Novel Stereoscopic View Generation by Image-Based Rendering Coordinated with Depth Information

Maiya Hori, Masayuki Kanbara, and Naokazu Yokoya

Graduate School of Information Science, Nara Institute of Science and Technology,
8916–5 Takayama, Ikoma, Nara 630–0192, Japan

Abstract. This paper describes a method of stereoscopic view generation by image-based rendering in wide outdoor environments. The stereoscopic view can be generated from an omnidirectional image sequence by a light field rendering approach which generates a novel view image from a set of images. The conventional methods of novel view generation have a problem such that the generated image is distorted because the image is composed of parts of several omnidirectional images captured at different points. To overcome this problem, we have to consider the distances between the novel viewpoint and observed real objects in the rendering process. In the proposed method, in order to reduce the image distortion, stereoscopic images are generated considering depth values estimated by dynamic programming (DP) matching using the images that are observed from different points and contain the same ray information in the real world. In experiments, stereoscopic images in wide outdoor environments are generated and displayed.

1 Introduction

A technology that enables users to virtually experience a remote site is called telepresence[1]. The telepresence system has to provide rich visual sensation so that user can feel like existing at the remote site. In general, methods to provide a user with rich visual sensation are divided into two approaches: A model-based rendering (MBR) approach[2,3,4] and an image-based rendering (IBR) approach[5,6,7,8]. In the MBR approach, since virtual scene images are generated from 3D model with the 3D shape and its reflectance properties of an object, it is difficult to automatically reconstruct large-scale virtual scene such as an outdoor environment for telepresence. On the other hand, the IBR approach can render a scene consisting of complicated shapes and reflectance properties because synthesized images are generated from captured images. Therefore, the IBR approach is often used for telepresence of an outdoor environment[9]. In the IBR approach, Ikeda et al.[10] have proposed an immersive telepresence system using high-resolution omni-directional videos. This system can show user a high-resolution virtual image, however the user's viewpoint is restricted on a camera path and user can see only monocular images.

B.K. Ersbøll and K.S. Pedersen (Eds.): SCIA 2007, LNCS 4522, pp. 193–202, 2007.

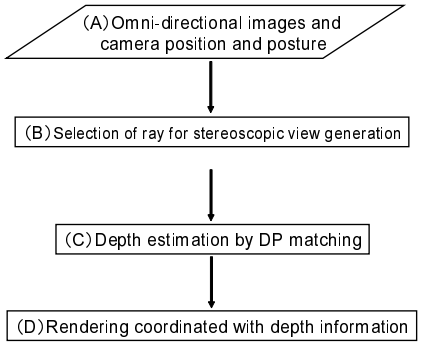


Fig. 1. Flow diagram of the proposed method

In this paper, we propose a method for generating stereoscopic images from omni-directional image sequences for telepresence in outdoor environments. In order to generate stereoscopic images, it is necessary to generate a novel view image from a set of images captured on the camera path. The conventional methods of stereoscopic view generation suffer from the distortion due to vertical parallax[11,12]. Our method employs the light field rendering[13] same as the conventional method, and tries to reduce the distortion of generated image by rendering coordinated with depth information. The depth value is estimated by dynamic programming (DP) matching[14,15,16] with sum of squared distances (SSD) between two images that are observed at different position and are captured a same ray in the world. We can generate stereoscopic images from omni-directional images that are captured along a free camera path in wide outdoor environments by using a vehicle equipped with a high accuracy position and posture sensor.

This paper is structured as follows; Section 2 explains a method for generating the stereoscopic view in detail. In Section 3, we demonstrate experimental results of the stereoscopic view generation in outdoor environments. Finally, Section 4 describes conclusion and future work.

2 Stereoscopic View Generation

This section describes a method of stereoscopic view generation from omni-directional images sequences in outdoor environments. Figure 1 shows a flow diagram of the proposed method. First, a pair of omni-directional image sequences and extrinsic camera parameters including position and posture of camera are acquired in outdoor environments(A). Next, positions of binocular viewpoint are determined in a process of stereoscopic view generation(B). Parts of images needed for stereoscopic image are collected from stored omni-directional images based on the relationship between the rays from the novel viewpoint and the captured omni-directional images. To reduce the distortion due to vertical parallax, the depth from novel viewpoint is estimated by DP matching(C). Finally,

Fig. 2. Omni-directional multi-camera systems mounted on a vehicle

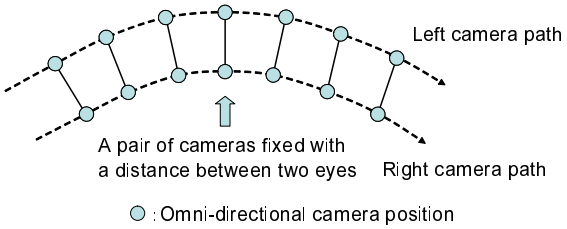


Fig. 3. Omni-directional camera position (top view)

stereoscopic images are rendered in the image plane coordinated with depth information(D).

2.1 Omni-Directional Images and Camera Positions

In this study, it is necessary to acquire a lot of rays required for stereoscopic view generation. We capture omni-directional image sequences in using a pair of omni-directional cameras fixed with a distance between two eyes as shown in Figure 2. Omni-directional cameras and their paths are illustrated in Figure 3. The images obtained from the fixed two omni-directional cameras enable us to generate the stereoscopic image when the view direction is parallel to the camera path. In addition, the configuration of cameras makes it possible to capture the images necessary for depth estimation mentioned in Section 2.3. The position and posture of each camera should be acquired at the same time with the high accuracy.

2.2 Selection of Ray for Stereoscopic View Generation

In this study, stereoscopic images at novel viewpoint are generated from pre-captured omni-directional images with a light field rendering approach. Omni-directional images exist at discrete points on the camera path as shown in Figure 4. In Figure 4, a perspective image can be generated from four omni-directional images captured from T_2 to T_5 and is vertically divided into four areas. As illustrated in Figure 5, we generate binocular stereoscopic images so that the center of left and right eyes is located on a circle whose diameter is the distance between two eyes. When the view direction is parallel to the camera

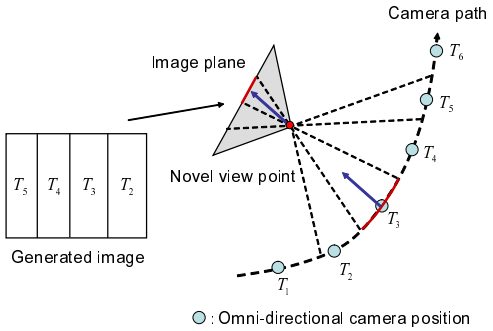


Fig. 4. Novel view image generated from pre-captured omni-directional images (top view)

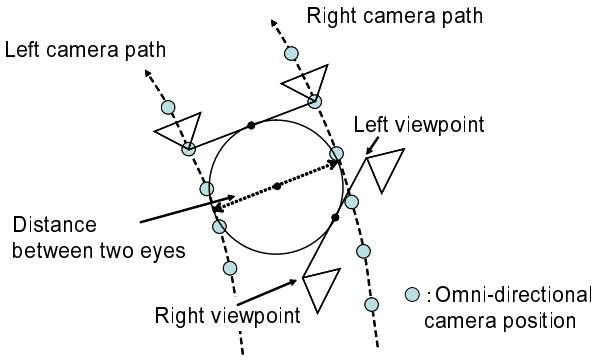


Fig. 5. Alignment of novel binocular viewpoints (top view)

path, each of stereo pair is made from a part of only one omni-directional image. When there is not ray to generate stereoscopic images because omni-directional images are captured at discrete positions, the omni-directional image of the nearest position is used. Some omni-directional images may contain the same ray to generate stereoscopic images. In this case, the omni-directional image captured at the near position from novel viewpoint is selected. In addition, mutual occlusions of camera bodies occur in the omni-directional images for generation of stereoscopic images because the omni-directional images are captured by using adjacent two omni-directional cameras. In this case, the omni-directional image captured by the other camera is used.

2.3 Depth Estimation by DP Matching

When a novel view is rendered considering without depth information, image distortion occurs in the boundary between subimages selected from different omni-directional images. The distortion appears in the generated image when the distances from an object to the omni-directional cameras differ from each other,

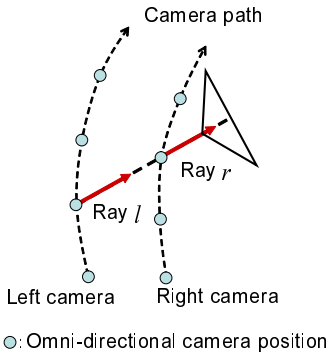


Fig. 6. Omni-directional images containing the same ray (top view)

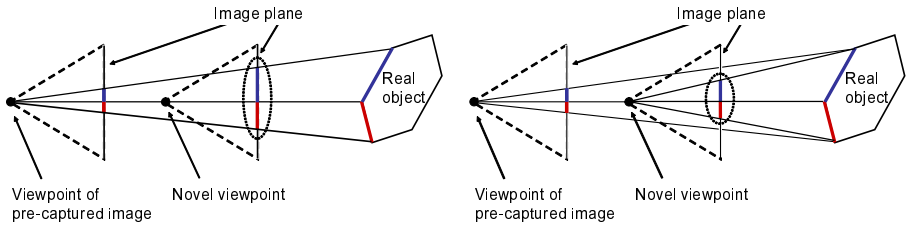
because the cameras capture the object from different position. We reduce the distortion in generated image by rendering coordinated with depth information. The depth value is acquired by DP matching[14,15,16] between edges in two images that are captured at different points and containing the same ray as shown in Figure 6. Similarity measure of DP matching is defined as follows:

$$g(L_i, R_j) = min \begin{cases} g(L_{i-1}, R_j) + d(L_i, R_j) \\ g(L_{i-1}, R_{j-1}) + d(L_i, R_j) \\ g(L_i, R_{j-1}) + d(L_i, R_j) \end{cases}, \tag{1}$$

where $L_i (i = 1 \sim I)$ represents i-th edge in the image captured in the left camera and $R_j (j = 1 \sim J)$ does j-th edge in the image captured in the right camera. $d(L_i, R_j)$ denotes the distance between feature vectors of two edges. By calculating a path to minimize $d(L_i, R_j)$, both edges are matched. Here, the distance between feature vectors of two edges is defined by SSD. When the plural edges can be corresponded with one edge, the edge which has a smallest value of SSD in plural edges assumes the corresponding edge. In this study the window size of SSD is 25×25 pixels.

2.4 Rendering Coordinated with Depth Information

The depth values are obtained only on edges by the method above. Dense depth map is computed by linear interpolation using depth values on edges. A distortion of generated image due to a vertical parallax by rendering a novel view image with the depth value. Figure 7 illustrates rendering process of the conventional and the proposal methods. In conventional method as shown in Figure 7(a), since a real object is rendered without a depth value, the size of real object can not be correctly represented in the image. On the other hand, in the proposed method (Figure 7(b)), the real object can be correctly rendered because the real object is projected onto an image plane with a perspective projection whose center is a viewpoint of the novel image. Therefore, the proposed method can reduce the distortion of novel view image.

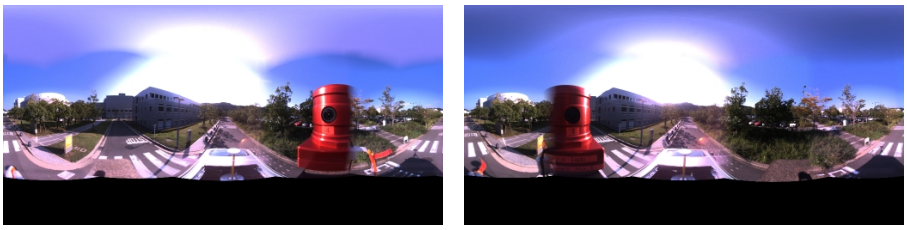


(a)conventional method (b)proposed method

Fig. 7. Rendering process (side view)

3 Experiments

To verify the validity of the proposed method, we have actually generated novel stereoscopic views of outdoor environments. In experiments, omni-directional movies and extrinsic camera parameters including position and posture are acquired by vehicle-mounted two omni-directional multi-camera systems and a position and posture sensor. As an omni-directional multi-camera system, we use Ladybug2 (Point Grey Research). The camera unit consists of six cameras: Five radically configured on horizontal ring and one pointing vertically. The camera system can collect synchronized movies at 30fps covering more than 75%



(a)Left camera (b)Right camera

Fig. 8. Omni-directional images

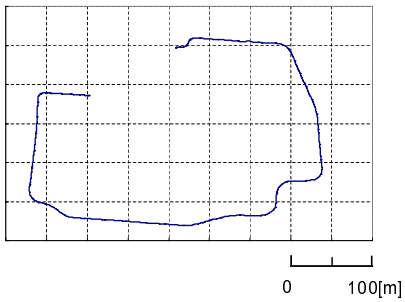
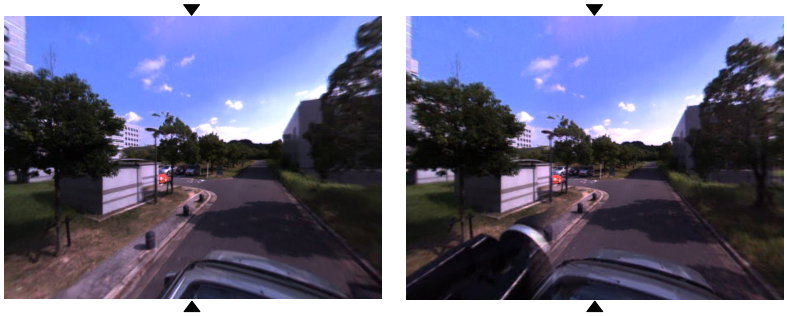


Fig. 9. Camera path



(a)Left camera (b)Right camera

Fig. 10. Images for depth estimation

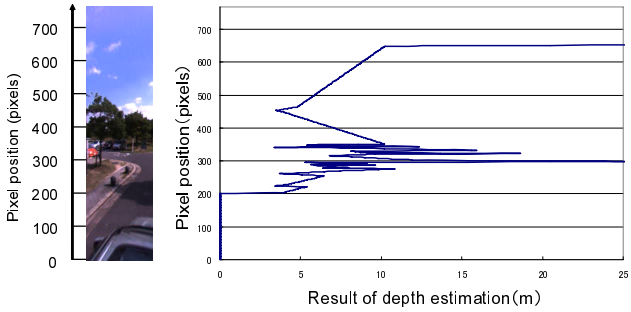
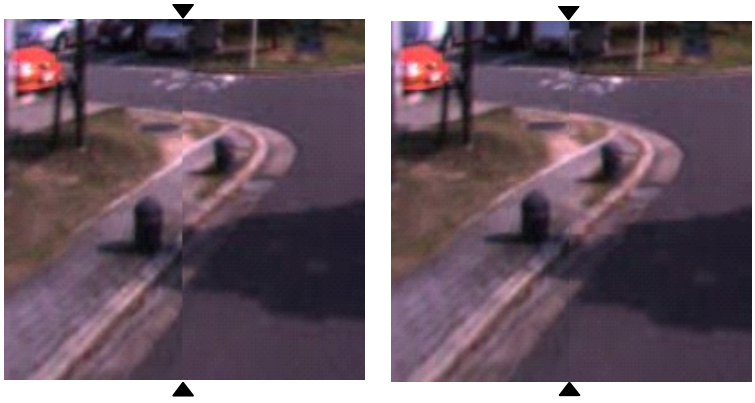


Fig. 11. Result of depth estimation



(a)Without depth information (b)With depth information

Fig. 12. Example of generated novel view images coordinated with and -out depth information

(i)View of going direction

(ii)View of 60 degrees turned from going direction

(iii)View of 120 degrees turned from going direction

(iv)View of 180 degrees turned from going direction

(a)View of left eye (b)View of right eye

Fig. 13. Generated stereoscopic images

of the full spherical view almost the same apparent point of view. The camera position and posture are measured by a hybrid sensor which consists of a real time kinematic global positioning system (RTK-GPS) LogPakII (Nikon-Trimble) and an inertial navigation system (INS) TISS–5–40 (Tokimec). This hybrid sensor can supplement the lowness of the measurement rate of GPS and the accumulation of measuring error of INS each other. The camera position and posture are measured at a high rate with high accuracy by this hybrid sensor. Input omni-directional images (resolution: 2048×1024 pixels) are shown in Figure 8. We captured the omni-directional movies along a free path as shown in Figure 9.

Figure 10 shows input images whose vertical center lines are observed along a same ray in the real world. Figure 11 shows the result of depth estimation on the line. In Figure 10, the distances between objects in an outdoor environment and camera position become far in the upper part. In the result of depth estimation as shown in Figure 11, the distances between objects and captured position are almost same as the real environment. From Figure 12, we can confirm that the distortion on the boundary was reduced by rendering coordinated with depth information. Using a PC (Pentium D 3.0GHz, memory 3.0GB), calculation cost is about 5.8 sec for one image generation. We generated binocular stereoscopic images in off-line processing when the view direction is turned by five degrees at a time. Examples of stereoscopic views (resolution: 1024×768 pixels) are given in Figure 13. We confirmed that these images can correctly show a stereoscopic view with glasses and display for a stereoscopic vision. Using generated and stored images, we can see stereoscopic view and free looking around following the indication of a user interactively were possible.

4 Conclusion

In this paper, we have proposed a method for generating novel stereoscopic view from omni-directional image sequences in wide outdoor environments. The proposed method can reduce the distortion, which is generated by conventional method in[11], by rendering coordinated with depth information that is estimated by DP matching between two images which are captured a same ray. In experiments, a user could look around a scene in outdoor environments and well perceive parallax in generated stereoscopic view. When omni-directional image sequences are captured in an outdoor environment, some moving objects such as human or vehicle are often observed. In order to generate a novel stereoscopic image, we should investigate a method for eliminating moving objects from the omni-directional image sequences.

References

1. Moezzi, S.: (ed): Special Issue on Immersive Telepresence, IEEE MultiMedia 4(1), 17–56 (1997)
2. El-Hakim, S.F., Brenner, C., Roth, G.: A Multi-sensor Approach to Creating Accurate Virtual Environments. Journal of Photogrammetry & Remote Sensing 53, 379–391 (1998)

3. Zhao, H., Shibasaki, R.: Reconstruction of Textured Urban 3D Model by Fusing Ground-Based Laser Range and CCD Images. IEICE Trans. Inf. & Syst. E-83-D(7), 1429–1440 (2000)
4. Asai, T., Kanbara, M., Yokoya, N.: 3D Modeling of Outdoor Environments by Integrating Omnidirectional Range and Color Images. In: Proc. Int. Conf. on 3-D Digital Imaging and Modeling (3DIM), pp. 447-454 (2005)
5. Adelson, E.H., Bergen, J.R.: The Plenoptic Function and the Elements of Early Vision. In: Landy, M., Movshon, J. (eds.) Computer Models of Visual Processing, pp. 3–20. MIT Press, Cambridge (1991)
6. McMillan, L., Bergen, J.: Plenoptic Modeling: An Image-Based Rendering System, In: Proc. SIGGRAPH'95, pp. 39–46 (1995)
7. Gortler, S., Grzeszczuk, R., Szeliski, R., Cohen, M.: The Lumigraph. In: Proc. SIGGRAPH'96, pp. 43–54. ACM Press, New York (1996)
8. Shum, H.Y., He, L.W.: Rendering with Concentric Mosaics. In: Proc. SIGGRAPH'99, pp. 299-306 (1999)
9. Chen, E.: QuickTime VR -An Image-Based Approach to Virtual Environment Navigation. In: Proc. SIGGRAPH'95, pp. 29–38. ACM, New York (1995)
10. Ikeda, S., Sato, T., Kanbara, M., Yokoya, N.: Immersive Telepresence System with a Locomotion Interface Using High-Resolution Omnidirectional Videos. In: Proc. IAPR Conf. on Machine Vision Applications, pp. 602–605 (2005)
11. Yamaguchi, K., Yamazawa, K., Takemura, H., Yokoya, N.: Real-Time Generation and Presentation of View-Dependent Binocular Stereo Images Using a Sequence of Omnidirectional Images. In: Proc. 15th IAPR Int. Conf. on Pattern Recognition (ICPR2000), vol. IV, pp. 589–593 (2000)
12. Ono, S., Ogawara, K., Kagesawa, M., Kawasaki, H., Onuki, M., Honda, K., Ikeuchi, K.: Driving View Simulation Synthesizing Virtual Geometry and Real Images in an Experimental Mixed-Reality Traffic Space. In: Int. Sympo. on Mixed and Augmented Reality, pp. 214–215 (2005)
13. Levoy, M., Hanrahan, P.: Light Field Rendering. In: Proc. SIGGRAPH'96, pp. 31–42. ACM, New York (1996)
14. Bellman, R.: Dynamic Programming. Princeton University Press, Princeton (1957)
15. Bellman, R., Dreyfus, S.: Applied Dynamic Programming. Princeton University Press, Princeton (1962)
16. Sakoe, H., Chida, S.: A Dynamic Programming Algorithm Optimization for Spoken Word Recognition. IEEE Tran. on Acoust. Speech and Sinal Proc. ASSP-26(1), 43–49 (1978)