Immersive Telepresence System Using High-resolution Omnidirectional Movies and a Locomotion Interface

Sei Ikeda, Tomokazu Sato, Masayuki Kanbara and Naokazu Yokoya

Graduate School of Information Science, Nara Institute of Science and Technology,
8916–5 Takayama, Ikoma, Nara, 630-0101 Japan

ABSTRACT

Technology that enables users to experience a remote site virtually is called telepresence.[1] A telepresence system using real environment images is expected to be used in the field of entertainment, medicine, education and so on. This paper describes a novel telepresence system which enables users to walk through a photorealistic virtualized environment by actual walking. To realize such a system, a wide-angle high-resolution movie is projected on an immersive multi-screen display to present users the virtualized environments and a treadmill is controlled according to detected user's locomotion. In this study, we use an omnidirectional multi-camera system to acquire images real outdoor scene. The proposed system provides users with rich sense of walking in a remote site.

Keywords: Telepresence, High-resolution Omnidirectional Movie, Multi-screen Display, Treadmill

1. INTRODUCTION

Technology that enables users to experience a remote site virtually is called telepresence.[1] A telepresence system using real environment images is expected to be used in a number of fields such as field of entertainment, medicine and education. Especially, telepresence using an image-based technique attracts much attention because it can represent complex scenes such as outdoor environments. Our ultimate form of telepresence is a system in which users can naturally move and look anywhere by their actions in a virtualized environment reproduced from a real environment faithfully. However such an ideal system does not exist today.

Conventional telepresece systems have two important problems. One is that high human cost is required to acquire images and to generate virtualized environments in the case of large-scale outdoor environments. The other is concerned with presentation of virtualized environments. Chen[2] has developed a method to generate a panoramic image by rotating a camera, and to present a part of a panoramic image as a virtualized environment in a standard display. Thereby, users can turn their view directions freely. He has additionally realized a system that enables users to turn their view directions and to move their view positions simultaneously by acquiring many omnidirectional images at various positions in advance. However, the image acquisition task takes much time and effort because multiple images should be captured to generate a panoramic image. Moreover, standard displays are not suitable to give a feeling of virtually walking in remote sites. Some works[3,4,5,6,7,8] have improved the method above. In these works, omnidirectional camera systems are used to reduce human cost in acquisition of images. In the telepresence system developed by Onoe, et al.,[3] multiple users can look around a scene of remote site in real time using a head mounted display. They used an omnidirectional video stream acquired by an omnidirectional video camera. More recent works[5,6,7] have improved the resolution of omnidirectional movies to be higher than Onoe's. However, users can not control their view positions in virtualized environments. Some other works[4,8] have addressed this problem. Kotake, et al.[4] used a multi-camera system radially arranged on a moving car to acquire high-resolution images of an outdoor scene. In

E-mail: {sei-i, tomoka-s, kanbara, yokoya}@is.aist-nara.ac.jp

(a) camera head (b) storage unit

Figure 1. Omnidirectional multi-camera system: Ladybug.

this system, an immersive three-screen display is used to provide users with the feeling of high presence in remote sites. In this case, a game controller is used to move the view position. Therefore, the system can not provide users with the sense of walking in a virtualized environment of a real outdoor scene.

In this paper, we propose a novel telepresence system which enables a user to move by actual walking and change his view point in a photorealistic virtualized environment using a high-resolution omnidirectional movie. For this system, first, movies of outdoor scenes are acquired by an omnidirectional multi-camera system (OMS). After calibrating the OMS geometrically and photometrically, a virtualized omnidirectional movie is generated by using an image-based representation from the captured multiple movies. Generated virtualized environment movies are projected on an immersive multi-screen display according to user's locomotion detected on a treadmill.

2. IMMERSIVE TELEPRESENCE SYSTEM USING HIGH-RESOLUTION OMNIDIRECTIONAL MOVIES

This section describes a method for realizing a telepresence system with a locomotion interface. Realization of the system requires three processes: acquisition of images of a real scene, generation of a virtualized environment and presentation to users. For the first process, we use an OMS: Ladybug[9] constructed of six cameras to acquire high-resolution movies and to reduce human cost in images acquisition. Ladybug is shown in Figure 1. In the second process, virtualized environment movies are generated from captured multiple image sequences. The OMS is calibrated geometrically and photometrically in order to generate a virtualized environment. The final process is to present it to users. We use an immersive three-screen display with a treadmill to present a generated virtualized environment to users. The following sections describe details of these processes.

2.1. Acquisition of Images of a Real Dynamic Scene

An OMS has an important advantage that human cost in acquisition of images can be reduced because of the following two reasons. One is that the OMS has a wide field of view. The other is that the total resolution of images captured by the OMS is usually higher than an omnidirectional camera system using a single camera. Therefore, any scenes can be usually captured with one time capturing without considering camera setting.

For the acquisition of images, movies are captured by an OMS Ladybug mounted on a moving car, as shown in Figure 2. This camera system obtains six 768×1024 images synchronously at 15 fps; five for horizontal views

Figure 2. An OMS mounted on a car.

and one for a vertical view, as shown in Figure 1 (a). Figure 1 (b) shows a storage unit, which consists of four hard disks. The camera system can collect movies covering more than 75% of the full spherical view with a little different apparent points of view. In this work, the speed of the car is kept constant because the replay speed of a generated virtualized environment should be controlled according to user's locomotion regardless of variation of the car speed in the presentation process.

2.2. Generation of Virtualized Environment

In this section, movies are generated according to the shape of screen of an immersive projection display by using geometric and photometric calibration techniques.[7] In this research, an OMS is calibrated geometrically and photometrically to automatically generate movies of virtualized environments. The following describes the calibration of an OMS and the generation of virtualized environment movies.

Calibration for OMS

In the geometric calibration, intrinsic and extrinsic parameters are estimated. The intrinsic parameters are concerned with the internal structure of each camera. The extrinsic parameters mean relative of positions and orientations among all the camera units. Therefore the extrinsic parameters should be determined in a unified coordinate system. To estimate those parameters, 3D positions of many markers are measured by using a calibration board and a total station. 3D coordinates of three corners of the calibration board are measured by the total station, and then all 3D positions of the markers on the board are calculated by linear interpolation among its corners. Intrinsic and extrinsic parameters of each camera are estimated using the obtained 3D positions of markers and their 2D positions in the captured images.[7]

In the photometric calibration, the limb darkening of each camera and color balances among multiple cameras are corrected in order. The limb darkening is a gradually decreasing effect of brightness in peripheral regions in images acquired by a wide-angle camera. The strength of the effect is calculated from estimated intrinsic parameters. In the color balance correction, we assume a linear relation between the radiance of the surface and the irradiance of the image. The difference of color balances between camera units is measured by a histogram matching of each RGB channel.

Generation of Virtualized Environment Movie

This step is based on re-projecting calibrated input images to a virtual image surface which corresponds to the screen shape of an immersive projection display. In advance, the limb darkening and the color balance of input images are corrected. Then, corrected images are projected on a projection surface by using the intrinsic

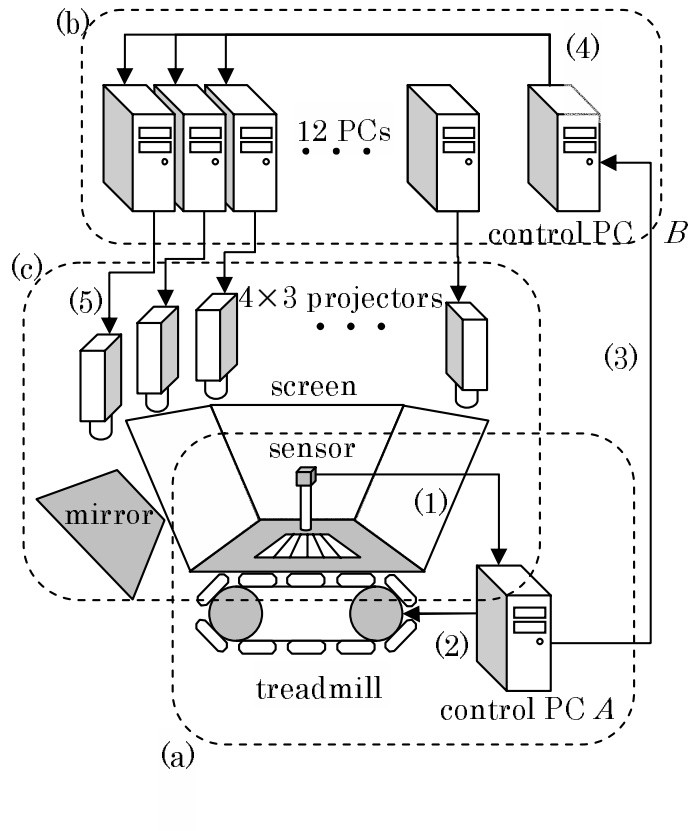

Figure 3. Components of the proposed system.

and extrinsic parameters obtained by the geometric calibration. Assuming a horizontal 13% overlap region between two adjacent cameras, the total horizontal resolution of Ladybug is about 3340 pixels.

Since the centers of projection of six camera units of the OMS are different from each other, the single viewpoint perspective projection model is not applicable for this system. However, when the distance of a target from the system is sufficiently large, the centers of projection can be considered as the same. Therefore, we assume that the target scene is far enough from the OMS and set the projection surface far enough from the camera system. A frame of a virtualized environment movie is generated by projecting all the pixels of all the input six images onto the projection surface. Note that a blending technique is used for generating a smooth image, when a point on the projection surface is projected from multiple images of different cameras.

2.3. Presentation to Users

This section describes a method for presenting virtualized environment movies generated in the previous section. As shown in Figure 3, our system is composed of (a) a locomotion interface, (b) a graphics PC cluster and (c) an immersive three-screen display. The locomotion interface detects user's motion as input data to the system, and sends calculated displacement information to the PC cluster. The PC cluster draws twelve images synchronized with the speed of user's walk on a treadmill because each screen image is generated by four projectors. As output data, these movies are displayed according to the user's motion so that the scene in presented movies is appropriately changed according to the user's walk. The frames of the movie are

interpolated by using a blending technique between frames when a user walks slowly. The system components are described in some more detail below.

(a) Locomotion Interface

This interface is composed of a treadmill (WalkMaster), a couple of 3-D position sensors (Polhemus Fastrak) and PC A (Intel Pentium 4 2.4 GHz) for control as illustrated in Figure 3. User's locomotion is detected by two 3-D position sensors fixed on user's legs (Figure 3(1)). The treadmill is controlled by PC A based on position information from the sensors (Figure 3(2)). The belt of the treadmill is automatically rotated so that the center of gravity of two sensors coincides with the center of the belt area.[10] Although a user can walk toward any direction on this device, only the forward and backward direction is used for the present telepresence system. PC A calculates the displacement of user's position and sends it to the graphics PC cluster (Figure 3(3)).

(b) Graphics PC Cluster

The graphics PC cluster is composed of twelve PCs (CPU: Intel Pentium 4 1.8 GHz, Graphics Card: Geforce4 Ti4600) and a control PC B (Intel Pentium 4, 1.8 GHz). Each graphics PC is networked through 100Mbps LAN and is controlled by PC B. PC B sends frame indexes to twelve PCs using the UDP protocol simultaneously (Figure 3(4)). Each machine draws synchronized frame images according to the user's motion (Figure 3(5)). Note that the images are accumulated in local hard disk in advance and only the frame index is carried via network.

(c) Immersive Projection Display

The immersive projection display is composed of three slanted rear-projection screens (Solidray VisualValley) and twelve projectors. To obtain a wide field of view, the screens are located in user's front, left and right sides. Consequently, they cover 3/4 of whole circumference. To realize high-resolution image projection, each screen image is made by four projectors (horizontal 2 $\times$ vertical 2). The actual resolution of each projector is 1024$\times$768 (XGA) pixels. Because there are some overlapping areas projected by multiple projectors and some areas are not projected on the screen, the resolution of each screen is about 2 million pixels.

3. EXPERIMENT

In experiments, omnidirectional movies were obtained by Ladybug[9] put on a moving car. Figure 4 shows an example set of input images. Since the car speed was fixed approximately at 20km/h and the frame rate of Ladybug is 15fps, about 2.7 frames of the omnidirectional movies are captured in 1m interval. Twelve movies corresponding to the projectors are generated as shown in Figure 5. The resolution of each movie is set as 480$\times$360 so as to be higher than the actual resolution of Ladybug with assumption of the 15% overlap regions projected by two adjacent projectors. It has been confirmed that the geometric and photometric discontinuities among adjacent camera images could not be recognized except in some scene areas very close to the camera system, where the geometric discontinuities are perceived due to the violation of single view point constraint.

Next, the subjective evaluation was conducted using the generated movies above shown in Figure 6. The system can render the generated virtualized environment at 26 fps. Figure 7 shows three pictures obtained from the user's view position in three directions. We have confirmed that the geometric discontinuities between regions projected by different projectors and synchronization errors could not be recognized except for the borderlines between two screens. This result means that the virtualized environment movies are accurately generated according to the shape of the display surface. We have also confirmed that the proposed telepresence system provides us with the feeling of rich presence in remote sites in this experiment. However, poor presence

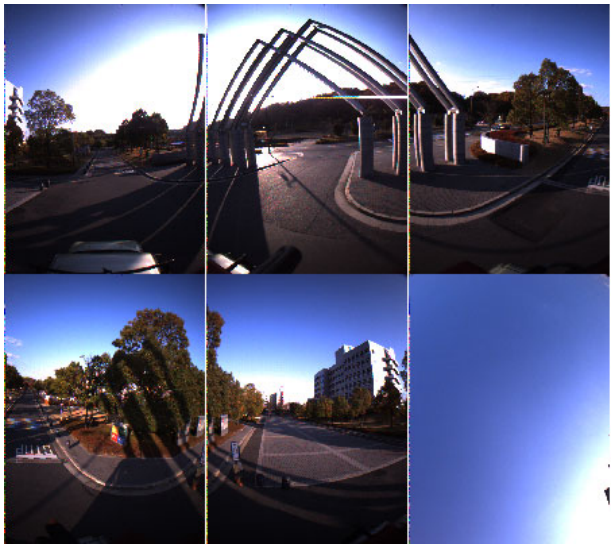

Figure 4. Sampled frames of image sequences acquired by six camera units of Ladybug.

Figure 5. Sampled frames of a composite movie generated by twelve graphics PCs.

was felt due to the limitation that the user's view position in a virtualized environment can not move in two dimensions. We also felt unnatural in the control of the treadmill when a user begins to walk, because the motion of upper part of the body is not considered in motion measurement; that is, the displayed image is not actually synchronized with head motion but with leg motion.

4. SUMMARY

In this paper, a novel telepresence system using an immersive projection display and a treadmill is proposed. This system can interactively present the feeling of walking in remote sites by showing a virtualized environment generated from real outdoor scene images. For construction of a virtualized environment, omnidirectional high-resolution movies are acquired by an omnidirectional multi-camera system calibrated geometrically and photometrically. The proposed system presents the calibrated movies to users according to their locomotion by using the treadmill and 3D position sensors.

We can confirm that the geometrical and photometrical calibration of the omnidirectional multi-camera system is successfully achieved. The experiment has shown that the proposed telepresence system provides us

Figure 6. Appearance of the system.

(a) left front (b) front (c) right rear

Figure 7. User's view.

with the feeling of rich presence in remote sites. In future work, we will relax the limitation in movement of user's view in virtualized environments, combining some methods such as camera pass estimation[11] and new view synthesis.[12,13]

REFERENCES

1. "Special issue on immersive telepresence," *IEEE Multimedia* **4**(1), 1997.
2. S. Chen, "Quicktime VR: An image-based approach to virtual environment navigation," *Proc. SIGGRAPH '95* , pp. 29–38, 1995.
3. Y. Onoe, K. Yamazawa, H. Takemura, and N. Yokoya, "Telepresence by real-time view-dependent image generation from omnidirectional video streams," *Computer Vision and Image Understanding* **71**(2), pp. 154–165, 1998.

4. D. Kotake, T. Endo, F. Pighin, A. Katayama, H. Tamura, and M. Hirose, "Cybercity walker 2001 : Walking through and looking around a realistic cyberspace reconstructed from the physical world," *Proc. 2nd IEEE and ACM Int. Symp. on Augmented Reality* , pp. 205–206, 2001.
5. U. Neumann, T. Pintaric, and A. Rizzo, "Immersive panoramic video," *Proc. 8th ACM Int. Conf. on Multimedia* **71**(2), pp. 493–494, 2000.
6. W. Tang, T. Wong, and P. Heng, "The immersive cockpit," *Proc. Int. Workshop on Immersive Telepresence* , pp. 36–39, 2002.
7. S. Ikeda, T. Sato, and N. Yokoya, "High-resolution panoramic movie generation from video streams acquired by an omnidirectional multi-camera system," *Proc. IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent System* , pp. 155–160, 2003.
8. M. Uyttendaele, A. Criminisi, S. B. Kang, S. Winder, R. Hartley, and R. Szeliski, "High-quality image-based interactive exploration of real-world environments," *Technical Report MSR-TR-2003-61 Microsoft Research* , 2003.
9. Point Grey Research Inc. http://www.ptgrey.com/.
10. H. Iwata, "Walking about virtual environments on an infinite floor," *Proc. IEEE Virtual Reality '99* , pp. 286–293, 1999.
11. T. Sato, M. Kanbara, N. Yokoya, and H. Takemura, "Dense 3D reconstruction of an outdoor scene by hundreds-baseline stereo using a hand-held video camera," *Int. Journal of Computer Vision* **47**(1-3), pp. 199–129, 2002.
12. M. Irani, T. Hassner, and P. Anandan, "What does the scene look like from a scene point?," *Proc. 7th European Conf. on Computer Vision* **2**, pp. 883–897, 2002.
13. A. Fitzgibbon, Y. Wexler, and A. Zisserman, "Image-based rendering using image-based priors," *Proc. 9th IEEE Int. Conf. on Computer Vision* **2**, pp. 1176–1183, 2003.