# An Immersive Telepresence System with a Locomotion Interface Using High-resolution Omnidirectional Movies

Sei IKEDA, Tomokazu SATO, Masayuki KANBARA and Naokazu YOKOYA
Nara Institute of Science and Technology
Vision and Media Computing Laboratory
8916-5 Takayama, Ikoma, Nara, Japan
{sei-i, tomoka-s, kanbara, yokoya}@is.aist-nara.ac.jp

## Abstract

*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 of a real outdoor scene. The proposed system provides users with rich sense of walking in a remote site.*

## 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 entertainment, medicine, 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 telepresence 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. 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 recent works [3, 4, 5, 6, 7] have improved the method above. 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. However, users can not control their view positions in virtualized environments. 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 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 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, generation of a virtualized environment and presentation to users. For the first process, we use an OMS: Ladybug [8] constructed of six cameras to reduce human cost in acquisition of images. In the second process, virtualized environment movies are generated from captured multiple image sequences. Geometric and photometric calibration of the OMS is required 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.

For the acquisition of images, movies are captured by an OMS Ladybug mounted on a moving car in outdoor, as shown in Figure 1. This camera system obtains six $768 \times 1024$ images at 15 fps; five for horizontal views and one for a vertical view. In this work, the speed of the car is kept constant for simplification of the control 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 an immersive screen of a telepresence system by using a method of [7]. Geometric and photometric calibration of an OMS is required to automatically generate movies of virtualized environments. The following paragraphs describe the calibration of an OMS and the generation of virtualized environment movies.

#### Calibration for OMS

In the geometric calibration, 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 all 3D positions of the markers on the board are calculated by linear interpolation among its corners. This method can virtually arrange markers around an OMS simultaneously. 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. The strength of limb darkening is calculated from estimated intrinsic parameters. The difference of color balances between camera units is measured by a histogram matching based on a linear relation between radiance and irradiance.

**Figure 1. An OMS mounted on a car.**

#### Generation of Virtualized Environment Movie

This step is based on re-projecting calibrated input images to a virtual image surface which corresponds to the shape of immersive screens for presentation. 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 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 2, 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 locomotion as input data, and sends calculated displacement information to the PC cluster. The PC cluster draws twelve images synchronized with the speed of user's walk because each screen image is generated by four projectors. As output data, these movies are displayed according to the user's motion. The scene in presented movies is appropriately changed according to the user's walking. The system components are described in some more detail below.

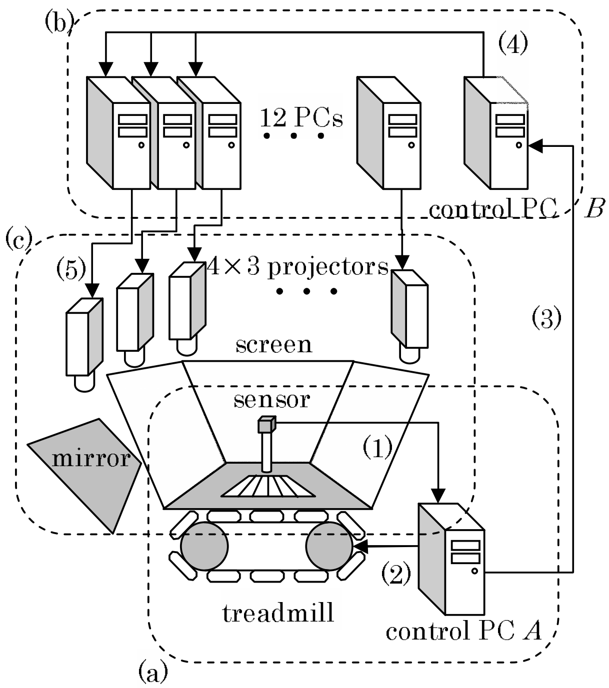


**Figure 2. Components of the proposed system.**

**Figure 3. A sampled frame of captured images acquired by six camera units of Ladybug.**

### 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 2. User's locomotion is detected by two 3-D position sensors fixed on user's legs (Figure 2(1)). The treadmill is controlled by PC A based on position information from the sensors (Figure 2(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 [9]. This virtually realizes an infinite floor. Although a user can walk in any direction on this device, only the forward and the backward direction are used for the present system. PC A calculates the displacement of user's position and sends it to the graphics PC cluster (Figure 2(3)).

### 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 2(4)). Each machine draws synchronized frame images according to the user's motion (Figure 2(5)). Simultaneously the frames are interpolated by a blending technique between frames when a user walks slowly. Note that the images are accumulated in local hard disk in advance and only the frame index is carried via network.

### Immersive Display

The immersive 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. To realize high-resolution image projection, each screen image is made by four projectors. The resolution of each projector is 1024×768 (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 potentially about 2 million pixels.

## 3. Experiment

In experiments, omnidirectional movies were obtained by Ladybug [8] put on a moving car. Figure 3 shows an example set of input images. Twelve movies corresponding to the projectors are generated as shown in Figure 4. The resolution of each movie is 480×360. It has been confirmed that the geometric and photometric discontinuities among adjacent camera images could not be recognized except in some close scene areas very close to the camera system due to its violation of single view point constraint.

Next, the subjective evaluation was conducted using the proposed telepresence system shown in Figure 5. The system can render the generated virtualized environment at 26

**Figure 4. A sampled frame of accumulated movies in twelve graphics PCs.**

**Figure 5. Appearance of the system.**

fps. We have confirmed that the proposed telepresence system provides us with the feeling of rich presence in remote sites in this experiment. We have also confirmed that the geometric discontinuities between regions projected by different projectors and synchronization errors could not be recognized. However, poor presence was felt due to the limitation that 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 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 a virtualized environment, wide-angle high-resolution movies are acquired by an omnidirectional multi-camera system calibrated geometrically and photometrically.

We can confirm that the geometrical and photometrical calibration of the OMS is successfully achieved. The experiment has shown that the proposed telepresence system provides us 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 [10] and new view synthesis [11].

## References


[1] "Special issue on immersive telepresence," *IEEE Multimedia*, vol. 4, no. 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*, vol. 71, no. 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*, vol. 71, no. 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] Point Grey Research Inc., http://www.ptgrey.com/.

[9] H. Iwata, "Walking about virtual environments on an infinite floor," *Proc. IEEE Virtual Reality '99*, pp. 286–293, 1999.

[10] 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*, vol. 47, no. 1-3, pp. 199–129, 2002.

[11] M. Irani, T. Hassner, and P. Anandan, "What does the scene look like from a scene point?" *Proc. 7th European Conf. on Computer Vision*, vol. 2, pp. 883–897, 2002.