Real-time Generation and Presentation of View-dependent Binocular Stereo Images Using a Sequence of Omnidirectional Images

Abstract

This paper presents a new method to generate and present arbitrarily directional binocular stereo images from a sequence of omnidirectional images in real time. A sequence of omnidirectional images is taken by moving an omnidirectional image sensor in a static real environment. The motion of the omnidirectional image sensor is constrained to a plane. The sensor's route and speed are known. In the proposed method, a fixed length of the sequence is buffered in a computer to generate arbitrarily directional binocular stereo images using a subset of plenoptic function. Using the method a user can look around a scene in the distance with rich 3-D sensation without significant time delay. This paper describes a principle of real-time generation of binocular stereo images. In addition, we introduce a prototype system of view-dependent stereo image generation and presentation.

1. Introduction

In recent years, the acquisition of a remote scene for telepresence or surveillance has become of great interest [1]. The important issue in telepresence is to produce the wide field of view and a stereoscopic effect. A telepresence system that expands the field of view has been developed by Onoe et al. [8, 9]. This system uses an omnidirectional image sensor, and generates view-dependent images without significant time delay. However the generated images are monoscopic. One of approaches to produce both the wide field of view and a stereoscopic effect is to use rotating stereo cameras. But this approach has a problem in terms of time delay from the change of viewing direction to the change of displayed image caused by rotating remote stereo cameras.

Recently, image-based rendering techniques have been proposed to generate a virtual environment from a set of photographs. QuickTime VR [3] is a system that uses these techniques. The user can alter viewing direction, but the viewpoint is fixed. A typical approach to relax this fixed position constraint is the plenoptic function [2, 6]. The Lightfield [5] and the Lumigraph [4] defined as a 4D plenoptic function make it possible to render images of a scene from any arbitrary viewpoint in a bounding box without finding depth values. Using the plenoptic functions an arbitrarily directional binocular stereo images can be obtained. However, such methods require a lot of images to render arbitrary views.

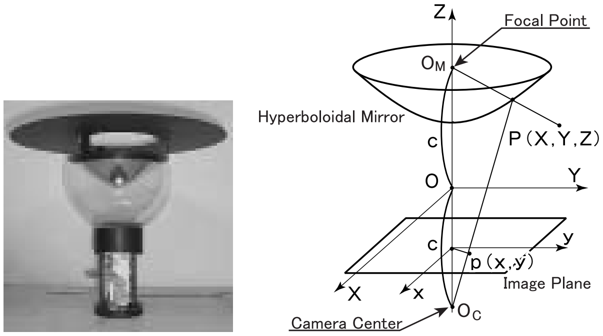
In this paper we propose a new method to generate an arbitrarily directional binocular stereo images from a sequence of omnidirectional images in real time by using a subset of plenoptic function. A sequence of omnidirectional images is taken by moving an omnidirectional image sensor in a static real environment. We constrain the motion of the omnidirectional image sensor to a plane and the sensor's route and speed are known. Compared with the Lightfield and the Lumigraph, the proposed method has smaller data size, as we constrain the viewpoint to the proximity of the sensor's route.

This paper is structured as follows. We first briefly describe the video-rate omnidirectional image sensor in Section 2. Real-time generation of arbitrarily directional binocular stereo images is then discussed in Section 3. We describe a prototype system of view-dependent stereo images generation and presentation, as well as experimental results in Section 4. We finally summarize the present work in Section 5.

2. Omnidirectional Image Sensor

We employ HyperOmni Vision [11] as an omnidirectional image sensor in our study. HyperOmni Vision uses a hyperboloidal mirror and a standard CCD video camera as shown in Figure 1(a). The geometry of HyperOmni Vision is illustrated in Figure 1(b). The hyperboloidal mirror has two focal points, an inner focal point O_M and an outer focal point O_C. The optical center of the camera lens is placed at O_C. Given a world coordinate (X, Y, Z) as shown Figure 1(b), the hyperboloidal mirror surface is represented as follows:

$$\frac{X^2 + Y^2}{a^2} - \frac{Z^2}{b^2} = -1 \ \ (Z > 0). \tag{1}$$



(a) Appearance (b) Geometrical configuration

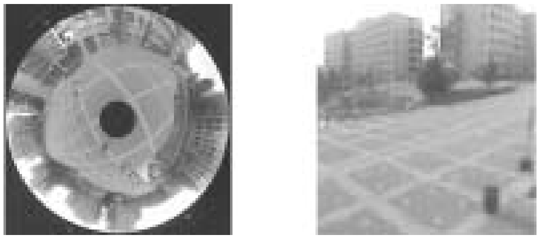
Figure 1. HyperOmni Vision ver.2A.

The inner focal point O_M of the mirror is at $(0, 0, c)$ and the outer focal point O_C is at $(0, 0, -c)$ in the world coordinate, where $c = \sqrt{a^2 + b^2}$.

A ray going toward the inner focal point O_M is reflected by the mirror and is focused on the outer focal point O_C. Thus, we can get a central projection of 360-degree filed of view onto the hyperboloidal surface by the CCD camera. A point $P(X, Y, Z)$ in world coordinate is projected to a point $p(x, y)$ on an image plane. This projection is represented as:

$$
\begin{aligned}
x &= \frac{Xf(b^2 - c^2)}{(b^2 + c^2)(Z - c) - 2bc\sqrt{X^2 + Y^2 + (Z - c)^2}}, \\
y &= \frac{Yf(b^2 - c^2)}{(b^2 + c^2)(Z - c) - 2bc\sqrt{X^2 + Y^2 + (Z - c)^2}}.
\end{aligned}
\tag{2}
$$

An omnidirectional image taken by HyperOmni Vision is shown in Figure 2(a). The generation of perspective or panoramic images from the omnidirectional image is relatively easy, because the projection is a central projection and can be converted by a simple mapping [8]. Figure 2(b) is a perspective image computed form the omnidirectional image shown in Figure 2(a). This transformation is performed by calculating correspondences of all points in the perspective image to points in the omnidirectional image.



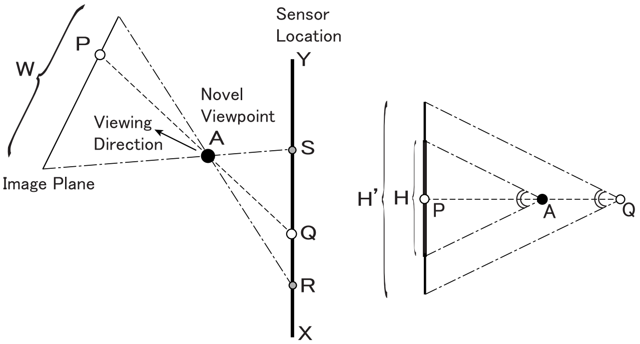
(a) (b)

Figure 2. Omnidirectional image (a) and perspective image (b).

3. Real-time Generation of Novel Binocular Views

3.1. Rendering a Novel View

A sequence of omnidirectional images that is taken by HyperOmni Vision moving straight in a static scene is used. Route and speed that HyperOmni Vision moves on a straight line are known. Given a sequence of omnidirectional images, we can render a novel view at an arbitrary point located in the plane on which a sequence of omnidirectional images is taken.



(a) top view (b) side view

Figure 3. Illustration of generating novel views.

The basic idea of novel view generation is illustrated in Figure 3. HyperOmni Vision moving from the point X to the point Y takes a sequence of omnidirectional images. Let us describe how to generate a perspective image of W by H pixels from the viewpoint A. The generation is done by computing all pixels on the image plane. For example, when computing pixels on the vertical line that includes the point P, pixels are taken from the omnidirectional image captured at the point Q where the line $\overline{\text{PA}}$ and the line $\overline{\text{XY}}$ intersect, because the ray AP is the same as the ray QP captured at the point Q [4, 5]. The vertical line that includes the point P is rendered by generating perspective image of 1 by H' pixels from the omnidirectional image captured at the point Q and scaling the resulting image to H pixels. It should be noted in this case that the focal length is $|\text{QP}|$ and the vertical field of view is the same as the vertical field of view with which the vertical line of H pixels is generated from the point A. Similarly, we compute the entire image of viewpoint A by using omnidirectional images obtained from the point R to the point S.

This method causes some vertical distortions in the image generated. If the distances between the image sensor and objects in a scene are known, we can reduce the vertical distortion. However, we actually do not know the distance and cannot correct the distortion. The vertical distortion decreases as the distance between the objects in a scene and the sensor increases [10].

A sequence of omnidirectional images is captured at video-rate. This means that the omnidirectional images on the line $\overline{\mathrm{XY}}$ are obtained at discrete points. Thus, there is a case that omnidirectional image required to generate a novel view does not exists. In such a case, we generate a novel view by using an omnidirectional image of the nearest neighbor.

3.2. Generation of Stereo Images

We generate arbitrarily directional binocular stereo images from a sequence of omnidirectional images by using the method described in the previous section. Here, we assume that both left and right eyes are located on the plane where a sequence of images is taken. Images for left and right eyes can be computed by specifying two different viewpoints with a certain distance.

3.3. Reduction of Correspondence Calculation

In order to generate binocular stereo images by using the proposed method, we must calculate the correspondence between stereo images and omnidirectional images. Such a computation is time consuming and is difficult to be performed in real time. We have developed the following method to achieve real-time generation of stereo images.

Omnidirectional images exist at discrete points on the route of HyperOmni Vision, say T_1 to T_5 in Figure 4. An image to be generated can be vertically divided according to omnidirectional images used for rendering the region. In Figure 4, a perspective image is generated from three omnidirectional images captured at T_2, T_3 and T_4 and can be vertically divided into three areas. We space a grid in every area as shown in Figure 4 and compute the exact transformation only for grid points. Stereo images are efficiently computed from clipped omnidirectional images and grid points by using an image warping technique with bi-linear interpolation [7].

The flow of an efficient stereo image generation algorithm is as follows.

1. The destination stereo images are vertically divided according to the number of omnidirectional images used.
2. A $m \times n$ grid is spaced in every divided area, where $m = \lceil \frac{w}{W} \times 16 \rceil$, $n = 12$.
   (w is a width of area, and W is a width of computed image.)
3. The exact transformation is computed for every grid point.

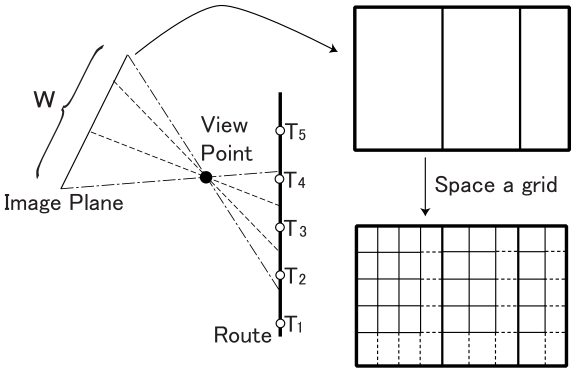


Figure 4. Division image and spacing a grid.

4. Hardware texture mapping generates the rectangular area in stereo images.

We can efficiently generate a pair of binocular stereo images of viewing direction by using this algorithm.

4. Prototype System for Immersive Telepresence

We have implemented the proposed method and constructed a prototype system of stereo images generation and presentation. The prototype system generates binocular stereo images from a sequence of temporarily recorded omnidirectional images in real time and continuously presents binocular stereo images of user's arbitrary viewing direction.

4.1. System Overview

The prototype system configuration is illustrated in Figure 5. In the system we have used a graphics workstation (SGI Onyx2 IR2 (MIPS R10000 195MHz, 16CPU)) with a video board (DIVO), HyperOmni Vision ver.2A, a robot (Nomad-200), a 3D magnetic tracker (POLHEMUS 3SPACE FASTRAK) and a head mounted display (HMD) (OLYMPUS Mediamask).

The flow of process in the system is as follows.

1. A robot with HyperOmni Vision is moved straight at constant speed by the workstation. The latest sequence of omnidirectional images is buffered in a workstation.
2. Viewing direction of a user is sent to a workstation from a 3D magnetic tracker attached to the HMD.
3. Binocular stereo images of user's viewing direction are generated in real time and are displayed on the HMD.

The process of generating binocular stereo images is executed using a single CPU of the workstation.

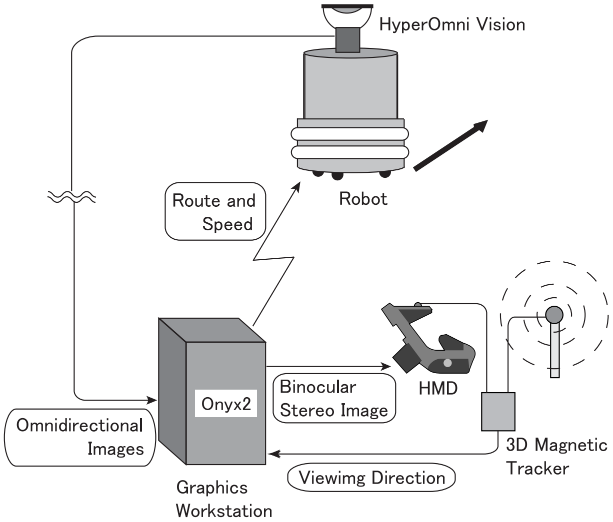


Figure 5. Hardware cofiguration of prototype system.

4.2. Experiments

In our experiment of binocular stereo image generation and presentation, the robot with HyperOmni Vision is moved straight at a constant speed of 17.5 cm per second. A sequence of omnidirectional images for the latest two second is buffered in the workstation memory. The system generates binocular stereo images that the center of eyes is located at the point of acquiring the central omnidirectional image among buffered images in the workstation. Therefore, a user looks around a scene from the viewpoint at the point that the robot passed one second ago.

The distance between eyes for binocular stereo images is set to 7cm. The horizontal field of view of the image is 60 degree. The computed images are of 640 × 480 pixels.

Figure 6 shows a sampled input sequence of omnidirectional images in the period of 35 second. Generated binocular stereo images are shown in Figure 7. In Figure 7, the binocular parallax can be clearly observed between left and right images of stereo pair. The parallax is easily recognized by watching a chair, which is closer to the sensor, in stereo images (a), (b) and (c) of Figure 7. Vertical distortions in stereo images are also observed. However it has been proven that a user can look around the scene with rich 3-D sensation without any sense of incompatibility.

A pair of stereo images is computed in 0.017 second. Images displayed on HMD are updated every 0.033 second (video-rate). In the present setup, we can generate stereo images when a user looks at an angle of from 40 to 140 degree and from –40 to –140 degree against the direction towards which the robot is moving. It should be noted that the shortage of omnidirectional images buffered in the workstation causes the fact that we cannot generate stereo images when viewing direction is close to the direction that robot is moving. The number of omnidirectional images used with respect to viewing direction is illustrated in Figure 8. The vertical axis in Figure 8 shows the number of omnidirectional images required for generating each of stereo pair. The number of omnidirectional images used for viewing direction from –40 to –140 degree is same as that for the direction from 40 to 140 degree as shown in Figure 8.

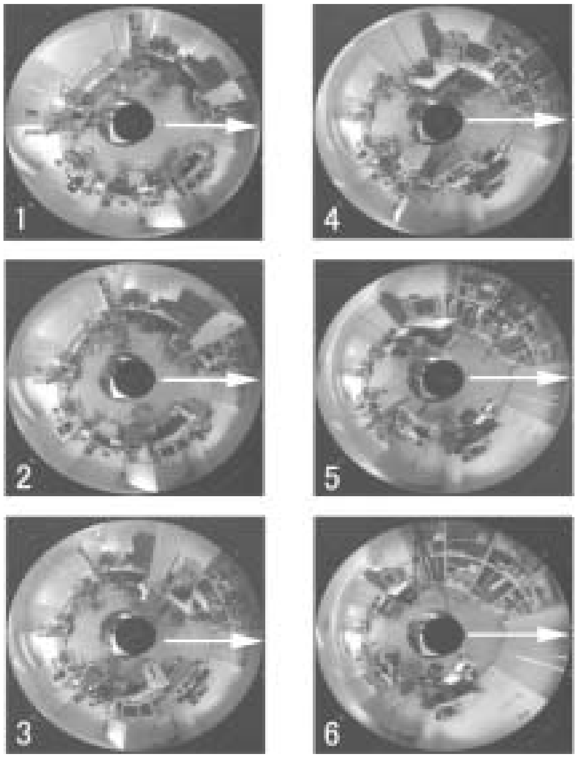


Figure 6. An input sequence of omnidirectional images (arrows show moving direction of sensor).

5. Conclusions

In this paper we have proposed a new method to generate novel binocular stereo views in arbitrary direction from a sequence of omnidirectional images in real time. We have constructed the prototype of stereo images generation and presentation system by using the proposed method. In our experiments, a user could look around a real scene in the distance and well perceive parallax in generated binocular stereo images.

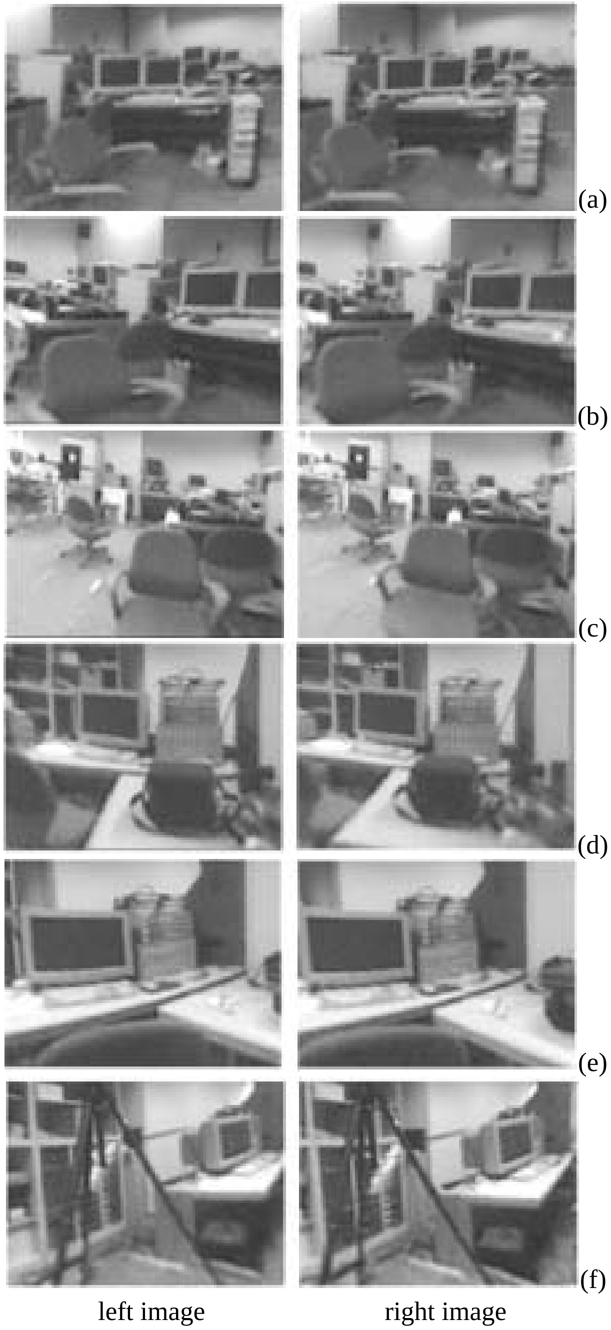


Figure 7. A sequence of generated binocular stereo images in the period of 35 sec. (viewing directions are (a) 52, (b) 96, (c) 122, (d) -56, (e) -81, (f) -118 degree from the moving direction, respectively).

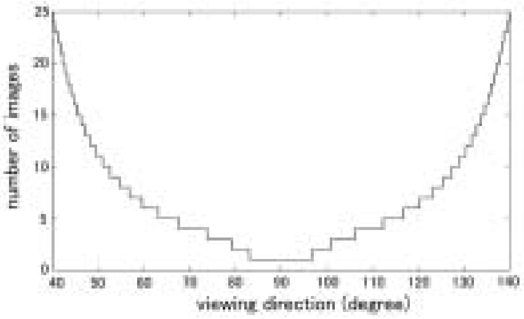


Figure 8. The number of omnidirectional images used with respect to viewing direction.

The future work is to generate stereo images in all directions and to compensate vertical distortions.

References

[1] Special issue on immersive telepresence. *IEEE MultiMedia*, 4(1):17–56, 1997.

[2] E. H. Adelson and J. Bergen. The plenoptic function and the elements of early vision. *Computational Models of Visual Processing*, pages 29–38, 1991.

[3] S. E. Chen. QuickTime VR - An image-based approach to virtual environment navigation. *Proc. SIGGRAPH 95*, pages 29–38, August 1995.

[4] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. *Proc. SIGGRAPH 96*, pages 43–54, August 1996.

[5] M. Levoy and P. Hanrahan. Light field rendering. *Proc. SIGGRAPH 96*, pages 31–42, August 1996.

[6] L. McMillan and G. Bishop. Plenoptic modeling: An image-based rendering system. *Proc. SIGGRAPH 95*, pages 39–46, August 1995.

[7] J. Neider, T. Davis, and M. Woo. *OpenGL Programming Guide*. Addison-Wesley Publishing Company, 1993.

[8] Y. Onoe, K. Yamazawa, H. Takemura, and N. Yokoya. Telepresence by real-time viewdependent image generation from omnidirectional video streams. *Computer Vision and Image Understanding*, 71(2):154–165, August 1998.

[9] Y. Onoe, K. Yamazawa, H. Takemura, and N. Yokoya. Visual surveillance and monitoring system using an omnidirectional video camera. *Proc. 14th IAPR Int. Conf. on Pettern Recognition(14ICPR)*, I:588–592, August 1998.

[10] K. Yamaguchi, K. Yamazawa, H. Takemura, and N. Yokoya. Realtime generation and presentation of arbitrarily directional binocular stereo images using a sequence of omnidirectional images. *IEICE Tech. Report, PRMU99-159*, pages 67–72, November 1999. (in Japanese).

[11] K. Yamazawa, Y. Yagi, and M. Yachida. Obstacle detection with omnidirectional image sensor hyperomni vision. *Proc. 1995 IEEE Int. Conf. on Robotics and Automation*, pages 1062–1067, May 1995.