ICAT '97

Efficient Sampling of 3-D Spatial Data using a Computer-Controlled Camera Gantry

Tomokazu MURAKAMI , Takeshi NAEMURA , Masahide KANEKO† *and* Hiroshi HARASHIMA

Dept. of Inform. & Commun. Eng., The Univ. of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113, JAPAN
TEL: +81-3-3812-2111 ext.6792
FAX: +81-3-5800-5797
E-mail: murakami@hc.t.u-tokyo.ac.jp

† KDD R&D Laboratories
2-1-15 Ohara, Kamifukuoka-shi, Saitama 356, JAPAN

Abstract

For the compression and transmission of a 3-D spatial data, it is desirable that the data can be represented independently of any kind of 3-D input/output method. From this point of view, we have examined a ray-based method as a neutral approach to represent the 3-D spatial data. We need all the rays in a 3-D space to synthesize arbitrary views according to the positions of observers. However, it is not a practical approach to sample all the rays in the space. In this paper, we propose a new efficient method for capturing the 3-D spatial data as a set of light rays. First, we evaluate the difficulty in interpolating the rays at any point for any direction in the space from the previously sampled rays. Secondly, we move a camera to the position where we estimated to be most difficult in the 3-D space in the interpolation process. Thirdly, we take a picture and decompose it into a set of rays. By using the proposed method, rays can be efficiently sampled and interpolated. Experimental results show the potential applicability of the method to the 3-D image communication.

key words : 3-D image communication, light ray, sampling method, multi-view images

1 Introduction

Recently 3-D image communication technology has been discussed and studied intensively. Until now, however, it has not reached to a comprehensive level. This is because several kinds of 3-D display technologies are still making rapid progress and the 3-D data format strongly depends on the corresponding display technology. In order to promote the advancement of 3-D image communication systems, it is desirable that the 3-D data can be represented independently of any kind of input/output method.

From this point of view, we have proposed the concept of the "Integrated 3-D Visual Communication System"[1] as illustrated in Fig.1. The key feature in this new concept is a display-independent neutral representation of a 3-D spatial data. The flexibility of this concept will promote the progress of 3-D image communication systems before 3-D display technology reaches maturity. Any 3-D data format is converted into a neutral intermediate format and transmitted to receivers. At the receiver, the intermediate data is converted into 3-D images suitable for the specific display method.

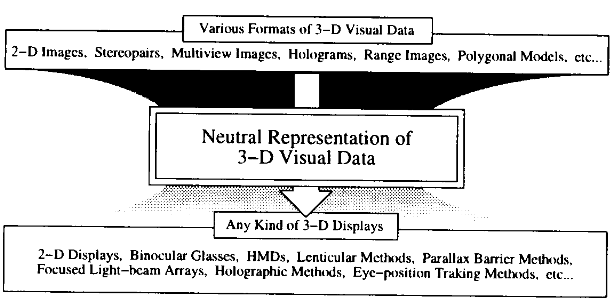


Fig.1: Basic concept of the "Integrated 3-D Visual Communication System"[4].

In order to realize the neutral representation of a

3-D data, a ray-based method has been examined[2]. By using this method, we can process a 3-D data irrespective of input/output devices and can synthesize virtual views one by one according to the positions of observers without analyzing precise structure of scene objects. However, in order to sample all the rays at once, a huge number of cameras must be aligned very closely.

In this paper, we propose a new method for capturing light rays based on the concepts of the "ray interpolation" and the "interpolation difficulty." First, we evaluate how it is difficult to interpolate rays at an intermediate viewpoint in a 3-D space from the previously sampled data. Secondly, we move a camera to the position where we estimated to be most difficult in the 3-D space in the interpolation process. Then we take a picture, and decompose it into a set of rays. Thus, the rays can be efficiently sampled and interpolated. Experimental results show the potential applicability of the proposed method to the 3-D image communication. This method enables efficient sampling of rays in a 3-D space and will contribute to the next generation 3-D communication systems.

2 Ray-based Representation of a 3-D Data

2.1 Efficient Representation of Rays

The ray-based method is originated from the idea that all the elements in a 3-D space can be visually represented by light rays emitted from these elements. A ray has such attributes as passing position, propagating direction and intensity(color). A set of light rays passing through any point in the 3-D space can be distinguished by their positions (X, Y, Z) and their directions (θ, ϕ) representing their azimuth and elevation angle of propagation. Thus, the rays can be recorded in a 5-D parametric space (X, Y, Z, θ, ϕ) which is called as "ray space"[2] or "plenoptic function"[3].

By assuming that rays go straight on without any variation in the propagating direction, we can represent all the rays in a 3-D space just by the rays intersecting a 2-D surface. In this case, the 3-D space is divided into a visual zone and a described zone by the surface as shown in Fig.2. By gathering appropriate rays on the surface, we can synthesize virtual views at any viewpoint in the visual zone. For this purpose, three coordinate systems have been proposed and examined[4]. In this paper, we utilize the planer coordinate system because its structure is simple for analysis.

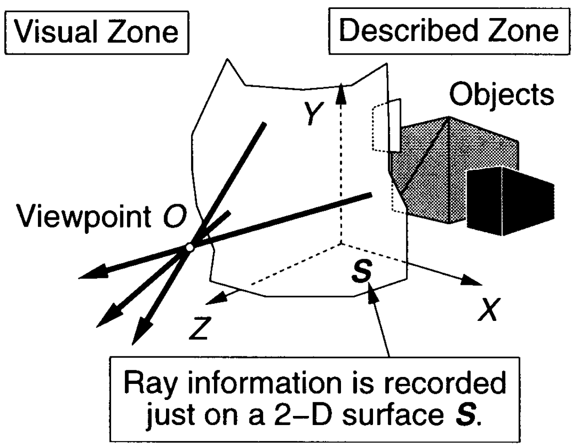


Fig.2: We can describe objects by gathering rays on a surface [4].

2.2 Geometrical Characteristics of Rays

By utilizing the planer coordinate system as a 4-D ray projection method, we can represent all the rays in a 3-D space on a reference plane[4, 5]. We set up the reference plane at the position $Z = 0$. Suppose that a ray emitted from an object at a point (X, Y, Z) and propagates in a direction (θ, ϕ), it intersects the reference plane at the position (P, Q) as shown in Fig.3. P and Q are defined by the following equation:

$$\begin{cases} P = X - Z \tan\theta \\ Q = Y - Z\frac{\tan\phi}{\cos\theta} \end{cases} \tag{1}$$

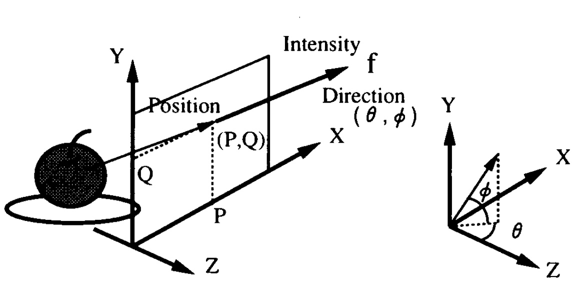


Fig.3: The intensity of a ray which passes a point (X, Y, Z) and propagates in a direction (θ, ϕ) is represented as $f(X, Y, Z, \theta, \phi)$. All the rays can be represented on a reference plane.

When we assume a 2-D image captured at a viewpoint by a pinhole camera directed perpendicularly to the reference plane, two variables (u, v) which correspond to image coordinates can be represented as $u = \tan\theta, v = \frac{\tan\phi}{\cos\theta}$. A 2-D image can be represented as $I(u, v)$. Then the following equations can be derived from Eq.(1).

$$\begin{cases} P = X - Zu \\ Q = Y - Zv \end{cases} \quad (2)$$

From the viewpoint of capturing rays, Eq.(2) indicates that when we take a picture at a position (X, Y, Z) in the 3-D space, a set of rays derived from Eq.(2) can be obtained from the (P, Q, u, v) space. On the other hand from the viewpoint of synthesizing views, we have to gather the set of rays derived from Eq.(2) for the observer whose viewpoint locates at the position (X, Y, Z). We can analyze the essential rays to synthesize the virtual views in the 3-D space by checking data on the reference plane according to Eq.(2).

3 Sampling Method of Ray Data

In order to synthesize virtual views from any viewpoint, we must represent all the rays in a 3-D space. However, since it requires a very large memory space and it is difficult to capture all the light rays in the 3-D space at once, it is not a practical way to represent all the rays. Here we propose the concepts of the "ray interpolation"[6] and the "interpolation difficulty." Several methods are examined for the interpolation of 3-D images, for example "lumigraph"[7]. The main purpose of our method is not to improve the precision of the interpolation of rays but to sample essential rays efficiently by utilizing the interpolation method.

In many cases, we will obtain a sparsely-sampled data space of rays from input images. In order to synthesize arbitrary views from this sparse data space, it must be filled up by an interpolation method. We sample some correct rays which are necessary and essential to synthesize the virtual views and interpolate other peripheral rays. As a result of interpolation, the data space will be filled with both the captured real rays and the interpolated virtual rays. Then, we can create the photo-realistic virtual world, which is represented in this data space.

The sampling method is illustrated in Fig.4. First, we evaluate the interpolation difficulty which means the difficulty in interpolation of rays at the intermediate viewpoints between the positions at which input images are captured previously. By evaluating the interpolation difficulty, we can find the best position to take a picture. Secondly, we move a camera to the position where we estimated to be most difficult in the 3-D space in the interpolation process. This position can be considered as the best camera position to gather the essential rays. Then we take a picture and decompose it into a set of rays. This method enables efficient sampling of rays in the 3-D space.

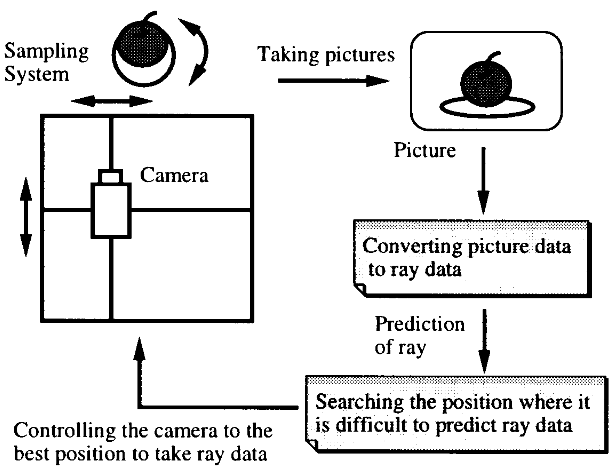


Fig.4: Sampling method utilizing the multi-view imaging system with a movable camera [6].

4 Viewpoint Search Method based on Ray Interpolation Estimation

4.1 System Overview

Figure 5 illustrates an overview of the sampling system. We use a camera gantry[8] to take pictures of scene objects on the front stage and the camera can be controlled to arbitrary positions in this gantry. Work station 1 (WS1) controls the camera and sends pictures to Work station 2 (WS2). WS2 predicts the most appropriate viewpoint to obtain the essential rays from input pictures and sends position data of the camera to WS1. By repeating the same process, we can sample rays in the 3-D space efficiently. WS1 and WS2 cooperate each other by the inter-process communication mechanism.

The outlook of the camera gantry of the sampling system is shown in Fig.6. We arranged several houses and buildings made by plastic blocks behind a white polystyrene wall. We can see the objects through the crevice of the wall. Two pictures are taken as the initial inputs at the left and right edges of the view area. Then we estimate the interpolation difficulty in the 3-D space from these two pictures, find the next position to take a picture and move the camera to the position.

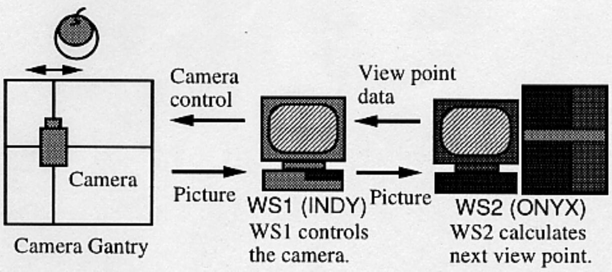


Fig.5: Overview of the ray sampling system using a computer-controlled camera gantry.

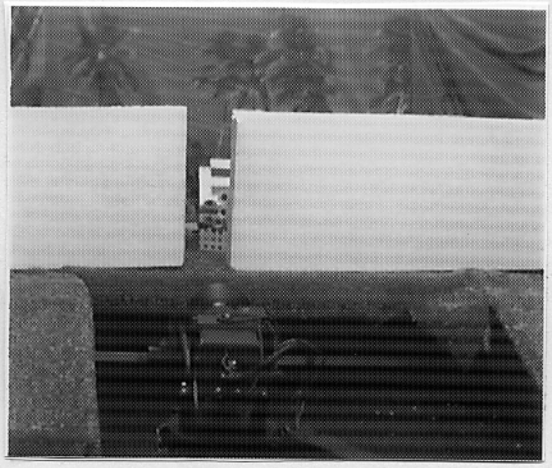


Fig.6: Outlook of the camera gantry.

4.2 Ray Interpolation Estimation

In order to synthesize the virtual view from every viewpoint in the 3-D space, we must sample rays which is necessary to interpolate peripheral rays. According to the theory of ray based representation, when the same position on an object is seen from two or more viewpoints, the rays emitted from that position can be interpolated and the virtual views between these viewpoints can be synthesized. Therefore we must control the camera to the positions where every part of the object can be seen in the captured picture from two or more positions. However, it is difficult to find the corresponding point on the object in other pictures. Even if a part of the object can be seen from two positions, another part of the same object being occluded by other objects may not be seen from one of the positions. When we find correspondences between the input pictures and interpolate rays by scanning color of the pixels in the input pictures, it may cause some errors in comparison with the precise geometric structure of the objects. However, from the viewpoint of synthesizing virtual views, the rays interpolated in this way resemble neighboring rays in color. And it is supposed that these errors do not stand out in the synthesized virtual views. Then, when we predict and interpolate rays from sparsely sampled rays, it is easy to interpolate the object which is flat and painted in one color and it is difficult to interpolate the object which is occluded by other objects and can not be seen in the previously captured pictures.

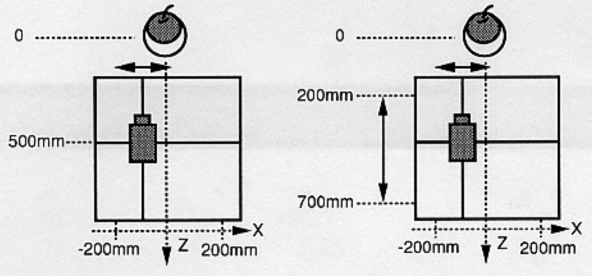
Then we evaluate the ray interpolation difficulty in the ray space. In order to sample rays in the 3-D space efficiently, we have to search the positions where it is difficult to interpolate rays and control the camera to such positions with higher priority. We check every possible ray in the ray space by scanning color of pixels in the input pictures, calculate the interpolation difficulty at every viewpoint and evaluate the positions where we can sample most rays which we estimate to be difficult to interpolate.

As for the scene objects, the white wall in front of the houses and buildings is flat and white at every part. Therefore, the rays emitted from the wall is easy to be interpolated. On the other hand the rays of objects which can be seen through the crevice of the wall is difficult to interpolate because they are occluded by the wall. Accordingly, the camera must be controlled to the front of the crevice and take pictures of the houses and buildings behind the wall.

We attempted two types of experiments. First, we limit the movable area of the camera to $(-200mm \leq X \leq 200mm, Z = 500mm)$ as shown in Fig.7(a) to simplify the analysis of the ray space. Secondly, we expanded the movable area to $(-200mm \leq X \leq 200mm, 200mm \leq Z \leq 700mm)$ as shown in Fig.7(b). In both experiments we estimated the interpolation difficulty at every viewpoint and searched the best position to sample rays in the 3-D space.

4.3 Viewpoint Search Method when the Camera Moves along a Line

For the first experiment, we utilise the concept of epipolar-plane images[9]. When we take pictures by cameras arranged in a line, we can construct an



(a) (b)

Fig.7: Movable area of the camera.
(a) For the first experiment.
(b) For the second experiment.

epipolar-plane image as shown in Fig.8. We estimate the ray interpolation difficulty in an epipolar-plane image as shown in Fig.9. The ray interpolation difficulty Q_s is calculated as follows.

$$Q_s(x) = \min_{dx \in D}\{s|I_0(x - sdx) - I_1(x - (1-s)dx)|, (1-s)|I_0(x - sdx) - I_1(x - (1-s)dx)|\} \quad (3)$$

Q_s indicates the ray interpolation difficulty at a view point s. I_0 and I_1 denote each of the input pictures. x denotes a position of pixel on a virtual picture I_s and dx denotes a disparity. In Fig.9, the vertical axis s relates to the position of a virtual picture I_s and the horizontal axis x relates to the position of pixel on a picture I_s.

We check all correspondences between the two input pictures by pixel-to-pixel matching and draw lines on an epipolar-plane image by the remainder value between the corresponding pixels. We plot the remainder between the corresponding pixels of the input pictures as a line in the epipolar-plane image. At the crossing point of the lines, smaller value is recorded on the image. By searching the minimum value of the remainder values for every disparity dx in a range D, we can estimate the ray interpolation difficulty at every viewpoint when a $I_s(x)$ was synthesized by an appropriate method. These processes mean checking every possible virtual ray when it is appropriately interpolated between the previously sampled rays. Because rays can be assumed as lines in the epipolar-plane image according to Eq.(2). Suppose a pixel at x on the picture I_s is interpolated from I_0 and I_1, when the remainder of intensity between $I_0(x-sdx)$ and $I_1(x-(1-s)dx)$ was small for disparity dx, the picture $I_s(x)$ is supposed to be interpolated in higher quality. $s, (1-s)$ are multiplied to the remainder between $I_0(x - sdx)$ and $I_1(x - (1-s)dx)$ to reset the ray interpolation difficulty to 0 at the positions at which the pictures were previously taken. In this way, an epipolar-plane image is filled by dots and the brightness of each dot indicates the ray interpolation difficulty at the point.

Then we check the sum of the brightness of the dots for each horizontal line in the image as the interpolation difficulty of each viewpoint. The larger Q_s is, the more difficult it is to interpolate rays at the position s. The position where Q_s becomes maximum is supposed to be the most suitable point to capture rays that are difficult to interpolate. Then we control the camera to the point and take a picture and recalculate Q_s again by using all input data. By repeating the same process and following the maximum point of Q_s successively, we can sample essential rays efficiently.

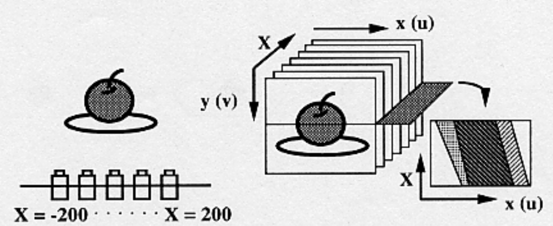


Fig.8: Pictures taken by cameras aligned mutually parallel compose an epipolar-plane image.

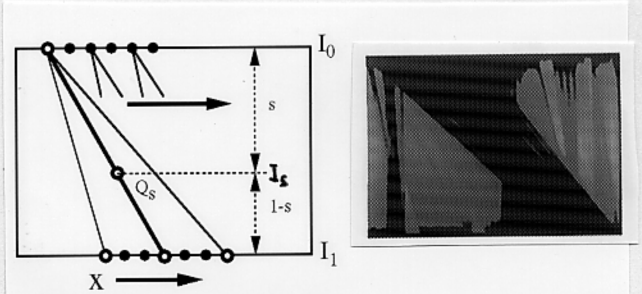


Fig.9: By checking correspondences between the two input pictures in a epipolar-plane image, we can estimate the ray interpolation difficulty at intermediate viewpoints.

4.4 View Point Search Method when the Camera Moves on a Plane

For the second experiment, we use a P-u plane derived from Eq.(2) for estimating the ray interpolation difficulty instead of an epipolar-plane image. Because the camera moves along the Z axis in the second experiment, the rays captured by the camera cannot be represented fully in an epipolar-plane image. We can project the rays in a P-u plane according to Eq.(2) when the camera moves freely on a plane to the directions X and Z. However, owing to the vertical disparity of ϕ, not all the rays can be projected in a P-u plane precisely. Suppose a picture captured at a position (X, Z) and a horizontal scanning line v of the picture, Q is determined by v and Z according to Eq.(2). Therefore, we set up a P-u plane for each Q value and estimate the ray interpolation difficulty in a P-u plane as shown in Fig.10. This method is an approximated one but it does not matter for a viewpoint search method.

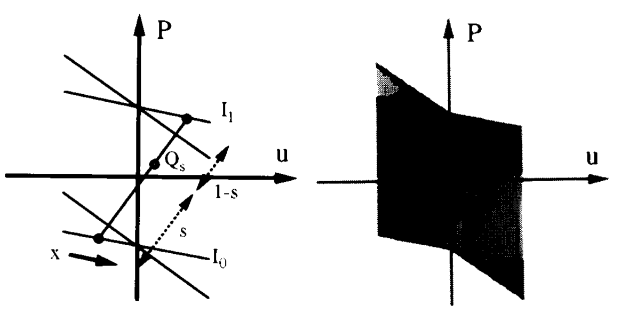


Fig.10: We project the lines of the input pictures to a P-u plane, check correspondences between them and estimate the ray interpolation difficulty at every viewpoint.

Ray interpolation difficulty Q_s is calculated by Eq.(3) in the same way for an epipolar-plane image. In this case, each of the input pictures is projected to a P-u plane as a line. We check all correspondences between the input pictures to check every possible interpolated ray between previously sampled rays in the same way with the previous experiment. We plot the remainder value between the corresponding pixels of the input pictures in the P-u plane. Then we analyze the crossing and overlapping of the lines in the P-u plane and calculate the sum of the remainder value at every viewpoint as the interpolation difficulty. Q_s is estimated for each viewpoint that corresponds to a line in the P-u plane.

The maximum point of Q_s is supposed to be the most suitable point to capture rays that are difficult to interpolate. Then we control the camera to the point and take a picture and recalculate Q_s again by using all input data. By repeating the same process and following the maximum point of Q_s, we can sample essential rays efficiently.

5 Experimental Results

5.1 Experiment 1 (A Case where Camera Moves along a Line)

For the first experiment, we limited the movable area of the camera to $(-200 \leq X \leq 200, Z = 500)$ to simplify the analysis of the ray space. The camera took pictures at $(X, Z) = (-200, 500), (200, 500)$ as the initial inputs. Then we estimated the ray interpolation difficulty Q_s between the two viewpoints. Q_s at each viewpoint is shown in Fig.11. The vertical and horizontal axes correspond to Q_s and the position of viewpoint, respectively. The maximum point of Q_s is supposed to be the most difficult viewpoint to interpolate rays. Since it takes too much time to calculate Q_s for all the viewpoints at intervals of $1mm$, we checked Q_s at several viewpoints which were selected appropriately by the Coarse-to-Fine method and searched the maximum point of Q_s. The circles, squares and triangles in Fig.11 indicate the sampled points.

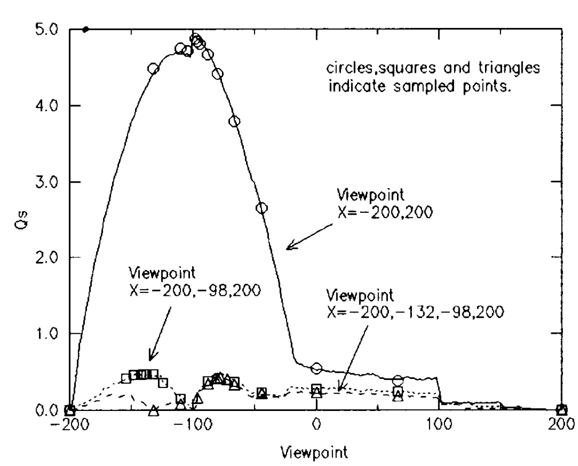
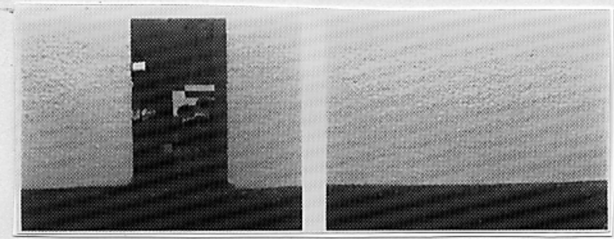


Fig.11: The ray interpolation difficulty Q_s at each viewpoint in $(-200 \leq X \leq 200, Z = 500)$. The circles, squares and triangles indicate the sampled points. Q_s is maximum at $X = -98$ for the initial inputs.

In order to capture rays efficiently, the camera must be controlled to the peak point of this graph successively and the camera moved to $X = -98$. The pictures captured at position $X = -200, 200, -98$ are shown in Figs.12(a),(b) and 13. The rays emit from the white polystyrene wall can be interpolated easily. Thus the camera moved near the crevice and took pictures of the houses and buildings behind the wall that are difficult to be interpolated.

Then we recalculated Q_s again by using all input data at $X = -200, 200, -98$. The graph of Q_s is shown in Fig.11 as a dotted line. Value of the ray interpolation difficulty becomes lower as the input data increases. By the same process as before, the next viewpoint was determined to $X = -132$. It is supposed that the camera was moved to capture essential rays which are necessary to interpolate other rays with higher priority. Repeating these processes, we sampled rays in 3-D space efficiently.



(a) (b)

Fig.12: Initial input pictures.
(a) Viewed at $(-200, 500)$.
(b) Viewed at $(200, 500)$.

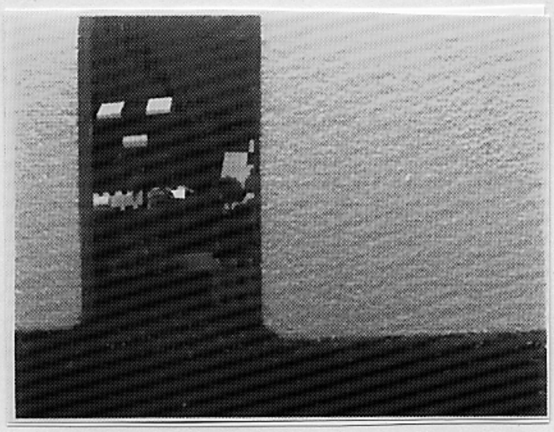


Fig.13: Selected view point at $(-98, 500)$.

5.2 Experiment 2 (A Case where Camera Moves on a Plane)

For the second experiment, we expanded the movable area to $(-200mm \leq X \leq 200mm, 200mm \leq Z \leq 700mm)$. For the initial input images, we moved the camera and took pictures at $(X, Z) = (-200, 200), (200, 200), (-200, 700), (200, 700)$, which are shown in Figs.15(a),(b),(c) and (d). Then we estimated the ray interpolation difficulty Q_s in the (X, Z) plane. Q_s at each viewpoint is shown in Fig.14. The bottom square in Fig.14 relates to the position (X, Z) of the camera and the vertical axis corresponds to Q_s. In the same way with the previous experiment, the camera is controlled to the peak point of this graph. Q_s is maximum at $(-130, 250)$ in this graph and the picture at this point is shown in Fig.16.

Likewise the first experiment, the camera was moved near the crevice and took pictures of the houses and buildings behind the wall. It is supposed that the camera was moved to collect essential rays to synthesize the virtual views in the visual zone.

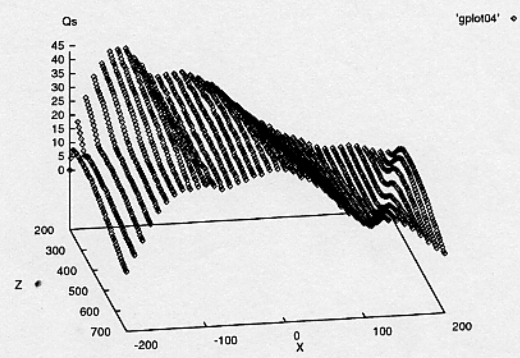
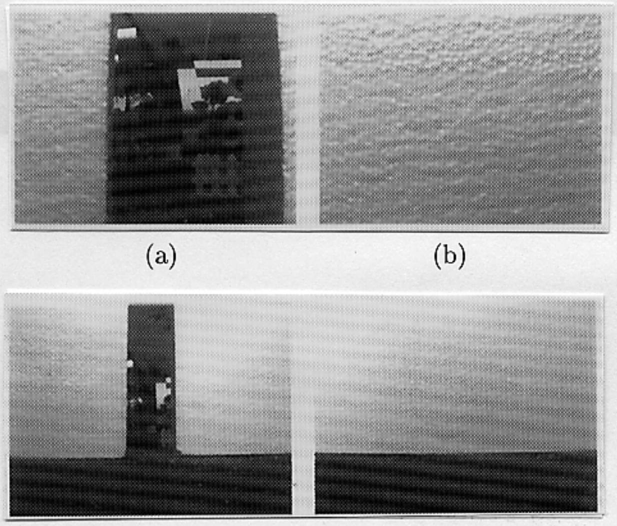


Fig.14: Ray interpolation difficulty Q_s at each viewpoint in $(-200 \leq X \leq 200, 200 \leq Z \leq 700)$. Q_s is maximum at $X = -130, Z = 250$ for the initial inputs.

In both of the experiments, the camera moved in front of the crevice of the wall and took picture of the objects behind the wall. It indicates that the camera was controlled to sample rays which are difficult to interpolate and by repeating this process, we can sample essential ray data efficiently. These experiments shows that the proposed viewpoint search method is useful.

6 Conclusion

In this paper, we examined a method of efficient sampling of rays by evaluating the interpolation dif-



(a) (b)

(c) (d)

Fig.15: Initial input pictures.
Viewed at (a)$(-200, 200)$, (b)$(200, 200)$, (c)$(-200, 700)$, (d)$(200, 700)$.

Fig.16: Selected viewpoint at $(-130, -250)$.

ficulty and applied the developed software of this method to an actual system composed of a camera gantry and work stations. Owing to the concept of the ray interpolation, we can synthesize a virtual view at any viewpoint from sampled rays. The proposed method provides the efficient way to sample rays enough to synthesize the virtual view from an arbitrary viewpoint in the 3-D space.

In the virtual reality systems, to present photorealistic 3-D graphics to an observer is one of the most interesting topics. The proposed method in this paper will help to handle the 3-D information easily and contribute to the next generation 3-D communication and virtual reality systems.

References

[1] H. Harashima, T.Fujii and T. Naemura : "Towards integrated 3-D visual communication systems," *Technical Report of IEICE*, HC92-58 (Jan. 1993). (*in Japanese*)

[2] T. Fujii : "A basic study on the integrated 3-D visual communication", Ph.D thesis of engineering, The Univ. of Tokyo (Dec. 1994). (*in Japanese*)

[3] E. Adelson and J. Bergen : "The plenoptic function and the elements of early vision", in M. Landy and J. Movshon, ed., *Computer Models of Visual Processing*, MIT Press (1991).

[4] T. Naemura, T. Yanagisawa, M. Kaneko and H. Harashima : "Efficient projection of ray space for a ray based description of 3-D real world," *Technical Report of IEICE*, IE95-119 (Feb. 1996). (*in Japanese*)

[5] T. Fujii, E. Kitou, M. Kaneko and H. Harashima : "Transformation of hologram and ray space for integrated 3D visual communication," *3D Image Conference '95*, 8-9, pp. 220-225 (July 1995). (*in Japanese*)

[6] T. Murakami, T. Naemura, M. Kaneko and H. Harashima : "Efficient Sampling of Ray Space Based on Ambiguity Estimation," *Proc. 1996 IEICE General Conference*, D-417 (March 1996).(*in Japanese*)

[7] S. Gortler, R. Grzeszczuk, R. Szeliski and M. Cohen : "The lumigraph", *Proc. ACM SIGGRAPH'96*, pp.43-54 (Aug. 1996).

[8] A. Ishikawa, T. Naemura, M. Kaneko and H. Harashima : "Development of movable monocular camera system for acquiring ray space information," *Proc. 1996 Information and System Society Conference of IEICE*, D-300, p.303 (Sep. 1996).(*in Japanese*)

[9] H.H.Baker and R.C.Bolles : "Generalizing epipolar-plane image analysis on the spatiotemporal surface," *International Journal of Computer Vision* 3, pp.33-49 (1989).

Dept. of Inform. & Commun. Eng., The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113, JAPAN
TEL: +81-3-3812-2111 ext.6792
FAX: +81-3-5800-5797
E-mail: murakami@hc.t.u-tokyo.ac.jp
URL: http://www.hc.t.u-tokyo.ac.jp/~murakami/