Real World Video Avatar: Transmission and Presentation of Human Figure

Hiroyuki Maeda[1)] Tomohiro Tanikawa[2)] Jun Yamashita[3)] Koichi Hirota[3)] Michitaka Hirose[3)]

Graduate School of Information Science and Technology, The University of Tokyo[1)]
SVR Research Center, TAO[2)]
Research Center for Advanced Science and Technology, The University of Tokyo[3)]
{maeda, tani, jun, hirota, hirose}@cyber.rcast.u-tokyo.ac.jp

1. Introduction

Video avatar[1] is one methodology of interaction with people at a remote location. By using such video-based real-time human figures, participants can interact using non-verbal information such as gestures and eye contact. In traditional video avatar interaction, however, participants can interact only in "virtual" space.

We have proposed the concept of a "real-world video avatar", that is, the concept of video avatar presentation in "real" space. One requirement of such a system is that the presented figure must be viewable from various directions, similarly to a real human. In this paper such a view is called "multiview". By presenting a real-time human figure with "multiview", many participants can interact with the figure from all directions, similarly to interaction in the real world. A system that supports "multiview" was proposed by Endo et al.[2], however, this system cannot show real-time images.

We have developed a display system which supports "multiview"[3]. In this paper, we discuss the evaluation of real-time presentation using the display system.

2. System Design

The conceptual idea is to capture a person from multiple directions and present each image in the same direction as the captured direction (see Figure 1). The method consists of the capturing process using multiple cameras and the presentation process using a revolving display. By revolving the display and updating the image according to the orientation of the display panel, each image captured from various directions can be presented in each direction. It is important to "reduce" the viewing angle of the display in order to make the image on the display panel visible from only the front. It should be noted that our approach is conceptually different from that of volumetric visual displays

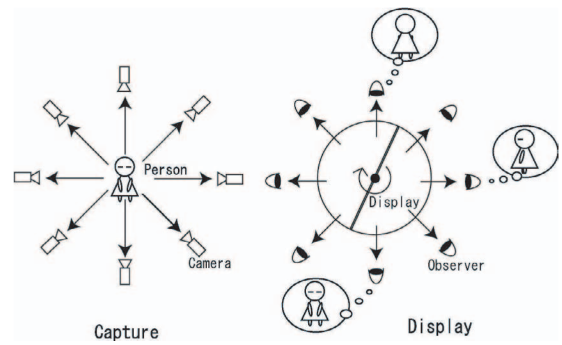


Figure 1. Conceptual Image

such as [4]; in these displays, display panels or screens are moved to scan a volume to generate a 3D image.

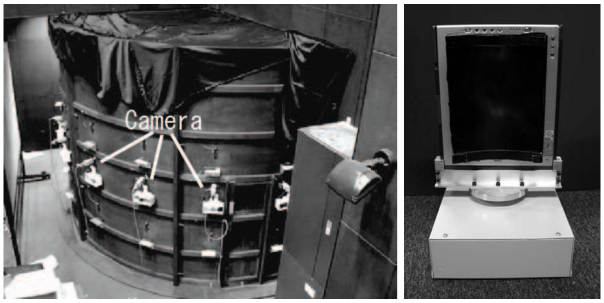
3. Implementation

3.1. Capture System

Figure 2(a) shows the exterior of the capture system. The capture system consists of a cylindrical chamber whose radius is 2000[mm], eighteen cameras, eighteen "node PCs", and one "central PC". The interior wall and floor of the chamber are painted blue for the convenience of chroma-key processing. The cameras are located on the wall at 20-degree intervals at a height of 1200[mm] from the floor. The optical axes of the cameras cross at the center of the chamber. Each camera is connected to one node PC via IEEE1394, and all node PCs are connected to the central PC via Gigabit Ethernet. Each node PC captures an image on a connected camera, discards a background of a human figure using chroma-key processing, converts the image to JPEG format, and sends it to the central PC. The central PC receives these images and sends to the display system via Ethernet.

IEEE Virtual Reality 2004 March 27-31, Chicago, IL USA
0-7803-8415-6/04/$20.00©2004 IEEE.

Proceedings of the 2004 Virtual Reality (VR'04)
1087-8270/04 $ 20.00 IEEE



(a) Capture System (b) Display System

Figure 2. Capture and Display System

Figure 3. Captured and Presented Images

3.2. Display System

The display system consists of a tablet PC (NEC) and a revolving mechanism driven by a stepping motor (see Figure 2(b)). The system is approximately 400[mm] in height, and the display panel is a 10.4-inch XGA LCD whose refresh rate is fixed at 60[Hz]. To reduce the viewing angle, a privacy computer filter (3M) is used. The filter is formed so that it is an arc whose center is identical to the ideal viewpoint, in order to make the entire area of the display panel visible from the ideal viewpoint. The radius of curvature of the filter is 200[mm]. The images obtained using the capture system are sent to the display system via IEEE802.11a wireless LAN. Received images are temporarily stored in memory and displayed with suitable timing.

3.3. Performance

In our implementation, the maximum updating rate of the image on the display panel depends on the resolution of images. The updating rate of 30[fps] was attained when we used images with the resolution of 120x90 pixels. The updating rate must be equal to the product of the revolutions per second of the revolving display and the number of directions for presenting different images. We determined the revolutions per second to be 5[rps] and the number of direction to be 6.

A figure of a person who was standing within the space of the capture system was presented by the display system (See Figure 3). We also measured the delay time from capture to display. The average delay time was approximately 183 [ms]: 84[ms] in the capture system, 12[ms] for transmission, and 87[ms] in the display system.

4. Discussion

The present display system does not have sufficient drawing capability. Hence the updating rate and the resolution of the display system cannot be efficiently increased. It is required to use a superior computer. Another problem is that the response speed of LCD is low. Today, some flat panel display devices which have superior response speed such as EL displays and plasma displays are available. They will be alternative candidates to the LCD.

If we consider applying our system for remote communication, a reduction of the delay time is important. The delay is mainly caused by the chroma-key processing in the capture process and the temporary storage of the images in the display process. The improvement of the algorithm is one way to reduce the time of the chroma-key processing. A stored image must wait until the display faces the appropriate direction for the image. One way to solve the problem is to increase the revolving speed, which requires to increase the updating rate of the display system. We consider using our method for other applications, for example, a presentation of a 3D CG model with which participants can interact using some devices in the real world.

References

[1] T. Ogi, T. Yamada, K. Tamagawa, M. Kano and M. Hirose, Immersive Telecommunication Using Stereo Video Avatar, Virtual Reality 2001, pp. 45-51, 2001.

[2] T. Endo, Y. Kajiki, T. Honda and M. Sato, Cylindrical 3D Display Observable from All Directions, Proceeding of PG'00, pp. 300-306, 2000.

[3] H. Maeda, K. Hirose, J. Yamashita, K. Hirota and M. Hirose, "All-Around Display for Video Avatar in Real World", Proceedings of The Second IEEE and ACM International Sysmposium on Mixed and Augumented Reality (ISMAR03), pp. 288-289, 2003.

[4] J. Napoli, D. M. Hall, R. K. Dorval, M. G. Giovinco, M. J. Richmond and W. S. Chun, 100 Million-Voxel Volumetric Display, AeroSense 2002 - for Cockpit Displays IX: Displays for Defense Applications, 2002