Live Video Integration for High Presence Virtual World

Tetsuro OGI, Toshio YAMADA
Gifu MVL Research Center, TAO
IML, The University of Tokyo
2-11-16, Yayoi, Bunkyo-ku, Tokyo 113-8656, Japan

Michitaka HIROSE
RCAST, The University of Tokyo
4-6-1, Komaba, Meguro-ku, Tokyo 153-8904, Japan

and

Kenji SUZUKI
Communications Research Laboratory
4-2-1 Nukui-kita, Koganei, Tokyo 184-8795, Japan

ABSTRACT

By using the live video image in the virtual world, a realistic dynamic scene can be constructed. In this study, the immersive projection display CABIN was connected with the Japan Gigabit Network, and the high presence virtual world was constructed using the live video images transmitted from the remote sites. In order to integrate the two-dimensional video image naturally with the three-dimensional virtual world, the real-time video based modeling and rendering technique was developed. In particular, the experiments on constructing the virtual stadium and the video avatar were conducted as examples of using the live video images of the distance view and the short-range view.

Keywords: Virtual Reality, Broadband Wide Area Network, Video Based Virtual World, Virtual Stadium, Video Avatar.

1. INTRODUCTION

In order to construct a realistic virtual world, several technologies that use photograph images have been developed and used. For example, by using the texture mapping or the image based rendering techniques, a three-dimensional realistic scene seen from various viewpoints can be generated without the detailed geometric models [1][2]. However, in these methods, it is difficult to represent a dynamic scene in which photographic objects are moving or changing the shapes, because these methods usually use the static images which were taken in the real world beforehand.

On the other hand, recent advances in the infrastructure of the broadband wide area networks have allowed us to transmit a large amount of data between remote places. For example, it has become possible to transmit a live video image from the remote site in real-time. Therefore, it is expected that the realistic dynamic scene can be generated, by integrating the transmitted video image with the virtual world. Although the video texture method has been often used to generate virtual worlds, it only synthesizes the video image as a two-dimensional data. In this study, the video based virtual world that constructs a three-dimensional realistic virtual world by integrating the live video image transmitted through the network was developed.

This paper describes the concept of the video based virtual world, and discusses the several experiments on constructing the dynamic virtual world using the live video image, such as the virtual stadium and the video avatar.

2. VIDEO-BASED VIRTUAL WORLD

The video-based virtual world that is proposed in this study is a technology to construct a highly realistic virtual world by using a live video image transmitted through the network. In this method, though the captured video image is a two-dimensional data, it should be integrated with the three-dimensional virtual world naturally. In other words, the video image should not be used as a simple two-dimensional data, such as the virtual television or the virtual screen, in the three-dimensional virtual world. In addition, it is desired that the generated virtual world can be displayed in the immersive displays such as the CAVE or the CABIN to generate a high presence virtual world [3].

In order to integrate the two-dimensional video image with the three-dimensional virtual world naturally, the effective video based modeling and rendering techniques should be used according to the captured video images. For example, when the video image of a distant view is captured, it can be simply texture mapped on the polygon face, because the binocular parallax has a small influence on the stereo view. Therefore, the integrated video image can be seen naturally, when the appropriate polygon is selected and the view dependent distortion is corrected. On the other hand, when the video image of a short-range view is used, the influence of the binocular parallax is very large to recognize the three-dimensional object in the scene. Therefore, it is necessary to generate the geometric model or the view dependent image to integrate it with the three-dimensional virtual world.

In order to generate a view dependent image from the two-dimensional captured images, several image based modeling and rendering techniques have been developed. However, in the video

Figure 1. Immersive Projection display CABIN

based virtual world, all the modeling and the rendering process must be completed in real-time. Therefore, in this study, the effective video based modeling and rendering technologies that can be used to integrate the live video with the three-dimensional virtual world are discussed. In particular, the experiments on constructing the high presence virtual worlds using the live video images of the distant view and the short-range view were conducted.

3. NETWORKED VIRTUAL REALITY ENVIRONMENT

In this study, as a virtual reality display system, the CABIN at the University of Tokyo was used [4]. The CABIN is a CAVE-like immersive projection display that has five screens at the front, on the left, right, ceiling and floor as shown in Figure 1. Since the user in the CABIN is surrounded by the stereo images projected on the screens, he can experience a high presence virtual world with a wide viewing field. Therefore, the virtual world displayed in the CABIN itself should be constructed as a high presence model compared with the image displayed on the CRT.

In addition, in this study, the CABIN was connected to the Japan Gigabit Network to utilize the live video images transmitted from the remote sites. The Japan Gigabit Network is a nation wide optical fiber network equipped by the Telecommunications Advancement Organization of Japan in 1999, and it has bee used for research and development activities [5]. In this system, the live video image filmed by the video camera can be transmitted to the CABIN through the Japan Gigabit Network, and it can be used to construct a realistic virtual world. The following sections of this paper discuss the several examples of constructing high presence virtual worlds using the live video images transmitted through the network, such as the virtual stadium and the video avatar.

4. VIRTUAL STADIUM

Video Based Rendering for the Distant View Image

As an experiment of constructing the video based virtual world using the video image of the distant view, the virtual soccer stadium was constructed. This experiment aimed at enabling the user to watch the live soccer game in the virtual stadium displayed in the

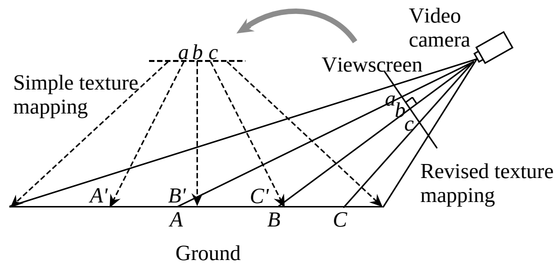

Figure 2. Distortion and correction in the video texture mapping

immersive projection display. Though the ground of the soccer stadium is not strictly flat because many players are playing on it, the scene seen from the distant viewpoint can be approximated as a two-dimensional plane. In the case of the distant view image, the three-dimensional information based on the binocular parallax or the motion parallax is not essential. Therefore, the video image of the soccer ground can be texture mapped on the simple two-dimensional plane.

However, in order to integrate the two-dimensional image with the three-dimensional virtual world, the distortion of the video image according to the camera position should be revised. Figure 2 shows the distortion of the captured video image and the correction method in the video texture mapping. Since the scene captured by the camera is a projected image onto the viewscreen perpendicular to the viewing direction, each point on the ground is projected onto the corresponding point on the screen plane respectively. On the other hand, the polygon face on which the video image is texture mapped should be positioned at the ground plane so that it can be seen at the correct position in the three-dimensional virtual world. Therefore, if the captured image was simply texture mapped on the polygon face of the ground plane, it would be distorted due to the difference of the angles between the ground plane and the projected plane.

Therefore, in this method, each point in the captured image was texture mapped at the correct point that was calculated back according to the camera position. For example, in Figure 2, the point *a* in the video image was texture mapped at the point *A* on the ground plane, and the point *b* was mapped at the point *B*. By revising the distortion of the captured video image due to the camera position, the user can see the correct video image in the three-dimensional virtual world. In the video based virtual world, this process should be completed in real-time.

Virtual Stadium

In this study, the experiment on constructing the virtual stadium by using the live video image that was transmitted from the actual stadium was conducted. Figure 3 shows the system configuration used in this experiment. Two HDTV cameras were placed at the seats in the National Kasumigaoka Stadium, and the filmed images were transmitted to the University of Tokyo and the Communications Research Laboratory through the Japan Gigabit Network and the HDTV Transmission Service. Since one HDTV camera can view the half area of the ground, the whole ground can be captured by using two HDTV cameras. Namely, one camera filmed the left half side, and the other camera filmed the right side

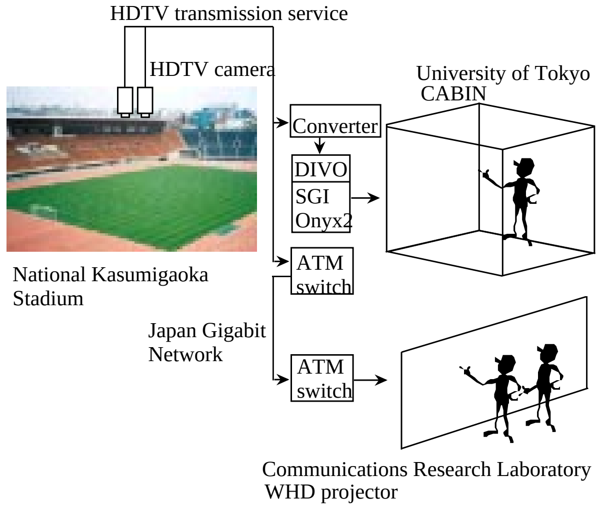

Figure 3. System configuration for the virtual stadium

Figure 4. Integrated video image in the virtual stadium

of the ground.

At the Communications Research Laboratory site, the high-resolution WHD (Wide/Double HD) projector display was equipped, and the transmitted HDTV images were directory displayed on it [6]. On the other hand, at the University of Tokyo site, the transmitted video images were used to construct the video based virtual world, and it was displayed in the CABIN. In this method, the three-dimensional polygon based computer graphics model of the soccer stadium was created, and only the ground part was texture mapped using the live video image. The transmitted HDTV images were converted to the NTSC images and they were captured by the SGI Onyx2 graphics workstation through the DIVO (digital video I/O option). From the captured images, only the ground parts were segmented and they were texture mapped on the ground part of the computer graphics model of the soccer stadium. When the video image was texture mapped on the ground, the above-mentioned video based rendering method was used to correct the distortion. Thus, the user in the CABIN was able to watch the live soccer game in the immersive virtual world.

Figure 4 shows the example of the integrated live video image in

Figure 5. Experience of virtual stadium in the CABIN

the virtual soccer stadium, and Figure 5 shows that the user is experiencing the virtual stadium in the CABIN display. In this experiment, since the NTSC video images were used in the virtual world, the resolution of the displayed image was not sufficient compared with the image displayed in the high-resolution WHD projector display at the Communications Research Laboratory. However, the user in the CABIN was able to experience the high presence sensation watching the live soccer game in the three-dimensional immersive virtual environment.

5. VIDEO AVATAR

Video Based Rendering for the Short-Range View Image

Next, as the example of using the live video image of the short-range view, the video avatar was developed [7]. The video avatar is a computer synthesized image of the human by using the live video. If the realistic image of the human was synthesized in the virtual world, the reality of the virtual world itself would be improved. In addition, in order to communicate naturally with the virtual human existing in front of the user, it should be represented using the three-dimensional stereo image. However, it is difficult to represent the natural image of the human by using the polygon based computer graphics model. Therefore, it is desirable to use the live video image captured by the video camera to generate the video avatar. In this study, the three-dimensional stereo video avatar technology was developed in order to synthesize the natural human image in the three-dimensional virtual world.

Figure 6 shows the basic process of making stereo video avatar using the live video image. In this method, the stereo video camera was placed in front of the user to film the user's image. From the stereo image, the depth data can be calculated using the stereo matching algorithm. 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, by connecting the pixels in the depth image using 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 model. This method can also be applied to represent the realistic three-dimensional object using the video image.

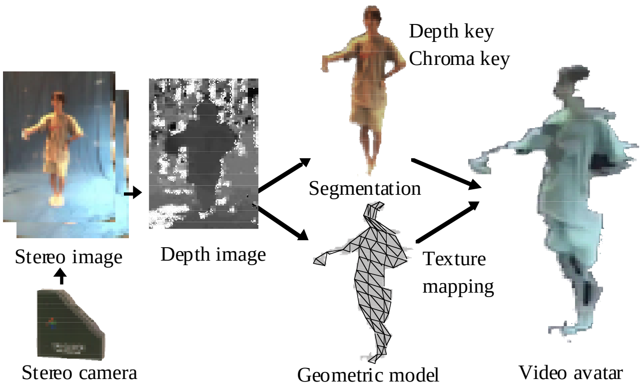

Figure 6. Process of making video avatar

Figure 7. Video avatar siperimposed on the virtual own

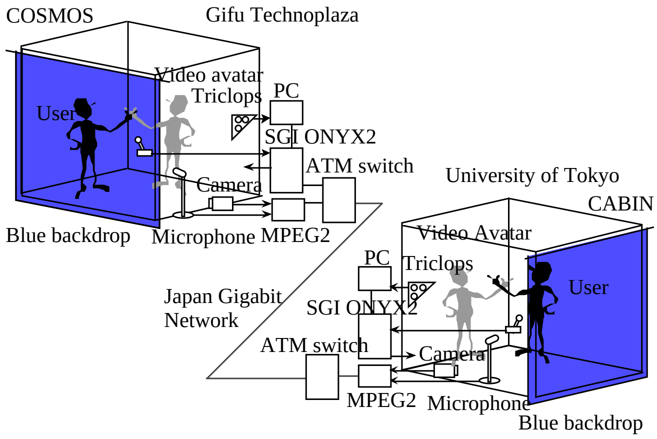

Figure 8. System configuration between immersive projection displays

Figure 9. Video avatar communication in the shared virtual world

Video Avatar Communication

By superimposing the video avatar on the virtual world, high presence virtual world with a live human can be constructed. For example, in this study, the blue back studio was connected to the CABIN through the Japan Gigabit Network to superimpose a video avatar in the immersive virtual world. Figure 7 shows the example of the video avatar superimposed on the virtual town. This video avatar data was synthesized in about 9.9Hz, and the user's figure could be represented in real-time in the virtual world.

This video avatar can also be used for the communication tool in the networked virtual environments. In this study, the immersive projection displays, the CABIN and the COSMOS were connected through the Japan Gigabit Network. The COSMOS is a six-sided display equipped at the Gifu Technoplaza. By transmitting the video avatar mutually between the CABIN and the COSMOS, the remote users can communicate with each other in the high presence shared virtual world. Figure 8 shows the system configuration of the networked immersive environment between the CABIN and the COSMOS. In this experiment, the Triclops Color Stereo Vision system made by Point Grey Research was used for the stereo camera system. Since this camera system consists of two pairs of stereo camera modules along the horizontal and vertical base lines, it can create an accurate depth image. Figure 9 shows the example of the video avatar communication in the shared virtual world. In this example, the remote users were discussing using the video avatar while sharing the same object in the high presence shared virtual world. This communication technology can be effectively used in the several fields of application, such as the collaborative design and the education between remote places.

6. CONCLUSIONS

In this study, the concept of the video based virtual world was proposed. This technology can generate a high presence virtual world by integrating the live video image transmitted from the remote site through the network with the computer graphics world. In this case, in order to integrate the two-dimensional video image with the three-dimensional world naturally, the video based modeling and rendering technology that can be used in real-time is necessary. In this paper, the experiments on constructing the virtual stadium and the video avatar were shown for the examples of using the video images of the distance view and the short-range view. Future work will include using high-resolution video images such as multiple HDTV images and evaluating the effectiveness of this technology.

ACKNOWLEDGMENT

We would like to thank Yoshiki Arakawa, Kenji Tanaka at the Communications Research laboratory for their help in the experiment on the virtual stadium. We also would like to tank Ken Tamagawa, Makoto Kano, Kazuhiko Hirose, Takashi Imamura at the University of Tokyo for their help in the experiment on the video avatar communication.

REFERENCES

[1] P.E. Debevec, G. Borshukov, Y. Yu, "Efficient View-Dependent Image-Based Rendering with Projective Texture-Mapping", 9th Eurographics Rendering Workshop, 1998.

[2] M. Hirose, T. Tanikawa, T. Endo, "Building a Virtual World from the Real World", Mixes Reality, Ohmsha, pp.183-197, 1999

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

[4] M. Hirose, T. Ogi, S. Ishiwata, T. Yamada, "Development and Evaluation of Immersive Multiscreen Display CABIN", Systems and Computers in Japan, Scripta Technica, Vol.30, No.1, January, pp.13-22, 1999.

[5] http://www.jgn.tao.go.jp/.

[6] K. Tanaka. K. Suzuki, K. Ohyama, Y. Arakawa, "WHD(Wide/Double HD) Transmission System", Correspondences on Human Interface, pp.53-56, 2000 (in Japanese).

[7] T. Ogi, T. Yamada, M. Kano, M. Hirose, "Immersive Telecommunication Using Stereo Video Avatar", Proceedings of IEEE Virtual Reality 2001 Conference, pp.45-51, 2001.