Arbitrary View Position and Direction Rendering for Large-Scale Scenes

Takuji Takahashi Hiroshi Kawasaki Katsushi Ikeuchi Masao Sakauchi
Institute of Industrial Science, University of Tokyo
7–22–1 Roppongi, Minato-ku, Tokyo, 106-8558, Japan

Abstract

This paper presents a new method for rendering views, especially those of large-scale scenes, such as broad city landscapes. The main contribution of our method is that we are able to easily render any view from an arbitrary point to an arbitrary direction on the ground in a virtual environment. Our method belongs to the family of work that employs plenoptic functions; however, unlike other works of this type, this particular method allows us to render a novel view from almost any point on the plane at which images are taken. Previous methods, on the other hand, have some restraints concerning their re-constructable area. Thus, when synthesizing a large-scale virtual environment such as a city, our method has a great advantage.

One of the applications of our method is a driving simulator in the ITS domain. We can generate any view on any lane on the road from images taken by running along just one lane. Our method, using an omni-directional camera or a measuring device of a similar type, first captures panoramic images by running along a straight line, recording the capturing position of each image. When rendering, the method divides the stored panoramic images into vertical slits, selects some suitable ones based on our theory, and reassembles them for generating an image.

The method can make a virtual city with walk-through capabilities. In that virtual city, people can move and look rather freely.

In this paper, we describe the basic theory of a new plenoptic function, analyze the applicable areas of the theory and the characteristics of generated images, and demonstrate a complete working system using both indoor and outdoor scenes.

1 Introduction

To create a large-scale virtual environment, such as entire town or city, is one of the most important applications in computer vision and computer graphics. Generally speaking, there are two main approaches applicable to the problem: image-based modeling (IBM) and image-based rendering (IBR). An IBM approach analyzes the geometry and surface attributes of the objects in the environment and, based on those analyses, creates new views. The IBM approach requires precise reflectance model analysis and careful image acquisitions. By capturing a series of images, we obtain the geometry and material properties of objects directly from the real world, and we can create images of real objects and scenes using computer graphics rendering software. This method, however, is still limited in its ability to create images of relatively complex or small objects such as trees in city scenes. So it is difficult to generate a photo-realistic landscape by using IBM.[5]

Generating a 3-D virtual world directly from real scene images without analyzing a reflectance model or using an explicit 3-D model is a promising technique. This method, referred to as image-based rendering (IBR), creates new views by re-sampling those prerecorded pixels in a timely manner. “Aspen Movie Map”[6] was the pioneering work of this IBR technology. This system consists of a computer-controlled laser disc, which records the images along the streets of the town of Aspen, CO. The user can walk along the street where images have been captured. However in this environment, the user can view the images only from the original viewpoint of the camera. Another representative system based on the IBR is “QuickTime VR”[7]. A series of captured environment images, pasted on a cylindrical environment, can generate a virtual world in which this method allows users to look around a scene from fixed points. This system, however, does not allow users to walk around in the environment.

One of the key concepts developed in the IBR is the plenoptic function. Originally, a 7D plenoptic function was proposed to define the intensity of light rays passing through the camera center at every location, at every possible viewing angle, for every wavelength and at any time[10]. It has since been shown that light source direction can also be incorporated into the plenoptic function for describing illumination environments. Ignoring time and wavelength, "Plenoptic Modeling"[11] is a continuous 5D plenoptic function. However, in fact it is difficult to apply these methods for rendering novel view image. Because in the original plenoptic function, the preparation of a huge amount of data is tedious, and concerning 5D plenoptic function it is hard to describe the correspondence between cylindrical-to-cylindrical mappings or between these cylinders and a new viewing position. On the other hand, some reasonable methods for rendering have been presented. "Lumigraph"[13] and "Lightfield"[12] and "Ray-space method"[9] systems are the 4D plenoptic function with clever parameterization. "Rendering with Concentric Mosaics"[14] is 3D plenoptic function which, as its name indicates, creates concentric mosaics.

In spite of the merits of the previously proposed plenoptic functions, viewing space is greatly restricted. Namely, the viewing spaces of "Lumigraph","Lightfield" and "Ray-space method" are inside a 3D box, and "Rendering with Concentric Mosaics" can render only inside a 2D circle. So these methods are not useful for rendering large-scale scenes, especially scenes of entire towns.

One of the works most closely related with ours, rendering large-scale scenes, is Hirose's[15]. They recorded images of the entire Marunouchi district in Tokyo, and played back images on demand from users so that the users had the impression of actually walking in the environment. However, that method can show only images captured along a path; the method cannot create new views.

This paper proposes a new plenoptic function resolving those issues. Namely, the strength of our method is that we can create any view from any position to an any direction on the ground, in a wide area of a city. This method first captures panoramic images through the use of either an omni-directional camera or a combination of standard TV cameras that run along a straight line, while recording the capturing position of each image. When rendering, the method divides the stored panoramic images into vertical slits, selects some suitable ones based on our theory, and reassembles them for generating images.

This paper is organized as follows. In Section 2, we explain the basic method for obtaining the new plenoptic function, and in Section 3, discuss some of the characteristics of the method. Section 4 presents results of experiments using a real scene both of an indoor area and a landscape of the city. Section 5 contains our conclusions regarding this method, along with discussions of future work.

2 Reconstruction of Arbitrary View from Panoramic Images

2.1 Capturing Panoramic Images

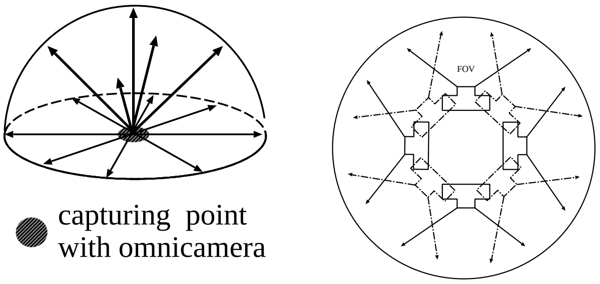
At each location, we construct a panoramic image,which contains all rays from the capturing location; we have a total horizontal viewing field of 360 degrees at each capturing location.

The simplest and easiest method for capturing panoramic images is to use an omni-directional camera. This type of camera has an orthographic lens that has a single effective viewpoint (see Fig.1-(a)). From the sensed omni-directional image, we can generate pure perspective images, and, thus can make panoramic images from the omni-directional image.

Using one omni-camera, we can take images with 360 degrees in a horizontal direction; those images cover the northern hemisphere of a viewing sphere. The images cover the directions above the image plane of the omni-directional camera.

Another method for capturing panoramic images is to arrange some cameras cylindrically as shown in Fig.1-(b) . These cameras' optical axes intersect at one point with rays around the center of cameras. So projecting these perspective images to cylindrical coordinates, we store cylindrical images at each location.

Hereafter 'panoramic image' means both of these two images.



(a) (b)

Figure 1: (a)Capturing with an omni-directional camera (b)Configuration of cameras to obtain cylindrical images

2.2 System for Capturing Images

One application of our system is to build a driving simulator. Here, we are mainly concerned with the case where a camera runs along a straight line. Namely, by running in a straight line, e.g., a lane on a road, we capture panoramic images with the position and direction of each panoramic image. Here the positional information is given by a GPS sensor.

We will denote the ground plane as the x-y plane. As shown in Fig.2, by moving from C_0 to C_n, we capture images and record their positions. Here we denote a panoramic image captured at (x_i, y_i) as $C_i(x_i, y_i)$. We can pick up arbitrary slits from each panoramic image for reconstruction purposes.

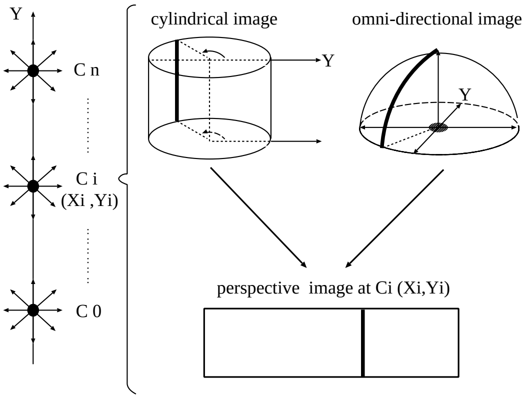


Figure 2: Capturing panoramic images

2.3 Reconstruction of Novel Views

Given a series of panoramic images on a straight line, we can construct a novel view from almost any arbitrary region. Consider the case of rendering the novel image at the point P, as illustrated in Fig.3. For constructing the view from P, we need rays around P from R_s to R_e as shown in the figure. By finding the slits corresponding to the rays from stored panoramic images and then collecting those slits, we can synthesize a new view. For example, the ray of R_s is substituted for the θ_i ray in the $C1$ panoramic image.

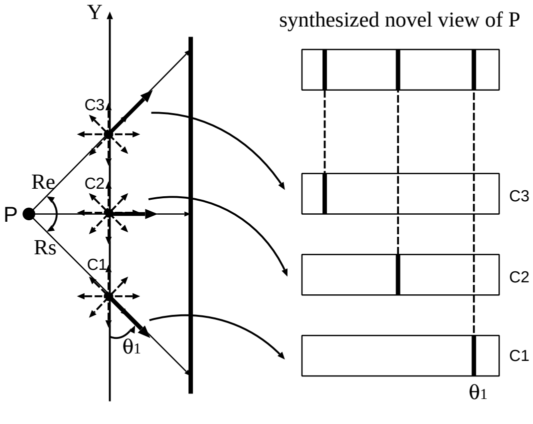


Figure 3: Reconstructing a novel view

3 Features

3.1 Region of Re-construction

In this section we discuss the areas where it is possible to reconstruct novel views. First, consider an x-y plane on the ground. Let us assume that the camera moves in a straight line for the interval from the position,C_0 to the position, C_n, see Fig.4. On or about the running line, we can render any rotation direction images. On the other hand, as the location of the viewing point is away from the camera running line, the limit of rotation angle is more restricted. In the Fig.4, the novel images from point P' can be rendered in the range as within arrows shown in the figure. Lastly, upon rendering the view from the position P, we can render a novel view at only one rotation angle. Although there is a difference concerning the flexibility of the rotation angle, the re-constructable region on the plane is combining two sectors as FOV is the angle at the circumference (shaded area in Fig.4).

Considering the actual use of this method as driving simulator, the required re-constructable area is usually limited near the camera's running lane. So, if the length of the image sampling section (from C_0 to C_n) became large, the rotation angle increases close to 360 degrees.

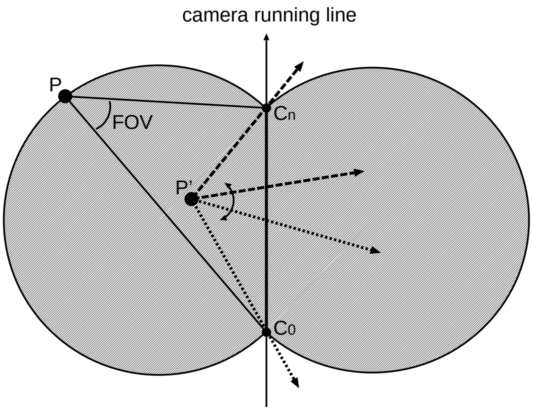


Figure 4: Reconstructable area on the ground

3.2 Singular Direction

With respect to reconstruction, we can classify a novel view into two cases: with or without a singular direction. Here we define the singular direction along the direction parallel to the running direction. Namely, the first case includes the ray parallel to the singular direction, while the second case does not. When reconstructing a view toward the moving direction from the driver's seat, we have to consider the first case, while for a side view, we consider the second case.

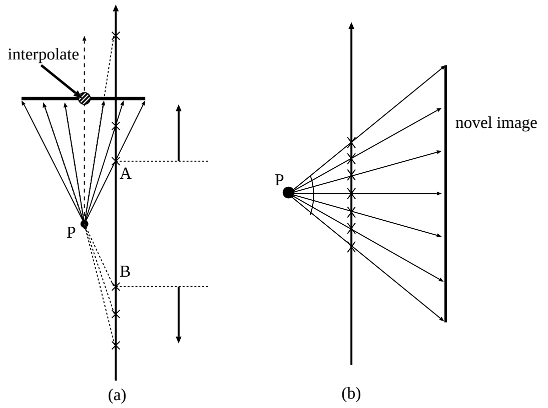


Figure 5: (a):With a singular direction (b):Without a singular direction

First, we will discuss the case with a singular direction. In this case, as shown in Fig.5-(a), we put together some slits from two parts of non-adjacent panoramic images. The right-hand side of the image comes from those ahead of the point A and the left-hand side of the image comes from those in back of the point B. Due to this discontinuity, there may be a distortion in the resulting image across the singular direction.

In addition, a ray parallel along the singular direction does not exist except moving for an infinite distance, so we have to interpolate this ray. In this paper we use the closest ray as substitution, because usually along the singular direction, no object or very distant objects exist in the ITS applications and this fact resulted in the fact that the ray along the singular direction is almost the same with small translation.

In the none singular case, a view to be rendered does not include the ray parallel to the singular direction. As shown Fig.5-(b), in this case, all necessary rays are contained in the series of panoramic images. Moreover, the rays are included in a continuous part; there is no distortion in this case, as contrasted to the singular case.

3.3 Vertical Distortion

Just as the concentric mosaics[14], this system has the effect of the vertical distortion. Also if the viewing direction is near the singular direction, the vertical distortion becomes significant because the distance between the capturing camera and the view point is large.

We can devise several methods to reduce this vertical distortion. If we know the distances between the camera optical center and the points in the scene, full perspective correction based on distances can be done. A novel view is rendered using optical flow. This method, however, requires a knowledge of geometry in the real scenes.

In many real scenes in the ITS application domain, it is reasonable to approximate that pixels along a vertical line have the same depth such as walls of buildings. In this case, we only need to estimate a depth value along each vertical line; the EPI analysis which can estimate the depth of vertical slit is most suitable. And for this purpose, “dynamic EPI” analysis [16] can be effectively used compared to usual EPI analysis because usual EPI analysis is done by static image analysis, while the dynamic EPI analysis uses the motion vector on the EPI plane. And as a result of this method, we can retrieve the depth value robustly and easily. Using this estimated value, we can scale the whole line uniformly.

4 Experimental Results

4.1 Simulation

We have implemented this system and experimented using panoramic images for simulation. Fig.6 is one of the panoramic images created for simulation. Fig.7, shows a sequence of images viewing singular direction running along the road. Fig.8 shows novel views rotating on one point.

All these rendered images contain singular directions, but the view of the singular direction is interpolated by substituting for the slit of the panoramic image which is located at the side of the novel viewing position. And we can have images successfully rendered as shown in the figures. Although, when we see these synthesized images precisely, there still remains slight discontinuity between rendered section and interpolated section, but it can be reduced by using a morphing technique. Additionally it has a lesser effect on the practical use of the application as a driving simulator.

4.2 Experiment using real scenes

We created many novel views from captured images in real scenes.

First, we describe experimental results for omni-directional images taken by a HyperOmni camera. Fig.9-(a) is an image of our laboratory scene captured by the HyperOmni-directional camera[17]. From this camera’s characteristics, we can easily generate a perspective image from omni-directional image. Fig.9-(b) is a part of perspective image which was generated from Fig.9-(a).

By using this camera, we have captured a sequence of omni-directional images as we move along a straight line. Novel views from position A, B, and C, as depicted in Fig.10, are rendered. Their locations are not on the moving line of the camera, close to the blackboard and the bookshelf in order. Fig.11-(a), Fig.11-(b), and Fig.11-(c) are shown rendered results viewed from at the locations A, B and C, respectively. Note that the relation between the left edge of the blackboard and the right edge of the bookshelf behind the blackboard differs in those images. More precisely, in Fig.11-(a), the right edge of the bookshelf and the left edge of the blackboard are not overlapped; in Fig.11-(b), where the view point is closer to the objects than that in Fig.11-(a), the left edge of the blackboard is lapped over the right edge of the bookshelf. In the Fig.11-(c), the right edge of the bookshelf is completely occluded by the blackboard. Eventually, these three rendered images show occlusion correctly corresponding to their viewing positions, which are not on the image capturing line.

Next are results for an outdoor scene, located in the landscape of the town YOKOHAMA. This experiment uses cylindrical images which are captured by some cameras constituted cylindrically as shown in Fig.1-(b) on the car. Fig12 is one of the synthesized panoramic images. The car runs on a public road in YOKOHAMA, and captures images. At the same time, the capturing positions of those images are recorded by GPS. Then we project them to cylindrical coordinates and store cylindrical images of each capturing point.

New viewing locations are shown in Fig.13. Notice that all these images are not from the image capturing line; we really rendered all of them, and projected them to the cylindrical coordinates. Our car ran on a road along the Y_1 axis as shown in Fig.13. Fig.14 shows a series of images viewing from the virtual running line-(1), namely moving parallel to X axis. On the other hand, Fig.15 are a series of images rendering from virtual running line-(2), namely, moving parallel to Y_2 axis. In other words, this sequence of rendered images is viewed from another driving line on the road.

See Fig.15-(a). Though the edge of the white building in the middle of the image connects smoothly, the pole's shadow on the ground is rendered to a stair-like appearance. This distortion is caused by vertical distortion mentioned in Sec.**3.3**.

Furthermore, there is a discontinuity at the slit joints caused by the sampling frequency. Of course, these discontinuities would vanish as the width of each slit becomes narrow enough, in other words, we have only to capture images continually. Additionally, the further away from the rendering view point the object is, the smaller the influence of sampling discontinuity (in the same way as the vertical distortions). In the town, the building is usually far enough from the road so that these discontinuities would be weak. These experimental results show that this system is very effective, when rendering large-scale scenes, as cityscape.

5 Conclusion and Future Work

This paper describes a new method for reconstructing a novel view in large-scale scenes, for example, in cities or towns. Mosaicing panoramic images captured by an omni-directional camera or a measuring device of a similar type, our rendering system can create any view from an arbitrary point to an arbitrary direction on the ground in a virtual environment.

First, we capture panoramic images running along a straight line, at the same time capturing positions of each image. And the rendering process is also very simple and easy to compute. We need only select some suitable slits from stored panoramic images, and reassemble them to generate an image from a novel observation point. In other words, once images are recorded along a straight path, an arbitrary view around the path can be constructed.

The strength of our method is that we can generate any view to any direction at any position on the ground. Because it produces a large scale space, such as a virtual city, our system is outstanding. This method can also correctly render occluded objects, and their locations. We plan to develop a driving simulator of the entire city of Tokyo using this method for the ITS purpose.

References

[1] S. B. Kang and R. Szeliski, 3-D scene data recovery using omnidirectional multibaseline stereo, in *Proceedings of Computer Vision and Pattern Recognition Conf.*, pages 364-370, 1996

[2] S. Peleg and M. Ben-Ezra, Stereo panorama with a single camera, in *Proceedings of Computer Vision and Pattern Recognition Conf.*, pages 395-401, 1999

[3] Y. Yu, P. Debevec, J. Malik and T. Hawkins, Inverse Global Illumination Recovering Reflectance Models of Real Scenes from Photographs, in *Proceedings of ACM SIGGRAPH 99*, pages 215-224, 1999

[4] P. E. Debevec, C. J. Taylar, and J. Malik, Modeling and Rendering Architecture from Photographs: A hybrid geometry-and image-based aproach, in *Proceedings of ACM SIGGRAPH 96*, pages 189-198. 1996

[5] Y. Sato, M. Wheeler, and K. Ikeuchi, Object Shape and Reflectance Modeling from Observation, in *Proceedings of ACM SIGGRAPH 97*, pages 379-387. 1997

[6] A. Lippman, Movie-Maps. An Application of the Optical Videodisc to Computer Graphics, in *Proceedings of ACM SIGGRAPH '80*, pages 32-43, 1990

[7] S. E. Chen, QuickTime VR - An image-based approach to virtual enviroment navigation, in *Proceedings of ACM SIGGRAPH'95*,pages 29-38, 1995

[8] S. K. Nayar, Omnidirectional Video Camera, in *Proceedings of Computer Vision and Pattern Recognition '97*, page 482 - 488, 1997

[9] T. Naemura, H. Harashima et al., Ray-Based Creation of Photo-Realistic Virtual World, *Virtual Reality and MultiMedia (VSMM'97)*, pages 59-68, 1997

[10] E. H. Adelson and J. Bergen, The Plenoptic function and the elements of early vision, in *Computional Models of Visual Processing*,pages 3-20,MIT Press,Cambridge,MA,1991

[11] L. McMillan and G. Bishop, Plenoptic modeling: An image-based rendering system, in *Proceedings of ACM SIGGRAPH'95*,pages 39-46, 1995

[12] M. Levoy and P. Hanrahan, Light field rendering, in *Proceedings of ACM SIGGRAPH '96*,pages 31-42, 1996

[13] S. J. Gorther, R. Grzeszczuk, R. Szeliski, and M. F. Cohon, The lumigraph, in *Proceedings of ACM SIGGRAPH*,pages 43-54, 1996

[14] Heung-Yeung Shum and Li-Wei-He, Rendering with Concentric Mosaics, in *Proceedings of ACM SIGGRAPH*, pages 299-306, 1999.

[15] M. Hirose and E. Takaaki, Building a Virtual World from the Real World, in *Proceedings of International Symposium on Mixed Reality.* (1999-3). 183-197

[16] H. Kawasaki, T. Yatabe, K. Ikeuchi and M. Sakauchi, Construction of a 3d city map using epi analysis and dp matching , Asian Conference on Computer Vision2000 (2000).

[17] Y. Onoe, K. Yamazawa, H. Takemura, and N. Yokoya, Telepresence by real-time view-dependent image generation from omnidirectional video streams, in *Computer Vision and Image Understanding, Vol.71, No.2*, pages 154-165, 1998.

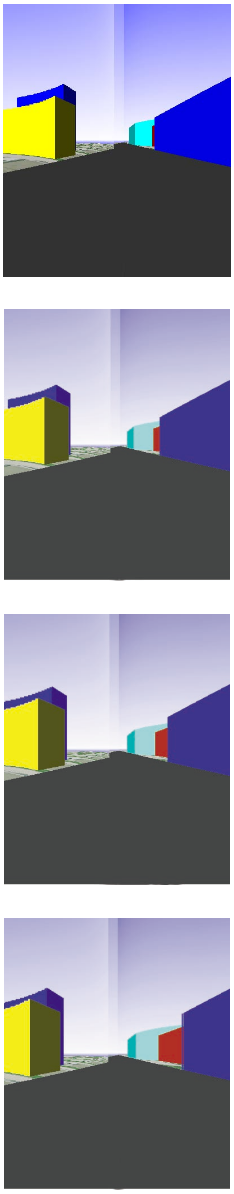


Figure 7: views moving along the road

Figure 8: views rotating on one point

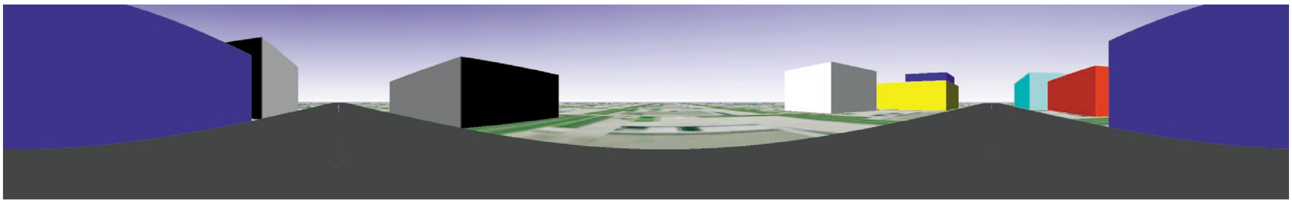


Figure 6: one of the panoramic images for simulation

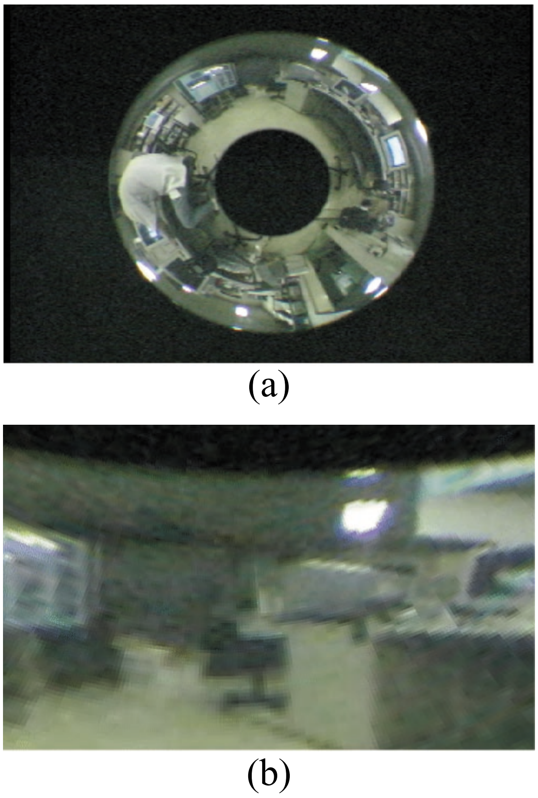


Figure 9: (a) an omni-directional image captured by a HyperOmni camera, and (b) a perspective image generated from (a).

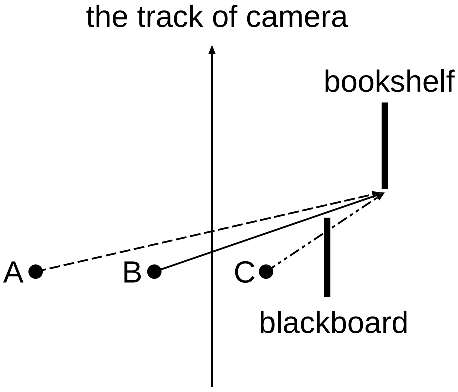


Figure 10: Rendered viewing positions (a),(b), and (c); synthesized images shown in Fig.11 are viewed from those points, respectively.

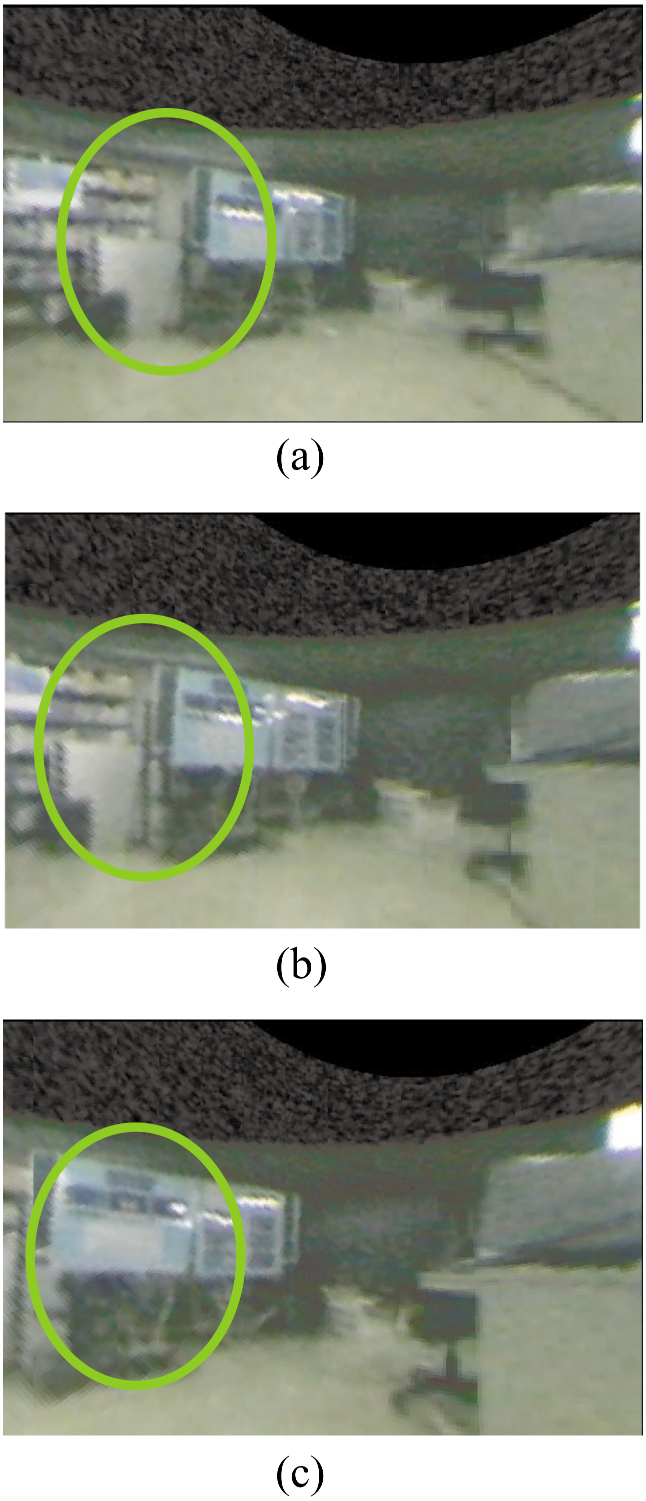


Figure 11: Rendered images; In (a), the left edge of the blackboard and the right edge of the bookshelf are separated. In (b), the left edge of the blackboard is lapped over the right edge of the bookshelf. in (c), the right edge of the bookshelf is occluded by the blackboard.

Figure 12: panoramic image in YOKOHAMA city

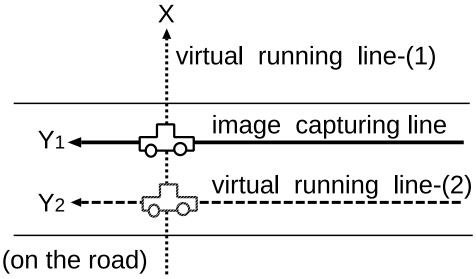


Figure 13: Y_1 axis is a driving line. A sequence of panoramic images is captured along this axis. Fig.14 show images rendered from virtual running line-(1), namely along X axis. Fig.15 show images rendered from virtual running line-(2), namely along Y_2 axis.

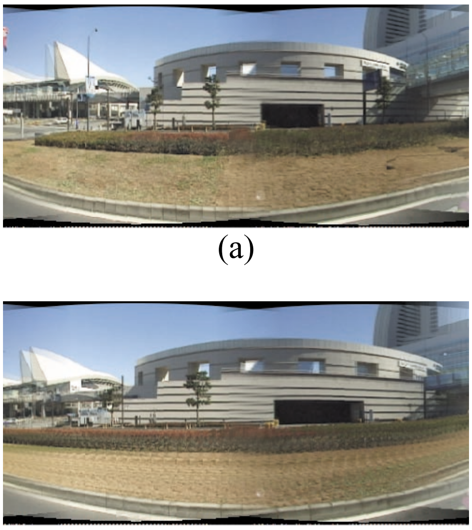


Figure 14: Rendered novel images viewed from virtual running line-(1)

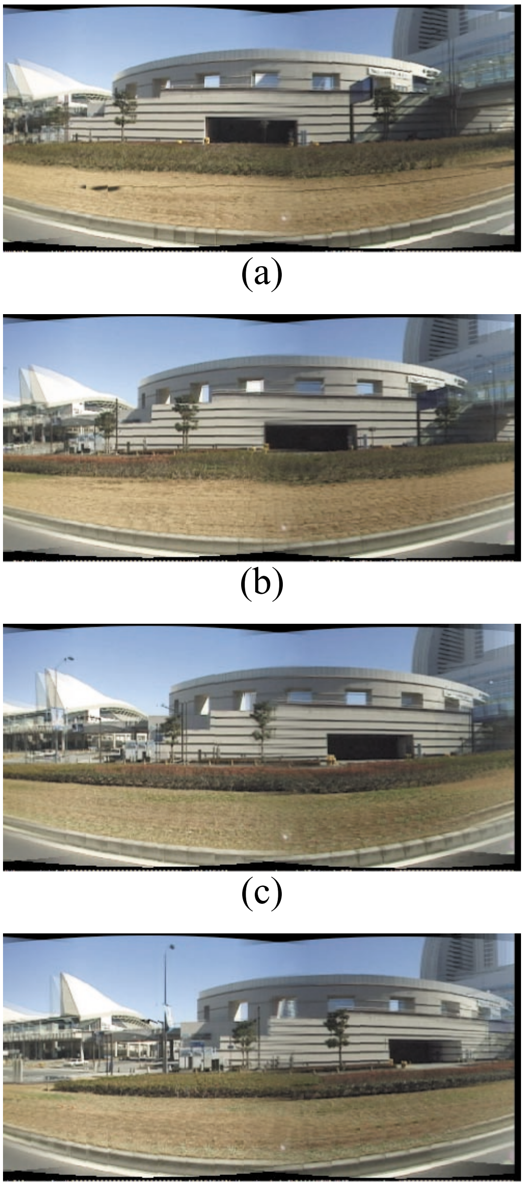


Figure 15: Rendered novel images viewed from virtual running line-(2). These are views on another driving line.