2004－CVIM－143　(5)
2004／3／5

任意回転のステレオ視用画像生成

千葉 直樹[†]　　Terence T. Huang [‡]
[†]三洋電機株式会社
[‡]Simon Fraser University
mail@naoki-chiba.com, terenceh@sfu.ca

概要　複数の画像から任意の視線の回転に対応したステレオ視用画像を生成する手法を提案する。全方位方向の写実的な画像を生成する仮想現実システムが注目を集めている。従来のステレオ視用画像を生成するシステムはユーザの視線の回転を水平方向（パン）に限る問題があった。この問題を解決し、チルトや光軸周りの回転を許容する手法を提案する。この手法はコンピュータ制御可能なパン・チルト回転台にカメラを設置して画像を撮影する。これらの画像から任意回転のステレオ視用画像を三次元的な奥行きを再構成することなく、生成する手法を述べる。実画像により、この手法の有効性を示す。

Stereoscopic Image Generation for Arbitrary Viewing Rotations

Naoki Chiba　　Terence T. Huang [‡]
SANYO Electric Co., Ltd.　　Simon Fraser University

mail@naoki-chiba.com, terenceh@sfu.ca

Abstract　We propose a method to generate stereoscopic images for arbitrary viewing rotations from a collection of images. Virtual reality systems, which can provide photo-realistic omni-directional views, are collecting a lot of attention. Existing methods for creating stereoscopic images, however, have a problem that limits viewing rotations to be only horizontal panning. To solve this problem, we have developed a method that can generate images not just for panning but also for other rotation orientations such as tilting and twisting. The method uses a single digital-camera mounted on a computer-driven pan-tilt unit whose rotating center has a fixed offset to the projection center. We present how to generate stereoscopic images from the images acquired with such camera configuration. Our method is completely image-based without recovering 3D depth. Experiments using real images show that our method is very effective to generate images for arbitrary viewing rotations. Our method can generate images supporting six degrees of freedom in 3D space; in other words, it reconstructs a 4D plenoptic function.

[‡]The research project is carried out during Terence T. Huang's internship at SANYO Electric Co., Ltd.

1 INTRODUCTION

Virtual reality systems, which can provide photo-realistic views, are collecting a lot of attention. One of the first systems is QuickTime VR [CHEN95]. After constructing a panoramic image from a collection of images acquired by a panning camera, the system can interactively playback sections of the 360-degree horizontal panorama. Methods extending the work to construct spherical mosaics have been proposed. They use images taken with a hand-held camera [SH97] or a camera mounted on a computer-controlled pan-tilt unit [WM96].

Methods for generating an arbitrary view can be categorized into two types. One is to build a 3D model by using computer vision techniques. The other is to generate a view by using acquired images, which is called image-based rendering (IBR).

Among IBR techniques for providing arbitrary views of an environment model, two approaches have been proposed. One is to use an omni-directional camera that can capture a horizontal 360-degree view at once by using a convex mirror [YY01]. The other approach is to stitch a collection of images captured with a standard camera, which is called image mosaics.

One of the research trends in image mosaics is to extend the number of degrees of freedom for viewing. In the following subsection, we review previous methods for generating stereoscopic views. The approach has a strong connection with the plenoptic function which is well-known in computer graphics research. After reviewing related work in this area, we present their problems to which we will show a solution with our method.

1.1 Stereo Panorama

By using a single rotating camera, methods for creating panoramic stereoscopic views have been proposed in [IYT92], [PBE99] and [NK00]. Ishiguro et al. used two vertical stripes of a camera swivelling around a rotation center with the optical axis remaining perpendicular to the tangent of the camera path in [IYT92]. Two horizontal panoramas were built by stitching the vertical stripes. They used rotation angles measured by the computer-driven turntable. Stereo pairs were obtained from the panoramas to estimate depth information for building depth maps by a mobile robot.

Peleg et al. extended the work to use a video camera on a levelled tripod in [PBE99]. After estimating camera motion between video frames, two stereo panoramas were constructed again by stacking dual stripes from each image. Instead of estimating depth, they generated stereoscopic views for a visualization purpose.

Nayar et al. created stereo spherical panoramas called *360 x 360 Mosaics* [NK00]. A “slice camera” employing a parabolic mirror and a telecentric lens enlarged the vertical field of view of each slit to 360 degrees. Each 180-degree half of the slit was used in constructing a spherical panorama for one eye viewpoint.

1.2 The Plenoptic Function and Novel View Generation

Adelson and Bergen have introduced the concept of the plenoptic function [AB91] that can describe the intensity of light seen at every viewpoint (V_x, V_y, V_z) at every viewing angle (θ, φ) for every wavelength (λ) at every time (t). Fig. 1 illustrates the function for one single viewpoint in one direction. The 7D function can be written as $p = P(\theta, \varphi, V_x, V_y, V_z, \lambda, t)$. If we can record a real scene as a smooth plenoptic function, then any arbitrary view of the scene can be “played back” at any time. If the scene is static without light changes and captured with a RGB camera, the last two dimensions can be omitted.

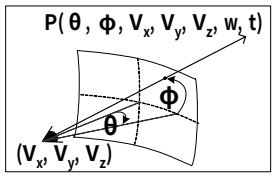


Fig. 1 The plenoptic function for one single viewpoint in direction (θ, φ)

McMillan and Bishop defined the 5D plenoptic function in their system based on a set of cylindrical panoramas imaged at various points in space [MB95].

If we assume that radiance along any ray remains constant within a closed space, we can reduce the number of parameters in the space. The *Light Field Rendering* [LH96] and *The Lumigraph* [GGSC96] presented a 4D parameterization of the plenoptic function if it can be constrained within a bounding box. Their representation uses a set of coordinates on two planes when a light ray crosses them. The function is $p = P(s, t, u, v)$, where (s, t) and (u, v) are coordinates on the planes.

Naemura et al. also showed reduction of the function parameterization to 4D if the space is divided by a 2D surface into an observing area and an observed area. When we capture images with cameras on the surface, we can reduce the number of dimensions by projecting light rays onto a planar screen, described by 2D parameters P and Q, facing each direction of rays. The function is described by $p = P(P, Q, \theta, \varphi)$. Since the shape of the surface is not restricted to be planar, it can also be spherical.

Shum and He’s *Concentric Mosaics* [SH99] showed that the plenoptic function can be further scaled down to three dimensions by confining viewpoints within a planar circular region. One of their suggested implementations in fact coincides with those in [IYT92, PBE99].

Recent upgraded systems include *Unstructured Lumigraph Rendering* by Buehler et al. [BBMGC01], *Dynamically Reparameterized Light Fields* by Isaksen et al. [IMG00], and *Plenoptic Stitching* by Aliaga et al. [AC01].

1.3 Limitations

The stereo panoramas are limited to reproducing only horizontal parallax. The lack of vertical parallax prevents the viewer from gaining stereo sensation if, for example, the virtual camera is twisted around the optical axis or is oriented to the polar regions. Furthermore, in the case of spherical panoramas [NK00], the output image quality tends to degrade significantly towards the poles. With current omnidirectional imaging apparatus such as parabolic mirrors, it is difficult to obtain a 360-degree image slice with uniform resolution.

Previous IBR systems based on the plenoptic function have some problems as follows. In *Light Field Rendering* and *The Lumigraph*, the allowable viewing space is affected by the geometric constraints of the acquisition system configured to work with the chosen parameterization for the plenoptic function. As Shum et al. have noted in [SH99], it is difficult to capture an inside-looking-out *Lightfield* or *Lumigraph* of a real scene, which has also been examined in [TS03]. While the above issue is loosened with *Concentric Mosaics*, which do permit an omnidirectional viewing range of a surrounding scene on a flat plane, the method again lacks both vertical field of view and vertical parallax.

1.4 The Proposed method

We propose a method to solve the above problems that can allow the user to explore a virtual environment with stereoscopic views. Unlike previous stereo panoramas which restrict viewing rotations to panning, it enables the user to rotate their heads in all orientations by panning, tilting and twisting. The proposed method also permits lateral motion within the viewing space.

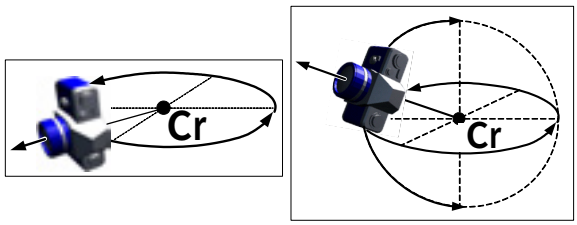
Our method is an extension of [IYT92, PBE99, SH99] and preserves the four-dimensionality of [NTKH97] by exploiting a special spherical parameterization of the plenoptic function and a novel acquisition technique with a computer driven pan-tilt-unit. We also show a method of rendering novel images that deals with approximation errors in the plenoptic function during resampling. Experiments using real images demonstrate that we can generate very high quality images for six degrees of freedom and stereo disparity regardless of view orientation.

The remainder of the paper is organized as follows. We present the configuration for image capturing and the procedure for novel view rendering. After showing the performance of our method with experiments, we conclude with a discussion on performance issues and future work.

2 ACQUISITION

2.1 Camera Configuration

Fig. 2 contrasts the camera configurations used in the previous method [PBE99] and in our method. We rotate a camera along the surface of a sphere by using a computer-driven pan-tilt unit (PTU). The offset between the path of camera's projecting center and the center of rotation is fixed. The PTU provides vertical rotation in addition to horizontal rotation as illustrated in Fig. 2 (b).



(a) Previous method [PBE99] (b) Our method

Fig. 2 Camera configuration.

2.2 Capture Routine

The capture routine is performed by a software program that sends rotation commands to the PTU to orient the camera at desired angles and shutter commands to the digital camera to capture images. The routine relies on an internal representation of the viewing sphere which provides pairs of precise spherical angles for each capture point. The sphere model can be a collection of vertices and polygons defining its geometric surface. Because we want to capture at uniformly spaced angles, we have opted a surface model based on the geodesic dome structure.

A geodesic dome can be constructed from scratch by subdividing a tetrahedron, an octahedron, or an icosahedron. One that is subdivided from an icosahedron distributes the most equal weight for each child face and thus gives the most uniform spherical surface structure. Another advantage of subdivided geodesic domes is that we can control the number of subdivisions to adjust the sampling frequency at discrete steps.

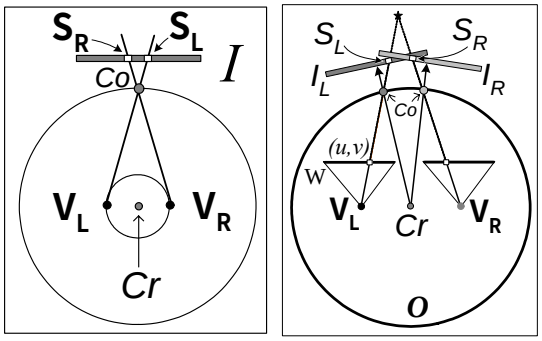
3 RENDERING

Once we have discretely sampled images, we can generate an arbitrary view of the scene. We describe a rendering routine that achieves this resampling procedure and how we deal with the involved approximation error.

3.1 Rendering Routine

Fig. 3 (b) shows a top-down view of a virtual image plane w and a sphere O which defines the path of the capturing camera's optical center. The coordinate pair (u, v) indicates a pixel position on w. The rendering routine is basically the mapping from captured image pixels to the virtual image pixels. One way to implement this mapping

function is to follow a geometrical approach instead of solving the algebraic relationships which might take a lengthy process. A simple implementation is through ray-tracing.



(a) Previous method [PBE99] (b) Our method

Fig. 3 Rendering a novel image. In (a), only two stripes from each source image are used to create the stereo cylindrical panoramas. In our method in (b), a ray-tracer determines corresponding source pixels for each virtual image element.

The output of the ray-tracer will directly tell us where the source pixel for each virtual image element should come from in terms of the capture angle of the source image and its location in local source image space. First, a ray is shot from the projection center V_L towards (u, v) in world coordinate space. Its intersection with O is returned as a spherical coordinate pair $(\theta_{u,v}, \varphi_{u,v})$ using Cr as the world origin. A second ray leaves the world origin towards the previous intersection point and finds where it crosses the corresponding source image. We can render the entire virtual image by following the same procedure for all (u, v).

In the case where we do not have an image for $(\theta_{u,v}, \varphi_{u,v})$, we apply the nearest-neighbour strategy and use the source image having the closest capture angle in the second step.

3.2 Handling Approximation Error

There is a source of approximation error due to the fact that we only have a finite number of source images. This approximation error is severe because it leads us to use a sample from a different ray direction during the ray-tracing procedure. [Fig. 4]

(a) (b)

Fig. 4 (a) shows an image rendered with approximation error. (b) shows the approximation error reduced with our technique

In Fig. 5, we can see that when we use an adjacent source image for ray R, we end up getting the pixel P_b that actually corresponds to object point B instead of object point A. One method for reducing this error is to choose pixel Pc which is on a ray closer to the object point A. The position of pixel Pc in local image coordinates can be obtained by scaling those of P_b with a factor described as the ratio of (f / (f + d)) if we note the similar triangles formed by $PcIcC$ and $PbIcX$.

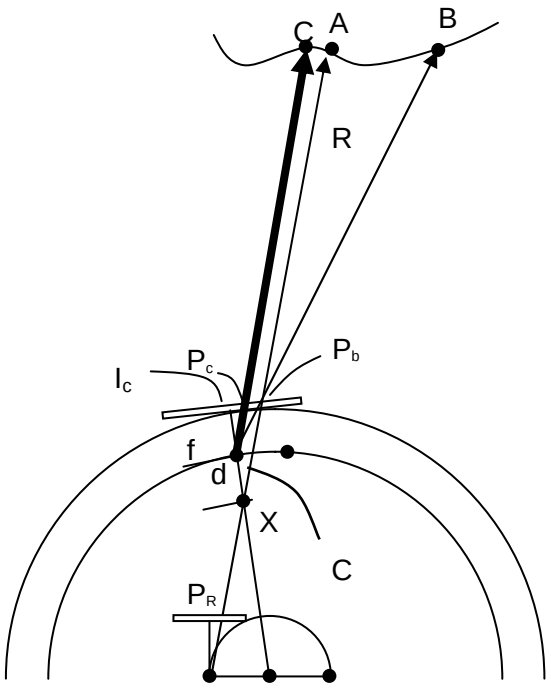


Fig. 5 Ray adjustment

4 EXPERIMENTS

4.1 Overview

First, we have obtained two sets of novel images rendered using the method described in [PBE99] and our own method. In each set, there are two stereo pairs of images, one showing no view rotation and the other showing a 90-degree twist. Next, we include experiments showing the effect of insufficient sampling rate for image acquisition and the effect of inaccurate alignment of the camera's optical axis.

For image capturing, we used an Olympus C-3040 digital camera with its Software Development Kit. We can control various functions of the camera (such as shutter, zoom, focus and etc.) and transfer captured image data through a USB cable. We fixed the shutter speed, exposure, and white balance to manual settings. Table 1 shows the specification of the camera. The effective focal length and the radial lens distortion were measured by using a variation of Zhang's calibration method.

Table 1

Focal length (35mm-film equivalent)	36 mm
Image Size (pixels)	1024 x 768
Image Format	JPEG

For the capture stage, we utilized a pan-tilt unit (model no. PTU-46-70 with nodal-gimbal upgrade option) manufactured by Directed Perception, Inc. The specification of the unit is as follows:

- Pan Range: 159˚ left and 159˚ right
- Tilt Range: 31˚ up and 47˚ down
- Resolution: 0.771 arc minutes (0.012857˚)

We have also made a special jig that offsets the camera from the rotation center with adjustable length from 0cm to 18cm, and interests the camera's optical axis with the rotation center. Fig. 6 shows a photo of the PTU and the jig.

Fig. 6 A photo showing the PTU and the camera jig

4.2 Comparison with a previous method

For comparison of previous work with our method, we have chosen the method described by Peleg et al. for creating stereo cylindrical panoramic images [PBE99]. The camera is offset 15cm from the rotation center of the PTU which moves in 0.5˚ steps. To create the stereo pair, we used two 22-pixel-wide vertical stripes from each image. The stripes are located 150 pixels to the left and 150 pixels to the right of the image center, respectively.

In the first pair of images without view rotation, we can observe significant horizontal parallax. However, the second pair does not show any parallax effect after the virtual view is rotated 90 degrees clockwise. [Fig. 7]

For the experiment of our method, we have captured about 5000 images of a conference room with the following capture configuration:

- Sampling step: approx. 0.7˚ in each direction
- Camera offset from rotation center: 15cm

With our proposed method, the images demonstrate clear stereo disparity in the case where there is no view rotation, as well as in the case where the viewing direction is rotated 90 degrees clockwise. [Fig. 8] Motion parallax along vertical translation is also demonstrated in Fig 11.

4.3 Insufficient Sampling Rate for Acquisition & Misaligned Camera

If we capture images less frequently by reducing the subdivision level of the spherical surface model for guiding the acquisition process, approximation errors will increase and the aliasing effect will become apparent. Similarly, if the camera's optical axis is not aligned well to intersect the rotation center, additional errors will be introduced in the rendered result. [Fig. 9, Fig. 10]

5 DISCUSSION

We have proposed a method that can generate stereoscopic views for arbitrary rotations. Unlike cylindrical stereo panorama, we do not restrict the user's head to horizontal rotation, panning. The experimental result of the 90-degree rotation in Fig. 8 showed that our method can recover a significant parallax which cannot be generated with the previous methods.

Our method is not restricted to stereoscopic image generation. We also showed the result of vertical translation of a single camera. Our method can provide motion parallax as well.

We can generate very high-quality images due to the large number of captured images. We used 84 captured images to render a 1024x768 virtual image whose focal length is 1000 pixels. The processing time to render the image was 42 seconds on a Pentium 4 2.3GHz with 4GB memory. We have also implemented a caching function to reduce the time of disk access. Without the disk caching, the processing time was 864 seconds.

As seen in Fig. 12, visible seam-lines are observable on close-ups because of finite number of images. To increase the quality of the generated images, we can think in two directions. One is to recover 3D depth for interpolating views. This is an active area of research in computer vision. One example is to apply multi-baseline techniques to recover 3D depth in cylindrical stereo panoramas [JOS04]. The other direction is to increase the number of source images. The experiment in Fig. 9 showed that the number of source images significantly affect the image quality. If we capture a greater number of images, we can easily increase the image quality. Since the PTU has its rotation resolution up to 0.001 degrees, we can add more images easily. This is our future work.

The drawback is its capturing time and rendering time. If we increase the number of capturing images, the capturing time increases linearly. Since it takes much time to capture, even a light change coming through windows can cause a problem. The scene that we can recover is limited to pure indoor scenes. Using a video camera might solve the problem partially because we can rotate the camera during capturing. This, however, affects the image quality, because video camcorders have limited resolution such as 640x480 pixels, which is much lower than that of a digital still camera. We need to wait for high-resolution video camcorders in future.

6 CONCLUSION

We have proposed a method that can generate stereoscopic views for arbitrary orientations. Unlike previous methods, the allowable rotations are not restricted to horizontal rotation. By using a collection of images captured by a single digital still camera rotated by a computer-driven pan-tilt unit, we can generate high quality views for any rotations by panning, tilting, and twisting.

We have described a system for capturing images and a procedure to generate a virtual view from them. We showed the performance against low sampling rate and misalignment of the optical axis to the center of rotation.

In our examples, we did not generate views for complete omni-directions. The method can be applicable for that purpose without any modification. Constructing a system to cover complete spherical viewing range is our future work.

References

[AB91] E.H. Adelson, J.R. Bergen. The Plenoptic Function and the Elements of Early Vision. Computation Models of Visual Processing, M. Landy and J.A. Movshon, eds., MIT Press, Cambridge, 1991.

[AC01] D. Aliaga, I. Carlbom. Plenoptic stitching: a scalable method for reconstructing 3D interactive walk throughs. Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pages 443-450, 2001.

[BBMGC01] C. Buehler, M. Bosse, L. McMillan, S. Gortler, M. Cohen. Unstructured Lumigraph Rendering**.** Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pages 425 – 432, 2001.

[CHEN95] S. E. Chen. QuickTime VR - An Image Based Approach to Virtual Environment Navigation. Computer Graphics: Proc. Of SIGGRAPH 95, pages 29–38, August 1995.

[GGSC96] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The Lumigraph. ACM SIGGRAPH, pages 43-54, 1996.

[IMG00] A. Isaken, L. McMillan, S. Gortler. Dynamically Reparameterized Light Fields. Proceedings of the 27th annual conference on Computer graphics and interactive techniques, pages 297-306, 2000.

[IYT92] H. Ishiguro, H. Yamamoto, and S. Tsuji. Omnidirectional Stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, pages 257–262, 1992.

[JOS04] W. Jiang, M. Okutomi and S. Sugimoto. Omnidirecional 3D Reconstruction Using Stereo Multi-Perspective Panoramas, Technical Report of IEICE, PRMU2003-211, pp. 77-82, 2004. (In Japanese)

[LH96] M. Levoy and P. Hanrahan. Light Field rendering. ACM SIGGRAPH, pages 31-42, 1996.

[MB95] L. McMillan and G. Bishop. Plenoptic Modeling: An Image-Based Rendering System. Computer Graphics: Proc. of SIGGRAPH 95, pages 39–46, August 1995.

[NK00] S. Nayar and A. Karmarkar. 360 x 360 Mosaics. Proc. of IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pages 388-395, 2000.

[NTKH97] T. Naemura, T. Takano, M. Kaneko, H. Harashima. Ray-Based Creation of Photo-Realistic Virtual World. Proc. of VSMM '97, pages 59-68 1997.

[PBE99] S. Peleg and M. Ben-Ezra. Stereo Panorama with a Single Camera. Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pages 395–401, June 1999.

[PS02] P. Peer and F. Solina. Panoramic Depth Imaging: Single Standard Camera Approach. International Journal of Computer Vision, vol. 47, pages 149-160, 2002.

[SH99] H.Y. Shum and L.W. He. Rendering with Concentric Mosaics. Proceedings of SIGGRAPH 99, pages 299-306, August 1999.

[SH99b]

[SK02] S. Seitz and J. Kim. The Space of All Stereo Images. International Journal of Computer Vision, vol. 48, no. 1, pages 21-38, 2002.

[SKS02] S. Seitz, A. Kalai, H.Y. Shum. Omnivergent Stereo. International Journal of Computer Vision, vol. 48, no. 3, pages 159-172, 2002.

[SS97] R. Szeliski and H. Y. Shum. Creating Full View Panoramic Image Mosaics and Environment Maps. Proc. of ACM SIGGRAPH'97, pages 251–258, August 1997.

[TS03] T. Todoroki and H.Saito. Arbitrary Viewpoint Image Synthesis via Light Field Rendering with Omni-directional Camera. IPSJ SIG Technical Report 2003-CVIM-138, pp. 45-50, 2003.

[WM96] T. Wada and T. Matsuyama. Appearance Sphere: Background Model for Pan-tilt-zoom Camera. Proc. Of ICPR, pages 718-722, 1996.

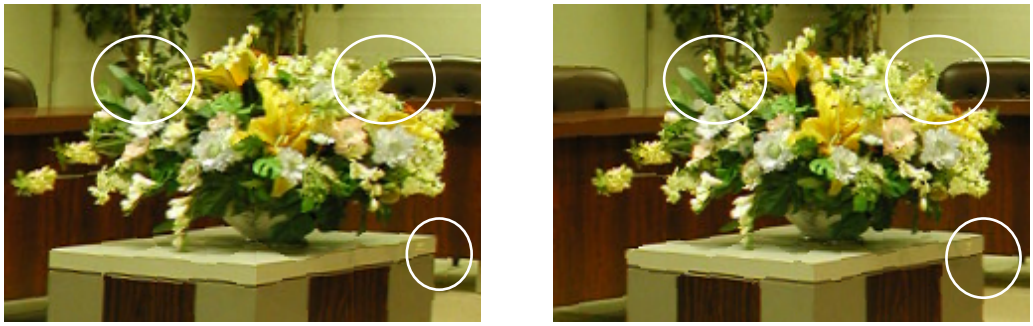
[YY01] T. Yagi and N.Yokoya. Omnidirectional Vision: A survey on Sensors and Applications. IPSJ Transactions on Computer Vision and Image Media No. SIG13(CVIM3), pages 1-17, 2001. (In Japanese)

(a) Without viewing rotation. Stereo disparity is visible.

(b) After the viewing orientation is rotated 90 degrees, no stereo disparity is visible due to the lack of vertical parallax.

Fig. 7 Novel image generation by using Peleg et al.'s method for creating cylindrical stereo

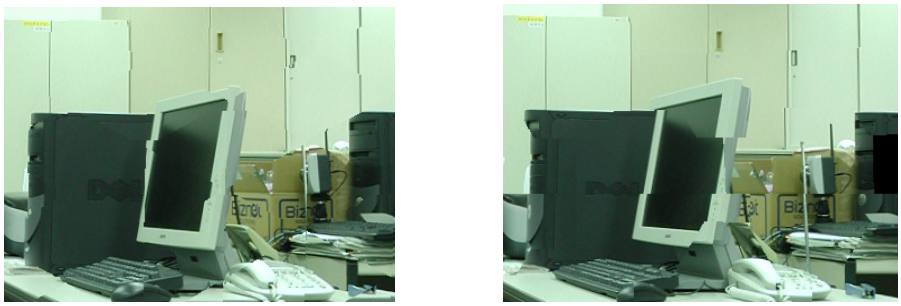
(a) A stereo pair of images without viewing rotation



(b) Without viewing rotation. Stereo disparity is significantly visible.

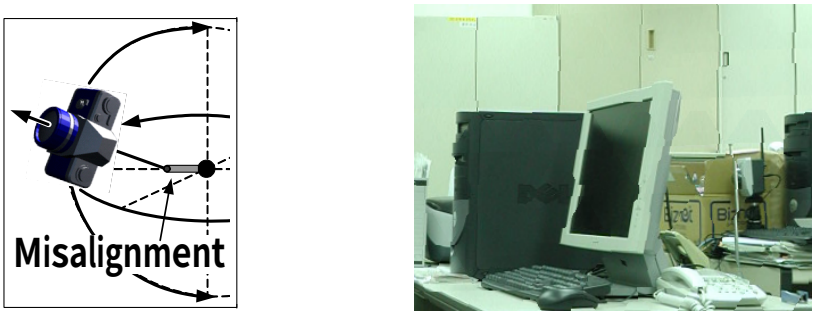
(c) Even after the viewing orientation is rotated 90 degrees, stereo disparity is still significantly visible.

Fig. 8 Novel image generation by using our method.



(a) (b)

Fig. 9 Image quality degradation: (a) and (b) demonstrate the effect of low sampling rate. The dataset for (a) has 1/3 of the number of source images while the one for (b) has only 1/9 of the number of source images.



(a) (b)

Fig. 10 Image quality degradation: (a) shows that the camera is misaligned such that its optical axis fails to intersect the rotation center. The image quality degrades as in (b).

Fig. 11 Novel images generated with our method showing vertical view translation. The vertical parallax is again significantly visible.

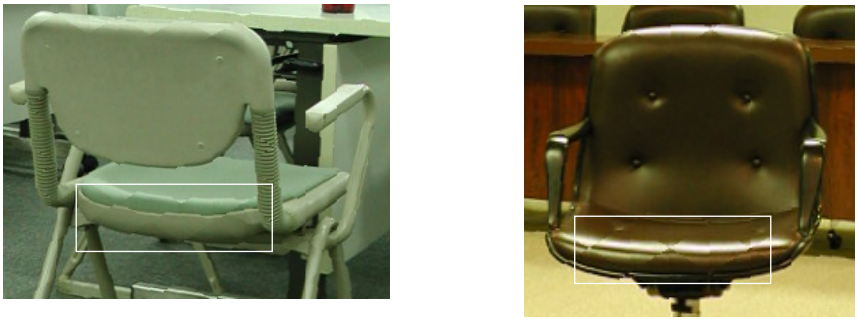


Fig. 12 Visible seams on the chairs due to a finite number of source images