MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan

15-4

Immersive Telepresence System with a Locomotion Interface Using High-resolution Omnidirectional Videos

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.naist.jp

Abstract

This paper describes a novel telepresence system which enables users to walk through a photorealistic virtualized environment by actual walking. In this system, a wide-angle high-resolution video is projected on an immersive multi-screen display and a treadmill is controlled according to user's locomotion. Calibration and motion estimation of a camera system are performed to enhance quality of presented images. 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. A telepresence system using real environment images is expected to be used in a number of fields such as entertainment, medicine and education. The telepresence system especially using an image-based technique attracts much attention because it can represent complex scenes such as outdoor environments. Our ultimate form of telepresence is an immersive system in which users can naturally move and look anywhere by their actions in a virtualized environment reproduced faithfully from a real environment. 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 [1] has developed an image-based telepresence system: QuickTime VR that generates a virtualized environment as panoramic images. In this system, users can see any directions and move their view positions in the environment through a standard display. Panoramic images are generated from multiple standard still images by using a mosaicing technique. The image acquisition task takes much time and effort to enable users to move their view positions in a wide area. Moreover, standard displays are not suitable for giving the feeling of virtually walking in remote sites.

In some recent works such as [2], omnidirectional video camera systems are used to acquire panoramic images at various positions and they have reduced human cost in acquisition of images. In a telepresence system developed by Onoe, et al. [2], multiple users can look around a scene of remote site in real time. They used an omnidirectional camera constructed of a standard single lens camera and a curved mirror. A part of omnidirectional video is displayed to users according to their view directions through a head mounted display.

Someother works [3, 4] have improved the resolution of omnidirectional videos using multiple cameras. Kotake, et al. [3] developed a telepresence system using a multi-camera system and an immersive three-screen display. They used a multi-camera system radially arranged on a moving car to acquire high-resolution panoramic videos of an outdoor scene. In the immersive display, users can see a wide field of view direction to get the feeling of high presence in remote sites. However, a game controller is used to move the view position. This system provides users with the sense of riding a carriage rather than the sense of walking in a virtualized environment of a real outdoor scene.

In this paper, we propose a novel telepresence system and a method to generate virtualized environments from real images. The system consists of a multi-projection display and a treadmill. It is designed to enable a user to move his view point freely by his actual walking in a photorealistic virtualized environment. For this system, the virtualized environment is generated as video streams by using computer vision techniques from real images acquired by an omnidirectional multi-camera system (OMS). In this method, shake effect and replay speed of video are corrected to improve the sense of walking. For these corrections, calibration and motion estimation of the OMS are performed in advance.

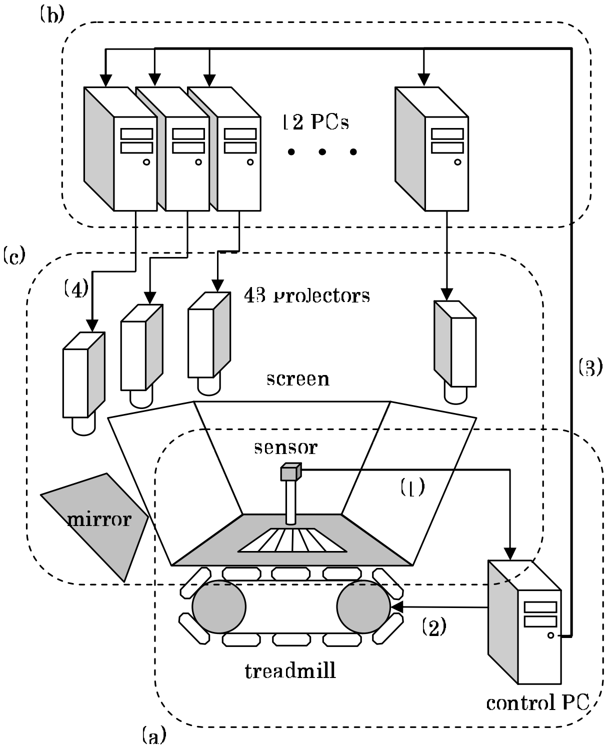

Figure 1. Components of the proposed system. (a) Locomotion interface, (b) graphics PC cluster, (c) immersive display.

2. Immersive Telepresence System with Locomotion Interface

This section describes a system to present a virtualized environment, as shown in Figure 1. 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 images are displayed according to the user's motion. The scene in presented images is appropriately changed according to the user's walking. The system components are described in some more detail below.

(a) Locomotion Interface

The locomotion interface is composed of a treadmill (WalkMaster), a couple of 3-D position sensors (Polhemus Fastrak) and a control PC (Intel Pentium 4 2.4 GHz) as illustrated in Figure 1. User's locomotion is detected by two 3-D position sensors fixed on user's legs (Figure 1(1)). The treadmill is controlled by the control PC based on position information from the sensors (Figure 1(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 [5]. 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. The control PC calculates the displacement of user's position and sends it to the graphics PC cluster (Figure 1(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). Each graphics PC is networked through 100Mbps LAN and is controlled by the control PC. The control PC sends frame indexes to twelve PCs using the UDP protocol simultaneously (Figure 1(3)). Each PC draws synchronized frame images according to the user's motion (Figure 1(4)). Note that the images are accumulated in local hard disk in advance and only the frame index is carried via network.

(c) 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. Generation of Virtualized Environments

This section describes a method to generate a virtualized environment images according to the shape of the immersive screen described in the previous section. This method is based on re-projecting calibrated omnidirectional images to a virtualized image surface which corresponds to the shape of immersive screens. In this image generation, shake effect of the acquired videos and variation of camera speed are reduced by using extrinsic camera parameters of OMS. The following paragraphs describe acquisition of images, the estimation of extrinsic camera parameters of OMS and the generation of virtualized environment videos.

Acquisition of Images

An OMS has an important advantage that human cost in acquisition of images can be reduced because OMS can capture wide field of view with high resolution. For the acquisition of images, an OMS : Ladybug [6] is employed and mounted on a moving electric wheel chair, as shown in Figure 2. This camera system obtains six 768x1024 images at 15 fps; five for horizontal views and one for a vertical

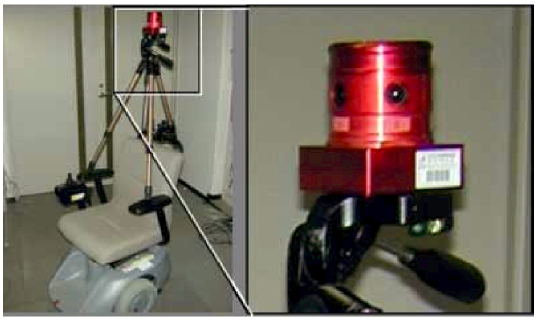

Figure 2. OMS mounted on an electric wheelchair.

view. In this work, the height of an OMS's view point is fixed to be almost same height with eyes of a standing human and the speed of the wheel chair is kept as constant as possible for simplification of the process of presentation to users because presentation of virtualized environments should be controlled according to user's locomotion regardless of variation of the car speed in the presentation process. However, it is difficult to control the position, orientation and speed of an OMS precisely. So, some of corrections are necessary.

Estimation of Extrinsic Camera Parameters

In this step, position and direction of an OMS are estimated from input multiple videos. We assume that relative positions and directions of camera units in the OMS are fixed. For the first step, these relative parameters and intrinsic camera parameters of each camera unit are measured by using our calibration method [4]. Limb darkening and color balance of camera units are also corrected in advance. After the calibration of intrinsic parameters of the OMS, extrinsic camera parameters of OMS are estimated by tracking both feature landmarks and image features in input video streams automatically [7]. Feature landmarks mean image features whose 3-D positions are known. In this method, the position and orientation of all the cameras are calculated and optimized simultaneously, so that projection errors of the feature landmarks and features points in images of all the cameras are minimized. This method makes it possible to estimate camera motion more precisely than conventional methods using a single camera system since the motion of OMS is constrained by feature landmarks and natural features existing in all the directions. In this estimation, a small number of feature landmarks need to be visible and to be specified manually in key frames for minimizing accumulative errors.

Video Generation Using Camera Parameters

This stage is based on re-projecting input images to a virtual image surface which corresponds to the shape of immersive screens for presentation. Input images are then projected on a projection surface by using the estimated extrinsic camera parameters of an OMS. This section first describes reduction of shake effect of acquired videos. The correction of replay speed is then described. Finally, the effect related to the violation of single view point constraint of the OMS is described.

Sake effect reduction: Shaking effect of video is caused by rotation and translation of a camera system. In this step, the rotation factor is canceled to reduce this effect by using the estimated extrinsic camera parameters of an OMS.

Replay speed correction: Frame indices of the input video are re-assigned so that the relation between user's locomotion and the displayed frame index become linear in the process of presentation to users. This correction can be easily performed by using translation parts of the estimated extrinsic camera parameters.

Projection of input camera Images: The single viewpoint perspective projection model is not applicable for this camera system since the centers of projection of six camera units of the OMS are different from each other. However, when the distance of a target from the system is sufficiently long, 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 video is generated by projecting all the pixels of all the six input images onto the projection surface. Although there is no exact definition of resolution for the above reason, the total horizontal resolution of an omnidirectional video acquired by Ladybug can be approximated as about 3340 pixels, assuming a horizontal 13% overlapping region between two adjacent cameras.

4. Experiment

We have acquired real images by using Ladybug in a outdoor environment. They are shown in Figure 3. Twelve movies corresponding to the projectors are generated from the acquired images 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 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 shake effect of the acquired video and variation of camera speed are successfully reduced. However, poor presence

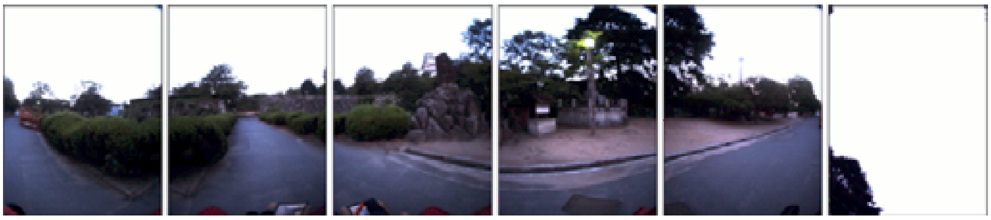

Figure 3. Sampled frame of captured videos acquired by six camera units of Ladybug.

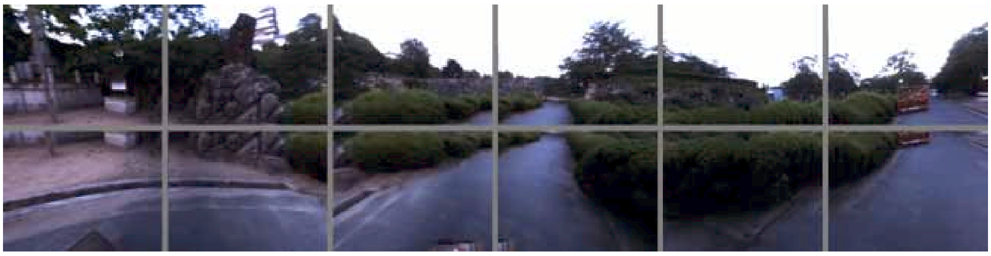

Figure 4. Sampled frame of generated video accumulated in twelve graphics PCs.

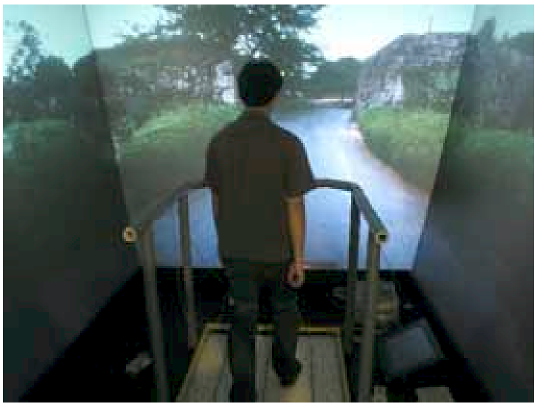

Figure 5. Appearance of the system.

was felt due to the limitation of user's view position in a virtualized environment : the user can not move in two dimensions in the presenta system. 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.

5. 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 videos are acquired by an omnidirectional multi-camera system. 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 using methods such as new view synthesis.

References

[1] S. Chen, "Quicktime VR: An image-based approach to virtual environment navigation," *Proc. SIGGRAPH '95*, pp. 29–38, 1995.

[2] 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.

[3] 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.

[4] 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.

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

[6] Point Grey Research Inc., http://www.ptgrey.com/.

[7] T. Sato, S. Ikeda, and N. Yokoya, "Extrinsic camera parameter recovery from multiple image sequences captured by an omni-directional multi-camera system," *Proc. 8th European Conf. on Computer Vision*, vol. 2, pp. 326–340, 2004.