Construction and Presentation of a Virtual Environment Using Panoramic Stereo Images of a Real Scene and Computer Graphics Models

Jun Shimamura, Haruo Takemura, Naokazu Yokoya and Kazumasa Yamazawa
Graduate School of Information Science,
Nara Institute of Science and Technology
8916-5, Takayama, Ikoma, Nara 630-0101, Japan
{jun-s, takemura, yokoya, yamazawa}@is.aist-nara.ac.jp

Abstract

The recent progress in computer graphics has made it possible to construct various virtual environments such as urban or natural scenes. This paper proposes a hybrid method to construct a realistic virtual environment containing an existing real scene. The proposed method combines two different types of 3-D models. 3-D geometric model is used to represent virtual objects in user's vicinity, enabling a user to handle virtual objects. Texture mapped cylindrical 2.5-D model of a real scene is used to render the background of the environment, maintaining real-time rendering and increasing realistic sensation. Cylindrical 2.5-D model is generated from cylindrical stereo images captured by an omnidirectional stereo imaging sensor. A prototype system has been developed to confirm the feasibility of the method, in which panoramic binocular stereo images are projected on a cylindrical immersive projection display depending on user's viewpoint in real-time.

1. Introduction

The recent progress in computer graphics has made it possible to construct various virtual environments such as urban or natural scenes. Moreover, the mixed reality technology which merges the real and virtual worlds seamlessly has recently become popular [6, 7].

The construction methods for virtual environments, such as large scale urban or natural scenes, are generally classified into two categories: polygon based and image based. The polygon based method which constructs an environment using polygonal representation of objects has the advantage of easily realizing interaction between a user and virtual objects, whereas the image synthesis time is dependent on the complexity of the constructed scene. This problem is particularly critical in simulation and virtual reality applications because of the demand for real-time feedback. A number of approaches have been proposed to overcome this problem [8], however the problem still exists, since scene complexity is potentially unbounded. On the other hand, the image based method constructs the environment using multiple real images [2, 5] or deforming real images [1] and requires less rendering time because the number of polygons is independent of the scene complexity. In addition, real images can present highly realistic sensations to a user with ease. However, such an approach has several drawbacks. Firstly, the implementation that enables a user to arbitrarily change his/her viewpoint and viewing orientation in a large environment is essentially impossible because of memory limitation. Secondly, the representation of depth and occlusion is difficult when the texture images are mapped to a planar or cylindrical plane. Finally, the realization of interaction between a user and virtual objects is difficult.

This paper proposes a hybrid method to construct a realistic virtual environment containing an existing real scene and computer graphics (CG) models. Our approach is based on acquiring full panoramic 2.5-D models of dynamic real worlds using a video-rate omnidirectional stereo imaging sensor [4]. CG objects are merged into the full panoramic 2.5-D model of real scene maintaining consistent occlusion between real and virtual objects. Thus it is possible to yield rich 3-D sensation with binocular and motion parallax. Moreover, CG objects can be manipulated in the virtual environment. We have developed a prototype of immersive mixed reality system using a large cylindrical screen, in which a user can walk through the virtual environment and can manipulate CG objects in real time.

2. Construction of full panoramic 2.5-D models of dynamic real scenes

To increase realistic sensation for a user, real images often would be used in virtual environment. Traditionally, the real images are mapped onto a planar or cylindrical plane.

However, the drawbacks of the method are that operationality and reality are decreasing because the user can not sense binocular nor motion parallax in constructed environment. In this section, we describe a novel method to construct a full panoramic 2.5-D models of dynamic real scenes that provides realistic sensation in mixed reality applications.

Figure 1 shows a flowchart for the construction of full panoramic 2.5-D models. First, a pair of panoramic stereo images are captured by an omnidirectional stereo imaging sensor at video-rate (Figure 1 A). Next, depth map of the real world are estimated from these stereo images (Figure 1 B). Finally, the full panoramic 2.5-D models are constructed from these depth images and then the texture images are mapped on the model (Figure 1 C).

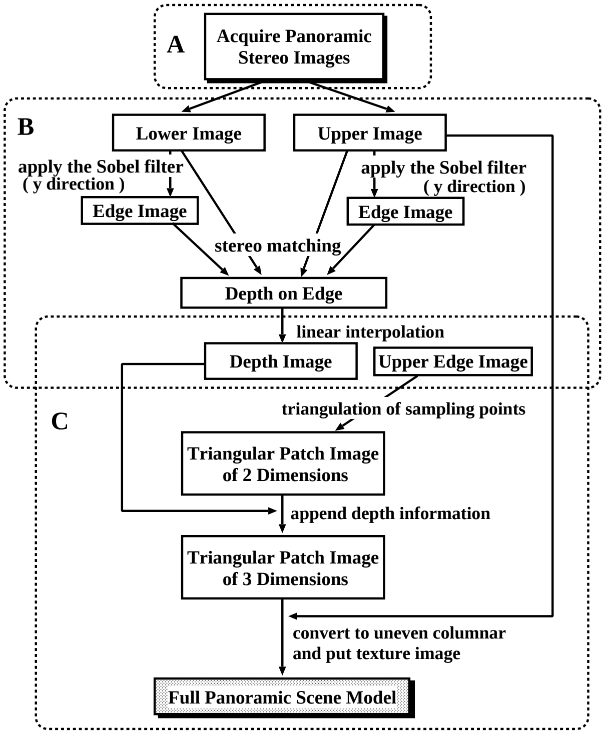
A
Acquire Panoramic Stereo Images
B
Lower Image
Upper Image
apply the Sobel filter (y direction)
apply the Sobel filter (y direction)
Edge Image
Edge Image
stereo matching
Depth on Edge
linear interpolation
Depth Image
Upper Edge Image
C
triangulation of sampling points
Triangular Patch Image of 2 Dimensions
append depth information
Triangular Patch Image of 3 Dimensions
convert to uneven columnar and put texture image
Full Panoramic Scene Model

Figure 1. Flow diagram of constructing a full panoramic 2.5-D model.

In the following sections, details of the imaging sensor, depth estimation and cylindrical model construction are described.

2.1. Capturing panoramic stereo images of a real scene

Omnidirectional stereo imaging

We use an omnidirectional stereo imaging sensor [4] that is composed of twelve cameras and two hexagonal pyramidal mirrors. The sensor component is designed so that the virtual lens centers of six cameras are located at a fixed point as shown in Figure 2 (a). It can take an omnidirectional image satisfying the single viewpoint constraint. In the system, two symmetrical sets of the component are used for omnidirectional stereo imaging. Each camera is a standard NTSC CCD camera with a wide-angle lens. The figure of the mirror is an equilateral hexagonal pyramid. The top of the mirror faces the six cameras and the base plane of the mirror is placed to be perpendicular to the line of sight of the cameras. The sensor, properly arranged, captures images of a real scene through the reflection on the pyramidal mirror. Figure 2 (b) shows an appearance of the sensor.

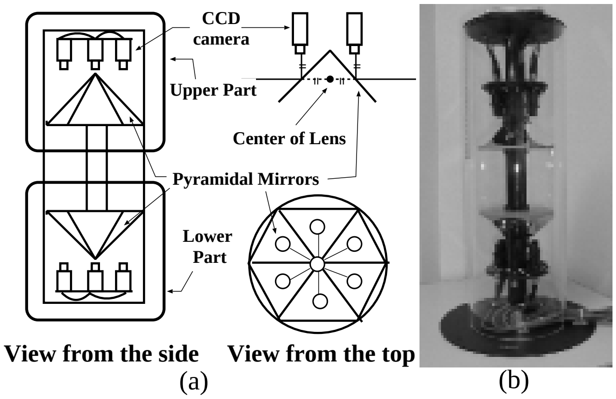

Figure 2. Geometry (a) and exterior (b) of omnidirectional stereo imaging sensor.

The omnidirectional stereo image sensor produces synchronized twelve video streams. The image captured by each camera is distorted because of using a wide-angle lens. The Tsai's calibration method [9] is applied to eliminate the distortion. Then, by projecting each image generated from upper and lower cameras onto a cylindrical surface, twelve sheets of images are combined into a pair of panoramic stereo images; upper and lower panoramic images. Consequently, a pair of panoramic stereo images which satisfy the vertical epipolar constraint is generated [4]. It should be noted that the stereo images here have vertical disparities, in contrast with familiar binocular stereo images with horizontal disparities.

Image analysis for representation of dynamic scene

In our approach, representation of dynamic real scene is realized by extracting images of static scene and images of dynamic event (moving object) from panoramic stereo images. The static scene image and moving object regions are extracted from panoramic images using existing techniques as follows.

1. Static scene image generation:

   A panoramic image of a static scene is generated by applying a temporal mode filter to a panoramic image sequence in a time interval. A stereo pair of panoramic static scene is obtained by applying this filter to both upper and lower images of omnidirectional stereo images.

2. Moving object extraction:

   Moving objects are extracted by subtracting consecutive image frames in time sequence.

2.2. Depth estimation from panoramic stereo images

In this section, the depth estimation from panoramic stereo images is described. The depth of an existing real scene is the principal factor in representing depth relationship between virtual and real objects correctly. We acquire panoramic depth based on stereo matching. However, there is high possibility of false matching caused by noises in performing stereo matching on the whole image. In order to acquire accurate depth map, some stereo techniques have been shown at a field of computer vision researches. But, we note that the precise depth information is not necessarily required in VR system. Thus, we estimate depth values only on edges, where the matching is thought to be reliable. Thereafter, intermediate data are approximated by linear interpolation. The following steps describe the method in more detail.

1. By adopting the Sobel filter, non-vertical edges are detected in the upper and lower images.
2. Stereo matching is performed and the depth values are computed. Note that only pixels on detected edges in the upper image are matched to those in the lower image, matching window size is 9×9 pixels, and similarity measure is the normalized cross-correlation having a high threshold. In the same way, the lower image as a reference image is matched to the upper image.
3. Matching errors are excluded by considering the consistency between upper-to-lower and lower-to-upper matchings.
4. To revise noise and lacking values at the upper edges, adopting the median filter (5×3). upper edges.
5. The depth values at the pixels between the edges are linearly interpolated to complete a dense depth map.

2.3. Generation of cylindrical model

By using a panoramic depth map estimated in Section 2.2, a cylindrical 2.5-D model for presentation of dynamic real scene is constructed using the following steps.

1. Edges are detected from the upper image and points on edges with reliable depth are sampled at a fixed interval (3 pixels). Then non-edge points are sampled at a fixed interval (31 pixels) over an entire region.
2. By applying the Delaunay's triangulation [3] to points extracted in Step.1, 2-D triangle patches are generated.
3. A 2.5-D triangular patch model is generated by assigning 3-D data of the depth image created in Section 2.2 to the vertices of 2-D triangles obtained in Step 2.
4. In order to compensate color balance of texture images captured by adjacent cameras of the sensor, a weight by which original intensity is multiplied is computed from averages of intensity at a connecting row. Then, the weight at the whole image is linearly interpolated. Finally, the texture is mapped onto the constructed 2.5-D cylindrical model.

3. Presentation of a mixed environment

In order to confirm the feasibility of the proposed method, we developed a prototype system for presenting a mixed environment of a real scene and CG objects. The cylindrical 2.5-D model is constructed on a graphics workstation, SGI Onyx2 (Infinite Reality2×2, 8CPUs MIPS R10000, 250MHz). Virtual objects are created by using a computer graphics software(Alias/WaveFront), and easily merged into a real scene model maintaining correct occlusion among real and virtual objects because the real scene model has depth information. For presenting the mixed environment to a user in our system, 3-D images are projected on a large cylindrical screen with the size of 6m in diameter and 2.4m in height of the CYLINDRA[1]) system, which has a 330-degree view covered by six projectors. Note that the projected images are a pair of panoramic stereo images with horizontal disparities. In the system, a user is able to change viewing position and orientation by a joystick device (SideWinder Precision Pro/Microsoft Inc.) and is able to experience stereo-scopic vision as well as motion parallax through liquid crystal shutter-glasses (SB300/Solidray Inc.). The hardware configuration of the system is illustrated in Figure 3.

Here we demonstrate an application to sight simulation. Figure 4 shows a pair of panoramic stereo images (3006×330 pixels) of “Heijo-kyo” (historical site in Nara) which contains the reconstructed “Suzaku-mon” gate captured by the omnidirectional stereo imaging sensor. In the experiment, the base-line of omnidirectional stereo imaging sensor is set to 25.0mm, and each focal length of CCD camera is set to 4.0mm. Figure 5 shows panoramic stereo images of a static scene generated from a sequence of dynamic stereo images including Figure 4. It can be clearly observed in Figure 5 that moving objects are eliminated from Figure 4 (moving object regions in Figure 4 are highlighted as white boxes in Figure 5). Depth map computed from panoramic stereo images in Figure 5 is shown in Figure 6 in which depth values are coded in intensities. A brighter

[1] CYLINDRA is an abbreviation for Cylindrical Yard with Large, Immersive and Novel Display for Reality Applications.

pixel is closer and a darker pixel is farther. A black pixel is a pixel of its depth is not computed from stereo images. The resolution of depth is about 1.0 meter at where 10.0 meter. It seems to be wrong precise, but it has enough information when representing depth relation among real and virtual objects in VR system. Computation time of generating the depth image is about 40 minutes with 8 CPUs. Figure 7 illustrates a bird's-eye view of texture-mapped full panoramic 2.5-D model constructed by applying the algorithm described in Section 2.3 to the depth image. Distance to a pixel of which disparity is not computed are set to infinity. The whole cylindrical model in Figure 7 consists of 13400 polygons. Figure 8 shows a mixed environment consisting of a static real scene and 3-D virtual objects (Four trees: 41340 polygons). From three different viewpoint images, it is clearly seen that motion parallax is presented in the system. Figure 9 shows examples of superimposing dynamic event layers onto the static scene, in which images of two time instances are rendered assuming two different viewpoints. A walking person is rendered as a dynamic event in this scene. Figure 10 shows a user performing the walk-through in the mixed environment using the CYLINDRA system. In the present system with this example, image updating rate is about 13 frames/sec, when 2 CPUs are used to compute stereo-scopic images of 6144×768 pixels.

It has been found that a user can feel real-time feedback and deep realistic sensation in the mixed environment constructed by the proposed method. In addition, we have confirmed that a user can handle virtual objects by using depth relationships among objects, and can sense binocular and motion parallax in the panoramic environment.

On the other hand, a user of the system have to stay relatively close to the center of cylindrical 2.5-D model in order to have better realistic sensation. When a user moves far away from the center, occluded area which is not observed from the panoramic image sensor appears in a scene and causes a sense of incompatibility. To solve this problem, a method of preparing multiple cylindrical 2.5-D models with different center position and switching among models based on a user's position will be effective.

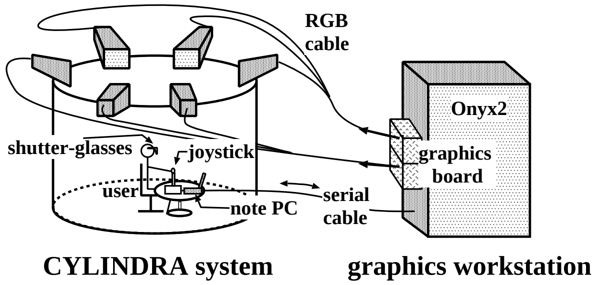

Figure 3. Hardware configuration of prototype system.

4. Conclusion and future directions

In this paper, we have proposed a novel method of constructing a large scale virtual environment. The constructed environment consists of cylindrical 2.5-D model of a real scene and polygonal CG objects. In order to represent approximate depth relationship among objects, depth values are appended to the cylindrical panoramic image. Consequently, the proposed method maintains real-time rendering because of constructing an approximate scene using real images and increases realistic sensation. Applying the proposed method, a user can virtually walk to arbitrary directions in real-time in the mixed world of real and virtual objects as same as in the real world, and can handle virtual objects smoothly by using depth relationships among objects. In the environment, a user can also feel deep realistic sensation of a mixed world.

As the future work, we will extend the model far larger by using multiple panoramic stereo images. We will also implement an algorithm that smoothly switch constructed models when user's viewpoint changes.

Acknowledgments

This work was supported in part by the Telecommunications Advancement Organization of Japan.

References

[1] G. U. Carraro, T. Edmark, and J. R. Ensor. Techniques for handling video in virtual environment. *Proc. SIGGRAPH98*, pages 353–360, 1998.

[2] S. E. Chen. QuickTime VR – An image-based approach to virtual environment navigation. *Proc. SIGGRAPH95*, pages 29–38, 1995.

[3] P. Heckbert, Eds. *Graphics Gems IV*. Academic Press Professional, Boston, 1994.

[4] T. Kawanishi, K. Yamazawa, H. Iwasa, T. Takemura, and N. Yokoya. Generation of high-resolution stereo panoramic images by omnidirectional imaging sensor using hexagonal pyramidal mirrors. *Proc. 14th IAPR Int. Conf. on Pattern Recognition (14ICPR)*, I:485–489, August 1998.

[5] A. Lippman. Movie-Maps: An application of the optical videodisc to computer graphics. *Computer Graphics*, 14(3):32–42, 1980.

[6] P. Milgram and F. Kishino. A taxonomy of mixed reality visual display. *IEICE Trans. on Information and Systems*, E77-D(12):1321–1329, December 1994.

[7] Y. Ohta and H. Tamura, Eds. *Mixed Reality –Merging Real and Virtual Worlds*. Ohmsha & Springer-Verlag, Tokyo, 1999.

[8] F. Sillion, G. Drettakis, and B. Bodelet. Efficient impostor manipulation for real-time visualization of urban scenery. *Computer Graphics Forum*, 16(3):207–218, 1997.

[9] R. Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, RA-3(4):323–344, August 1987.

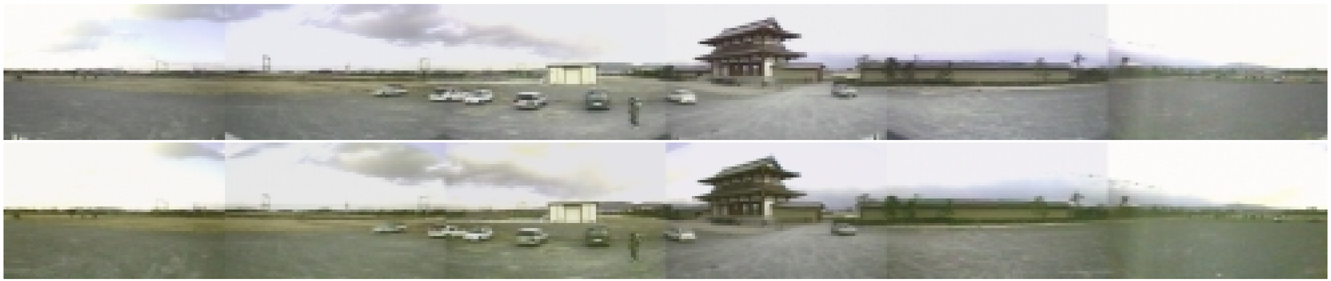

Figure 4. A pair of computed panoramic stereo images.

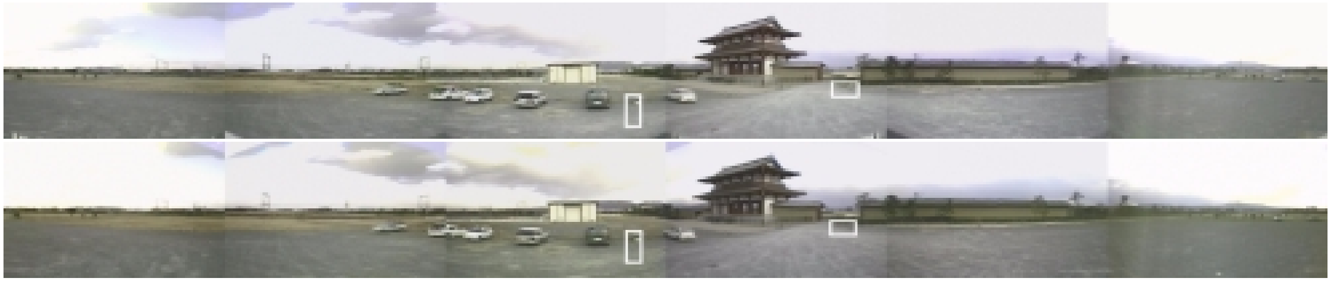

Figure 5. A pair of panoramic stereo images of a static scene without moving objects.

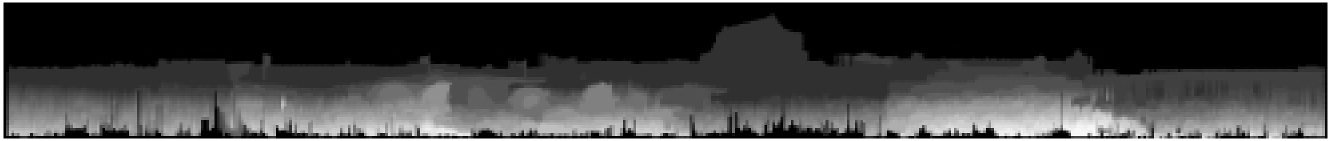

Figure 6. Panoramic depth map generated from stereo images.

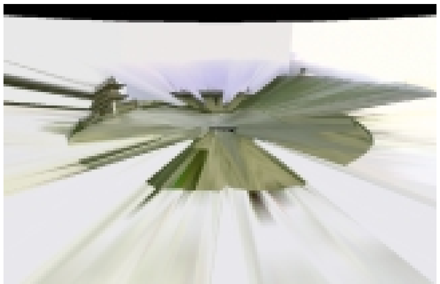

Figure 7. Bird's-eye view of full panoramic 2.5-D scene model.

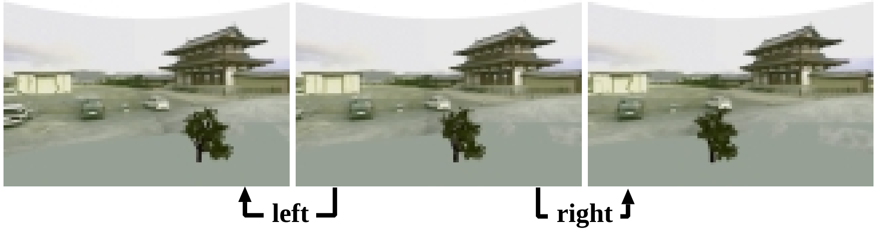

Figure 8. Mixed environment observed from different viewpoints(center: original viewpoint of sensor).

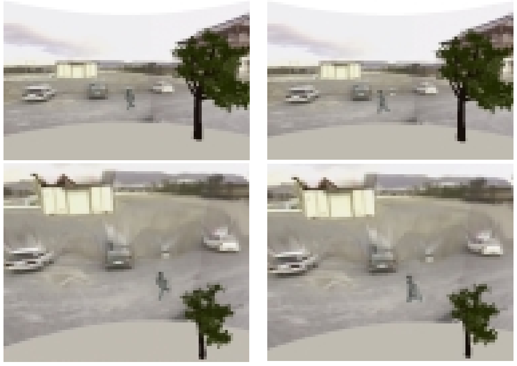

Figure 9. Superimposing dynamic event layers onto a static scene layer with virtual objects (top: original viewpoint; bottom: new higher viewpoint).

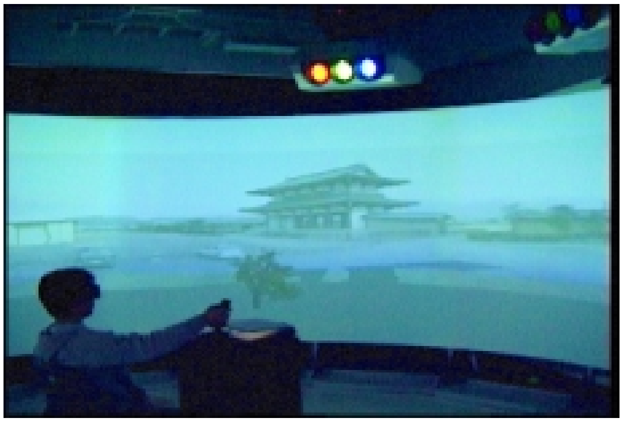

Figure 10. User's appearance in mixed environment using CYLINDRA system.