study, the Triclops Color Stereo Vision system made by Point Grey Research Inc. was used[16]. Since this camera consists of two pairs of stereo camera modules along the vertical and horizontal base lines, it can create an accurate depth image. The resolutions of the captured color image and the created depth image are 320x240 pixels and 160x120 pixels respectively, and the depth resolution is about 5.0cm. Since the field of view of this camera is 70 degrees, almost a full-length image of the user can be captured when it is

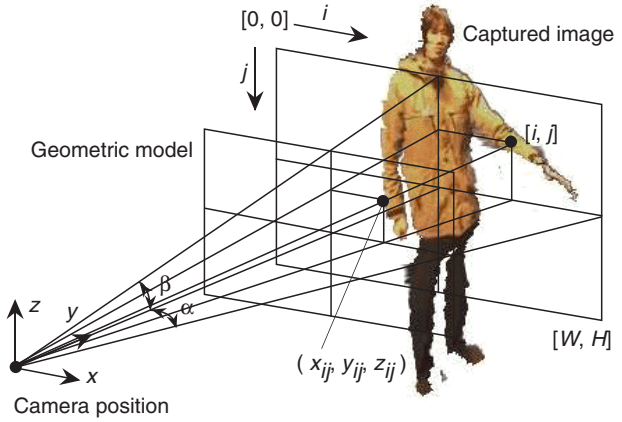

Figure 6. Geometric model of 2.5 dimensional video avatar

used in a multi-screen display such as CABIN.

Once the depth image is created, the user's figure can be segmented from the background by the threshold of the depth value. In practical applications, the chroma key can also be used in combination with the depth key to create a clear image of the avatar.

Additionally, the three-dimensional position of each pixel can be calculated using the depth data. When the depth data y_{ij} is known, the x_{ij} and z_{ij} coordinates of the pixel can be calculated from the following equations.

$$x_{ij} = y_{ij}\,(i\text{-}W/2)/(W/2)\,\tan\alpha \qquad (1)$$

$$z_{ij} = y_{ij}\,(H/2\text{-}j)/(H/2)\,\tan\beta \qquad (2)$$

where α and β are horizontal and vertical viewing angles of the stereo camera, and W and H are the width and height pixel numbers of the captured image as shown in Figure 6. By connecting the pixel positions using a triangular mesh, a three-dimensional geometric model can be created.

Thus, a stereo video avatar can be generated by texture-mapping the user's segmented image onto the geometric

Figure 7. Appearance of 2.5 dimensional video avatar seen from various directions

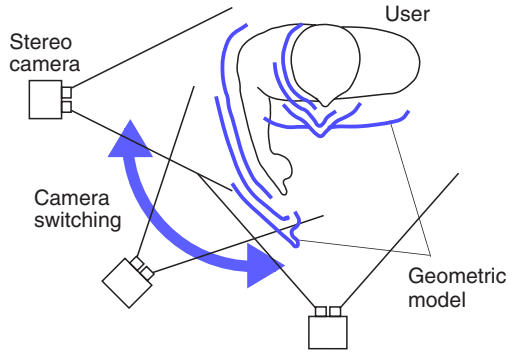

Figure 8. Change of geometric model of the video avatar by camera switching

model that has been created. Since this avatar only has surface model information for the side that faces toward the stereo camera, it can be called a "2.5 dimensional video avatar".

3.3 Switching avatar model

When a 2.5 dimensional video avatar is seen from a viewpoint close to the camera position, the visualized image of the avatar is well formed. However, when the viewpoint moves away from the camera position, the avatar's image becomes distorted. Figure 7 shows the appearance of a 2.5 dimensional video avatar seen from various directions. In this study, several stereo cameras were placed in the immersive projection display surrounding the user, and the closest camera to the other user's viewpoint was selected and used. By switching the selected camera, the modeled part of the avatar is changed as it tracks the other user's viewpoint, as shown in Figure 8, and therefore the distortion of the avatar's image is kept to a minimum. Thus, in effect, a quasi three-dimensional video avatar is generated.

A computer vision technique such as 'virtualized reality' that can create a complete three-dimensional model has been proposed as a method of creating the required geometric model from the video image[17]. However, this method cannot be performed in real-time, because it requires a significant computation time in order to perform the calculation that combines all of the video images captured by the multi-camera system. On the other hand, since our proposed method uses only one pair of stereo cameras at any given time, it can create the geometric model in real-time. When a Pentium III 700MHz PC is used, it can generate the video avatar at a refresh rate of about 9.9Hz, and the time delay was about 0.6 sec.

4. Communication capabilities of video avatar

In the multimedia virtual laboratory, a natural, accurate and convenient communication technology is required. Since the video avatar is a computer-synthesized image created by using live video images, it has the characteristics of both a natural communication tool and a computer supported tool. In this chapter, the communication capabilities of the video avatar are experimentally evaluated.

4.1 Natural representation of avatar

The proposed video avatar technology generates a quasi three-dimensional geometric model by switching some stereo cameras according to the observer's viewpoint. In this section, the condition for determining the camera intervals at which the appearance of the avatar's image would be naturally switched was evaluated experimentally. In this experiment, several stereo cameras were placed surrounding the user, and the intervals between these cameras were changed from 5 degrees to 90 degrees. The generated video avatar was visualized in the CABIN, and the subject in the CABIN was asked to evaluate on a five-grade system how natural the avatar's image was when switched.

The experiment made a comparison between using the 2.5 dimensional video avatar and using a two-dimensional video avatar in which video image was texture-mapped onto a flat plate. Figure 9 shows the results of this experiment for six subjects. In this graph, averaged values of the evaluated grades for each camera interval were shown. From these results, we can conclude that the subjects felt the 2.5 dimensional video avatar was more natural when compared with a two-dimensional video avatar. In particular, in the case of using the two-dimensional video avatar, when the subject of the video avatar extended his hand, the perceived hand position was variable according to the camera-switching, because the hand position was away from the center of the body. Judging from this graph, we can conclude that a camera interval angle of less than 45 degrees (i.e. where the evaluation grade is over 3.0) is desirable to represent a

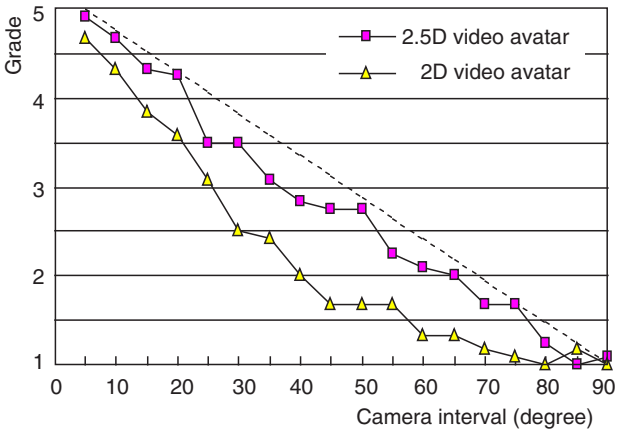

Figure 9. Experimental results of natural camera switching

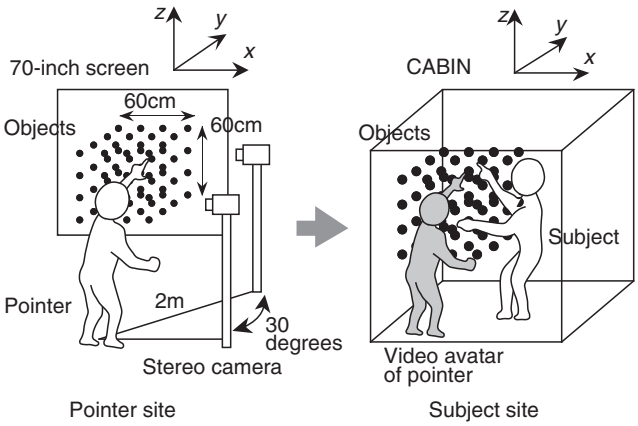

Figure 10. Experimental condition of finger pointing using video avatar

natural image of the avatar.

4.2 Accuracy of pointing task

Next, the accuracy of video avatar communication was experimentally evaluated. In the multimedia virtual laboratory, remote users held discussions while looking at the same data, such as a design model or scientific data in the shared virtual world. In this situation, finger pointing is an important function to realize a natural and accurate style of communication. Namely, it is necessary that one user can accurately recognize the position pointed at by the other user's fingertip during the collaborative work. Since the proposed video avatar has a geometric model, it can be used to indicate a three-dimensional position in the shared virtual world. In this section, a pointing experiment using the video avatar was conducted between the CABIN and a 70inch one-screen display connected via a fast Ethernet.

In this experiment, the user in front of the one-screen display was asked to point at one of a number of small balls placed at the grid points of a cubic lattice in the shared virtual world as shown in Figure 10. Two stereo cameras were placed at a 30 degrees interval at the 'pointers' site, and the video avatar of the pointer was sent to the CABIN site. Then, the observer in the CABIN was asked to orally identify which ball was pointed at while looking at the video avatar. The spacing between the balls was varied between 10cm and 30cm, and when the answer was wrong, the positional error was calculated from the distance between the object that was pointed at and the object given as the answer.

Five subjects undertook the test, and the results of the experiment are as shown in Table 1. From these results, the error along the x-axis was large, because the subjects mostly looked at the pointer from this direction, and the average error was 7.4cm. This error is thought to be unavoidable, because the depth value measured by the stereo camera

Table 1. Results of perceived position error in pointing experiment

Grid spacing	x(depth)	y(side)	z(height)	Total
30cm	2.4	0.6	0.0	2.6
20cm	6.4	1.2	0.8	7.5
10cm	9.4	3.6	1.8	12.1
Average	6.1	1.8	0.9	7.4
				(cm)

contains errors of about 10cm. Therefore, we can understand that the stereo video avatar itself functioned effectively to transmit the positional information in the shared virtual world.

4.3 Recording user's behavior

Since the video avatar is a computer-synthesized image, it can be used not only for natural communication but also as a computer-supported tool. For example, by using the image of the video avatar, past behavior of the user can be recorded in the virtual world. Figure 11 shows an example of recording the user's behavior. In this example, a snapshot of the user's figure was recorded at the position where it occurred in the virtual world using the image of the video avatar. In this picture, the left figure of the video avatar represents a record of past behavior, and the right one is the present user's figure.

By using this function, the user can record his actions while he moves within the virtual world. In applications requiring collaborative work such as using finger pointing, the user can indicate more than one object using the record of his figure. Therefore, the video avatar can also be used effectively as a computer-synthesized communication tool by incorporating a high presence image of the user.

Figure 11. Recording snapshot using video avatar

5. Application of video avatar communication

5.1 System configuration

The video avatar technology was applied to communications between immersive projection displays. Figure 12 shows the system configuration of the communication experiment that was conducted between CABIN at the University of Tokyo and COSMOS at the Gifu Technoplaza. The users in the CABIN and COSMOS can move their positions in the virtual world by using the joystick type input device respectively.

Two stereo cameras were placed at each site, and the selected camera was switched according to the positional relationship between the remote users. The PC captured the stereo image of the user, and the generated video avatar was transmitted to the observer's site through the graphics workstation. In this experiment, blue backdrops were hung at the entrances of CABIN and COSMOS in order to use the chroma key together with the depth key to segment a clear image of the user. In addition, an MPEG encoder and a decoder were used to transmit the scene to and from the opposite sites, and they were also used to transmit the voice of the video avatar.

5.2 Application examples

Several experiments were conducted on the networked immersive projection displays involving sharing the virtual world using the video avatar technology. For example, Figure 13 shows an example of sharing a wide area of the virtual world. In this example, remote users shared the virtual town, and one user guided the other user while apparently walking together. In this experiment, the video avatar was effectively used to mutually transmit the user's positions, and to represent body actions such as directing the way in the virtual town.

Figure 14 shows another example of applying the video avatar to collaborative work. In this example, users in remote locations held discussions with each other while apparently standing across a design model in the shared virtual world. In this experiment, remote users were able to hold effective discussions, including using the finger pointing function of the video avatar. In particular, they sometimes used the snapshot function of the video avatar to point at more than one object.

6. Conclusions

In this study, the video avatar was developed as a communication technology in a multimedia virtual laboratory constructed on the broad band network. This method can realize high presence communication in the shared virtual world by transmitting a three-dimensional live video of the user. Since the video avatar is a computer-synthesized image created using live video, it can be effectively used for natural, accurate and convenient communication. In this study, the communication capabilities of the video avatar were experimentally evaluated. In addition, the video avatar was

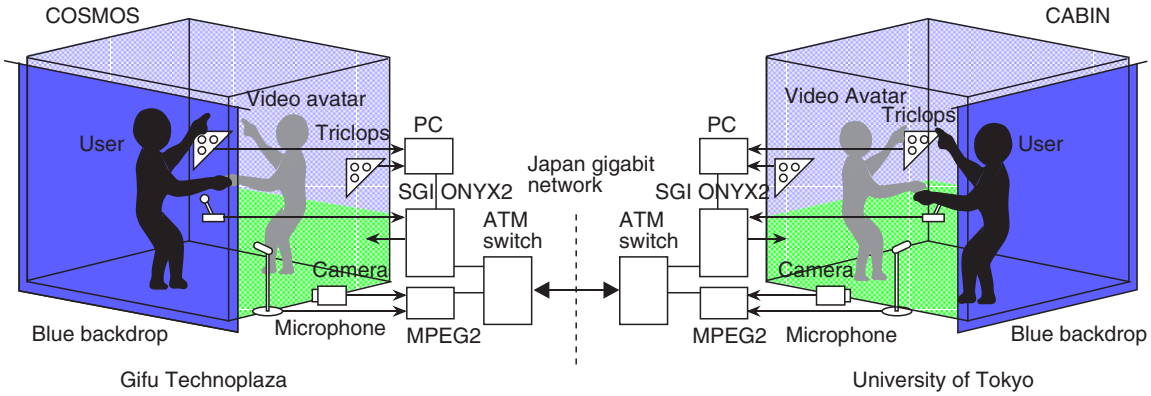

Figure 12. System configuration of the communication experiment using video avarar between CABIN and COSMOS

Figure 13. Application of the video avatar to the guide in the virtual town

Figure 14. Application of the video avatar to the collaborative design

applied to several communication situations, such as the guide in the virtual town and the collaborative design application to verify its effectiveness.

Although in this study the video avatar was applied to communication between two remote sites, it can also be applied to communication across several sites. Future work will include applying this technology to collaborative work among several sites in order to realize the multimedia virtual laboratory.

Acknowledgments

We thank Dr. Tom DeFanti and Dr. Dan Sandin of the University of Illinois at Chicago for their useful discussions. We also thank Dr. Hideaki Kuzuoka of the University of Tsukuba and Mr. Yoshiki Arakawa and Mr. Kenji Suzuki of the Communications Research Laboratory for their help in the communication experiments.

References

[1] Cruz-Neira C., Sandin D.J., DeFanti T.A., Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE, Proceedings of SIGGRAPH'93, pp.135-142, 1993

[2] Bullinger H., Riedel O., Breining R., Immersive Projection Technology- Benefits for the Industry, International Immersive Projection Technology Workshop, pp.13-25, 1997

[3] Leigh J., DeFanti T.A., Johnson A.E., Brown M.D., Sandin D.J., Global Tele-Immersion: Better Than Being There, ICAT'97, pp.10-17, 1997

[4] Leigh J., Johnson A.E., Vasilakis C.A., DeFanti T.A., Multi-Perspective Collaborative Design in Persistent Networked Virtual Environments, Proceedings of the IEEE VRAIS, pp.253-260, 1996

[5] Moroshima S., Yotsukura T., Face-to-face Communicative Avatar Driven by Vioce, ICIP'99, 1999

[6] Insley J., Sandin D.J., DeFanti T.A., Using Video to Create Avatars in Virtual Reality, Visual Proceedings of 1997 SIGGRAPH, pp.128, 1997

[7] Kobayashi M., Ishii H., ClearBoard: A Novel Shared Drawing Medium that Supports Gaze Awareness in Remote Collaboration, IEICE Trans. Communications, Vol.E76-B, No.6, pp.609-617, 1993

[8] Sugawara S., Suzuki G., Nagashima Y., Matsuura M., Tanigawa H., Moriuchi M., InterSpace- Networked Virtual World for Visual Communication, IEICE Trans. Information & Systems, Vol.E-77D, No.12, pp.1344-1349, 1994

[8] Capin T.K., Noser H., Thalman D, Pandzic I.S., Thalman N.M., Virtual Human Representation and Communication in VLNet, IEEE Computer Graphics and Applications, Vol.7, No.2, pp.42-53, 1997

[10] Hirose M., Ogi T., Ishiwata S., Yamada T., Development and Evaluation of the Imersive Multiscreen Display CABIN, Systems and Computers in Japan, Vol.30, No.1, pp.13-22, 1999

[11] Yamada T., Hirose M., Iida Y., Development of Complete Immersive Display: COSMOS, Proceedings of VSMM98, pp.522-527, 1998

[12] Asai K., Sugimoto Y., Saito F., Multi-Screen Display with Liquid Crystal Projectors, Proceedings of ISMCR'99, pp.253-258, 1999

[13] Hirose M., Ogi T., Yamada T., Tanaka K., Kuzuoka H., Communication in Networked Immersive Virtual Environments, 2nd International Immersive Projection Technology Workshop, 1998

[14] Arakawa Y., Kakeya H., Isogai M., Suzuki K., Yamaguchi F., Space-shared Communication based on Truly 3D Information Space, ICIP'99, 1999

[15] Hirose M., Ogi T., Yamada T., Integrating Live Video for Immersive Environments, IEEE Multimedia, Vol.6, No.3, pp.14-22, 1999

[16] Point Grey Research Inc. web page: http://www.ptgrey.com/

[17] Kanade T., Rander P.W., Virtualized Reality: Being Mobile in a Visual Scene, Proceeding of ICAT/VRST'95, pp.133-142, 1995