ICAT 2002
December 4-6, Tokyo, JAPAN

Issues in an Image Capture System for TWISTER

Junya Hayashi, Kenji Tanaka, Masahiko Inami, Dairoku Sekiguchi, Naoki Kawakami, Susumu Tachi

Graduate School of Information Science and Technology, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, JAPAN
{junya, tanaken, minami, dairoku, kawakami, tachi}@star.t.u-tokyo.ac.jp

Abstract

In this paper we report the development of TWISTER [3][8] (Telexistence Wide-angle Immersive STEReoscope), which realize the concept of "Mutual Telexistence". In contrast to conventional telexistence, where people in remote places can be appeared with a realistic sense of presence, in mutual telexistence, the figure of the observer him/herself can be displayed to remote users and all of the users can communicate as if they were in the same three-dimensional space. TWISTER is a cylindrical booth designed for mutual telexistence. It enables the user to view immersive and autostereoscopic images, and at the same time, it captures and reconstructs the image of the user inside with cameras positioned in-between the display unit. In previous reports on the development of TWISTER, we have introduced the development of display part [1]. In this report we discuss some methods to capture images of the user and render them in real time.

Key words: Mutual Telexistence, Virtual Reality, Multiple Cameras, Image Based Rendering

1. Introduction

There are a number of researches about face-to-face communication systems among the people in distant locations [1]. Some of them adopted displays, which can provide immersive stereoscopic images. These include IPT (Immersive Projection Technology) [2][4][5] and HMD (Head Mounted Display) based systems. These systems require special eyewear that hides the users' faces and at the same time makes it difficult to shoot them. In contrast to these conventional systems, where the images of people in remote locations can be seen only to the user, in mutual telexistence, the figure of the observer him/herself can also be displayed to remote users and all of the users can communicate as if they were in the same three-dimensional space (Fig. 1). We have designed TWISTER especially for this feature of mutual telexistence. It displays three-dimensional images of other people to a user inside the booth, while capturing the image of the user.

In order to realize the display of three-dimensional images without special eyewear, "Autostereopsis" in other words, we have adopted a "Rotating Parallax Barrier" method [8]. The current prototype, TWISTER III [19], can display immersive full-color stereoscopic motion pictures at the frame rate of 30fps (frames per second). We can also put image sensors in-between display units, because the gaps between two display units, consisted of only two LED arrays, are wide enough. In the future we are going to put 30 cameras shooting inwards on the corner of the cylinder.

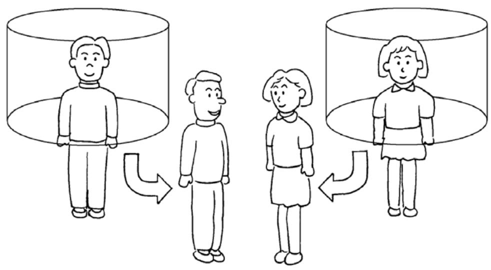


Fig 1: The Concept of TWISTER

Because one of the major elements for the sense of presence is the motion parallax, we have to design the system to be able to generate the images of the user inside the booth according to the user's eye position. We also have to generate the image in real time using limited number of cameras. We have already shown and experimented the method of obtaining the images from desired viewpoint with multiple cameras arranged on a line. The approach is based on an IBR (Image Based Rendering) method. By selecting necessary scan line in advance using analog circuit, we have achieved the real time image generation [9]

In this paper, we extend this method to an image capture system with circumferentially arranged cameras, based on discussion and experiments, and show the method to obtain images seen from arbitrary viewer's position of two degrees of freedom in one horizontal plane.

2. Background

As a prototype experiment of the image capture system for TWISTER, Kunita has built the system with 12 cameras arranged on a linear actuator. In this experiment, he generated the images seen from an arbitrary virtual viewpoint in real time. He selected the necessary light information from 12 cameras by scanning line, before digital processing by PC. Since necessary image data is selected in early stage, the real-time rendering of images from arbitrary points has been realized.

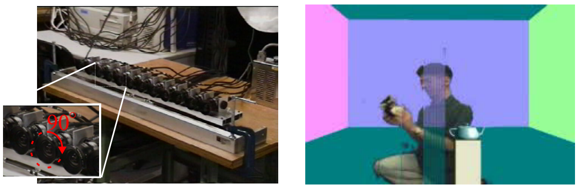


Fig 2. Image Capture Prototype System

With TWISTER, we have to extend this method in order to fit the cylindrical arrangement of the cameras of TWISTER.

3. Principle

Considering a certain closed surface, if we can obtain the information of all light rays from all the points on the surface, we can reconstruct the image of inside the closed surface seen from arbitrary viewpoints by assuming the going-straight nature of light and no attenuation of light. We will try to reconstruct the images of inside the TWISTER using this principle by arranging multiple cameras looking inside from the cylindrical surface.

First, we consider the case where the infinite number of cameras is everywhere located horizontally and perpendicularly on the cylinder. We can reconstruct the images seen from arbitrary viewpoints by using the principle directly. Next, considering the state where the infinite number of cameras is horizontally located. There, the picture seen from a virtual viewpoint, which is in the same level surface as cameras, can be reconstructed in approximation. However, when the object position is not known correctly, since there is an estimate error of the depth direction, a perpendicular error comes out on reconstructed images.

Considering TWISTER, we plan to put limited number of cameras on the cylindrical device horizontally. As the present TWISTER is not equipped with the mechanism for measuring the object’s position and form, it is necessary to assume an object’s geometric form. And the estimate error of the depth distance to the object will come out both horizontally and perpendicularly in reconstructed images. Therefore, it is necessary to consider how to assume object's position and shape, and from which camera we should get the surface information of the assumed object. From now on, we will call the assumed position and shape of an object “Virtual Screen”, as it is a projection plane of each camera’s output. We will suggest how to decide the virtual screen, and how to choose camera projecting on the virtual screen.

4. Theory

4-1. Issues

Limited number of cameras is arranged horizontally at equal intervals on the wall of the cylindrical booth, and they take the images of an object in the booth. Now, we will consider the method of reconstructing the picture of the object seen from an arbitrary viewpoint, which is outside the booth and on the same plane as the cameras.

At this time, position and shape of the object is not known, and we have to reconstruct the desired image as sufficient accuracy as possible. Specifically, as described in Section 3., the problems are (1) how to set a virtual screen, and (2) how to choose camera output for the texture of the virtual screen.

In addition, since the object’s position is restricted in the cylinder, the area where the object can exist is limited and we don’t need to think about background images. This is because we plan to put the captured object on a virtual environment, in which the background in actual environment does not have a meaning.

4-2. How to choose camera for each light ray

When an image capture system is composed of a limited number of cameras, it is necessary to approximate the incident light to a virtual viewpoint with the outputs of each camera. At this time, for each incident light to the virtual viewpoint, it is advantageous to choose the camera so that the angle of the incident light vector to the virtual viewpoint and the actual light ray to the camera can be as small as possible. The reason is that by choosing the camera in this way, we can minimize the error effect to the virtual screen, and we can also reconstruct the object’s color with fine accuracy including the reflective characteristic.

Fig. 3 shows how to choose camera to approximate the incident light ray to the virtual viewpoint. The positions of two adjacent cameras are C_1 and C_2, the position of a virtual viewpoint is P, and the intersection of the light ray and the virtual screen is Q in the figure.

We can divide space to some areas by choosing the camera that approximates each light ray to virtual viewpoint. The boundary of these areas is described as **the locus of Q, where PQ is the bisector of $\angle C_1C_2Q$.** We will obtain this locus here.

We set the position of 2 cameras $C_1(-1, 0)$ and $C_2(1, 0)$, the position of virtual viewpoint $P(p, q)$, and the intersection of the light ray and the virtual screen $Q(x, y)$. When Q is on the boundary of the areas, $\angle C_1QP$ is equal to $\angle C_2QP$. Therefore by using complex plane method,

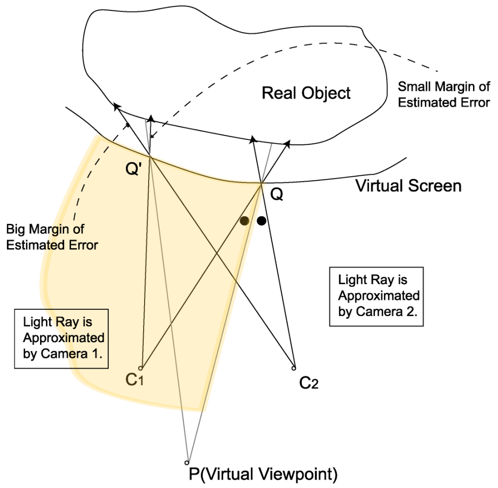


Fig. 3 Camera Selection for Approximation

$$\arg\left(\frac{(x+yi)-(q+pi)}{(x+yi)-(-1)}\right) = \arg\left(\frac{(x+yi)-1}{(x+yi)-(q+pi)}\right) \quad (1)$$

It can be described as

$$\mathrm{Re}\left[\left(\frac{(x+yi)-(q+pi)}{(x+yi)-(-1)}\right)\Big/\left(\frac{(x+yi)-1}{(x+yi)-(q+pi)}\right)\right] > 0$$
$$\mathrm{Im}\left[\left(\frac{(x+yi)-(q+pi)}{(x+yi)-(-1)}\right)\Big/\left(\frac{(x+yi)-1}{(x+yi)-(q+pi)}\right)\right] = 0 \quad (2)$$

And it can be solved as below.

$$\begin{aligned} f(x,y) &= -qx^3 + px^2y - qxy^2 + py^3 \\ &+pqx^2 - pqy^2 + (q^2 - p^2 - 1)xy \\ &+qx + py - pq = 0 \end{aligned} \quad (3)$$

$$\begin{aligned} g(x,y) &= x^4 + 2x^2y^2 + y^4 \\ &-2px^3 - 2qx^2y - 2pxy^2 - 2qy^3 \\ &+(p^2 - q^2 - 1)x^2 - 4pqxy - (p^2 - q^2 - 1)y^2 \\ &+2px - 2qy - p^2 + x^2 > 0 \end{aligned} \quad (4)$$

(In addition, when f (x, y)=0 and g (x, y)<0, PQ is the bisector of an exterior angle of $\angle C_1QC_2$.)

Fig. 4 shows the locus of Equation (3) when (p, q) is (-20, -20), (-20, -5), (-5, -20) and (-5, -5). Two circles in the figure are described as reference of the body of TWISTER and the safety shield (objects can only exists inside the shield). The radius of TWISTER's body and the distance from each camera to the center of TWISTER are both set to 600mm and the distance between adjacent cameras is set to 200mm. The dashed lines in the figure represent the straight line PM (M is the middle point of C_1C_2).

From the figure, we see that the obtained locus approaches to the straight line PM as the distance from cameras to Q becomes large, and the angle between PM and C_1C_2 approaches 90 degrees. Besides, although it is

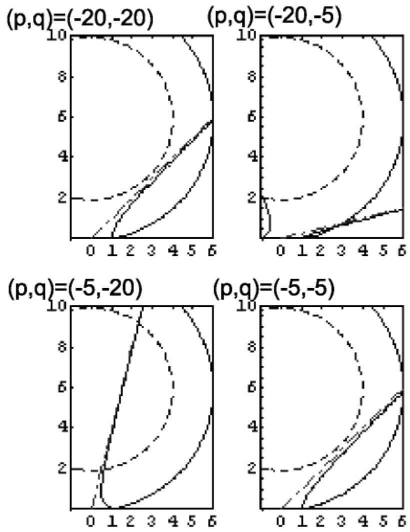


Fig. 4 The locus of Point Q

an obvious solution, in the case of the angle between PM and C_1C_2 is equal to 90 degrees, the locus corresponds to the straight line PM.

Apply to the TWISTER System

Considering TWISTER system, since the object is restricted to the inside of TWISTER, the angle between PC and C_1C_2 is close to 90 degrees, and the distance between adjacent cameras can be set small enough, compared to distance between the camera and the object. Thus, in the case of our image capture system we can approximate the boundary of areas, where each camera approximates the light, by straight line PM.

By using the straight line PM as the boundary of these areas, we can choose the camera, which approximate each light ray to the virtual viewpoint, without considering the position and the shape of the virtual screen (Fig. 5).

4-3. Error Evaluation of Virtual Screen

When we reconstruct the image of the object from the arbitrary viewpoint based on the outputs from a limited number of cameras, the more the virtual screen approximates the actual object well, the more an exact picture can be obtained. However, if there is an error between the object's position and the virtual screen, an error appears in output images. We will evaluate the error here.

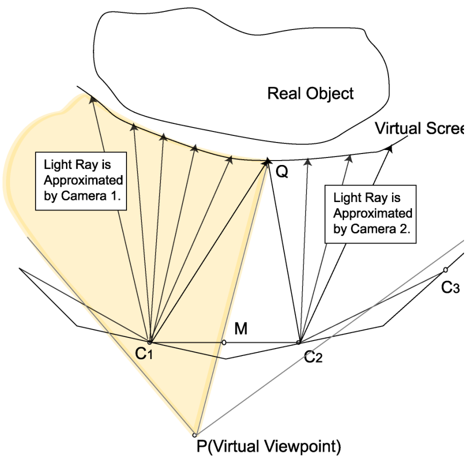


Fig. 5 Areas where each camera approximates light rays

As shown in Fig. 6, the position of the camera is C, the position of the virtual viewpoint is P, one of the points on the virtual screen projected by the camera Q, the angle between PQ and CQ is θ, and the error between the position of the object and the virtual screen is δ.

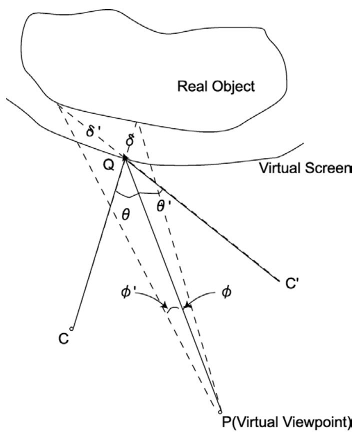


Fig. 6 Modeling Error

At this time, the amount of the angle direction error ϕ is described as below.

$$PQ\sin\phi = \delta\sin(\theta-\phi) \quad (5)$$

It is be solved as,

$$\phi = \tan^{-1}\left(\frac{\gamma\sin\theta}{1+\gamma\cos\theta}\right) \quad (6)$$

where $\gamma = \frac{\delta}{PQ}$, and γ is the relative length of the error compared to the distance between the virtual viewpoint and the virtual screen. Fig. 7 shows the graph of

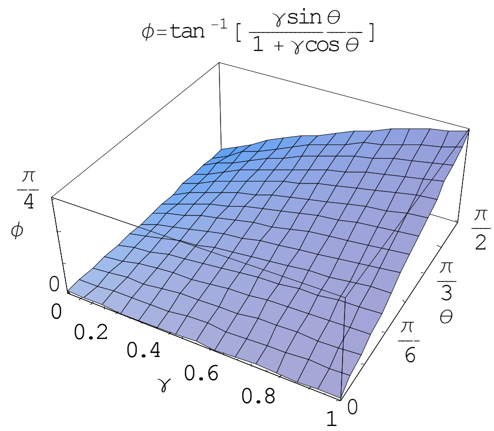


Fig. 7 Angle Error with Approximation of Virtual Screen

equation (6). From this graph, it turns out that the error can be made small by making the virtual screen close to the object or making the angle between PQ and CQ as small as possible.

From now on, Taylor series expansion of the equation (6) is used to analyze in detail,

$$\phi = \gamma\sin\theta - \gamma^2\sin\theta\cos\theta + o[\gamma^3] \quad (7)$$

Here, if θ is small enough and γ is the order of 0.1 or so, the error ϕ of the angle direction is

$$\phi = \gamma\sin\theta \quad (8)$$

Moreover, if we reconstruct the image of the object, with 45 degrees of field angle, seen from the virtual viewpoint and output it on a VGA (640 x 480) display, the number of error pixels is described as below (sin is approximated with angle in following calculation).

$$640 \div \frac{\pi}{4}\cdot\phi = \frac{2560}{\pi}\gamma\sin\theta[px] \quad (9)$$

For example, if $\theta = \frac{\pi}{30}$, the error is 42.6 γ [px] (If the position of the virtual screen is assumed at the center of TWISTER and the number of cameras is 30, the value of θ is $\frac{\pi}{30}$ when Q is on the boundary of the areas.).

In an output image, this error clearly appears on the boundary of the each camera's output area in the following two situations:

・when the virtual screen is front rather than the actual object's position, the same point will appear twice on a screen.

・when the virtual screen is back rather than the actual object's position, some point will disappear on a screen.

Hence if we want to prevent the error to appear on the screen, we should make the error on the boundary of the area under 1px. Taking account of that the error pixels on the boundary of the area are described as the sum of the approximation error of adjoining area,

$$
\begin{aligned}
& 42.6\gamma + 42.6\gamma' < 1 \\
\Leftrightarrow \quad & \gamma + \gamma' < 0.0117 = 1.2\% \\
\Leftrightarrow \quad & \gamma < 0.6\% \qquad (10)
\end{aligned}
$$

Therefore, for example, when the distance between the object and the virtual viewpoint is 2m, if we set the virtual screen within 1.2 cm from the object, the error from the approximation of the object's position can be ignored in practice (in this case, $\theta = \frac{\pi}{30}$).

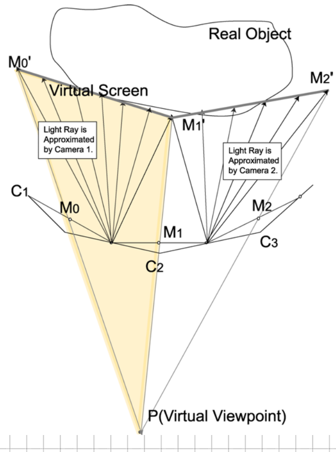
In addition, the value of γ is inversely proportional to $\sin\theta$, and almost directly proportional to the number of cameras. For example, if the number of a camera is increased 10 times, the maximum of γ can be made about 6%.

4-4. How to make Virtual Screen for TWISTER

Considering TWISTER, since we don't have the mechanism for measuring the object's position, we use the model as virtual screen. Although the shape of a plane, a cylinder, etc. can be considered as the shape of the model, we adopted a N-sided prism as a model's shape this time (N is the number of cameras attached on TWISTER at equal intervals) because of the easiness of the rendering procedures. In this way, each camera's virtual projection screen, which is separated by the boundary as mentioned in Section 4-2, is set as a plane parallel to the image surface of the camera and the output clipped from the camera by scanning line can be stuck on a virtual screen only by expansion and contraction. It makes rendering process very easy.

Here we describe in detail about the method of making such a N-sided prism virtual screen. In the following explanation, we explain by using the upper surface figure of the system. For example, an expression "triangular prism ABC" means a prism that is displayed as a triangle ABC in an upper surface figure.

First, we set the center of TWISTER O, virtual viewpoint P, the radius of TWISTER R, N cameras located on TWISTER at equal intervals Ci (i=0, … N-1) and the middle point of CiCi+1 Mi ($C_N=C_0$). And, we set the virtual screen as a N-sided prism $M_0'M_1'M_2'\ldots M_{N-1}'$, which is obtained from a N-sided prism $M_0M_1M_2\ldots M_{N-1}$ by a central similarity of center P. The ratio belongs to the value of δ, which is the distance between the N-sided prism virtual screen and the center of TWISTER. At this time, when the virtual screen moves with movement of virtual viewpoint's position P, the movement follows an envelope curve, a pillar with radius δ and center O. The area on the virtual screen



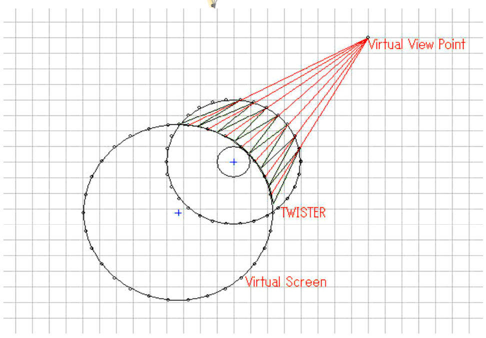


Fig 8: Cylindrical Virtual Screen of TWISTER

which camera C_i takes charge of is $M_i'M_{i+1}'$, which is the plane parallel to the image plane of Camera C_i.

4-5. Necessary Field Angle for Camera

In TWISTER, the area where the object can exist is limited in a safety shield, and cameras are arranged on the cylinder outside the shield. Here we set the distance from the center of TWISTER to each camera R_1 and a radius of the safety shield R_2. At this time, we need $2\sin^{-1}(R_2/R_1)$ for each camera's field angle to capture the whole images in the shield. In other words, if the each camera's field angle α is given, the area captured from all the cameras is the cylinder with radius $R_1 \sin(\alpha/2)$.

Additionally, considering TWISTER II, since the radius of the safety shield is 400mm and the distance from the center of TWISTER to each camera is 600mm, we need $\sin^{-1}(400/600) = 83.6[\mathrm{deg}]$ for the field angle of cameras to capture the whole images in the shield.

4-6. Summary of Theory

We have proposed the methods to determine the camera for each light ray incident to the arbitrary virtual viewpoints, and the method to make virtual screen, and have evaluated the error between the virtual screen and the object. As described in Section 4-2, when we adopt "the bisector model" for the camera selection, we can reduce the effects on the virtual screen from the approximation error of modeling. At the same time by adopting "the middle point model", which is the approximation of "the bisector model", we can select the camera independently of the determination of the virtual screen. Although, we have set the virtual screen as a N-sided prism for easiness of rendering, this idea can be applied to any virtual screen model.

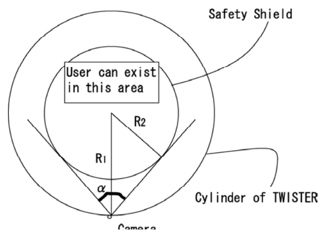


Fig. 9 Necessary Field Angle for Camera

5. Experiment

The experiment was conducted based on the above theory. We have selected the required light information by scanning line here. The rendering process is planned in the near future.

5-1. System Architecture

To imitate the situation that 30 cameras are arranged on a cylindrical device (TWISTER), 12 small CCD cameras (Toshiba IK-C40) were put in order at intervals of 12 degrees (Fig. 10). The synchronization is taken by the common Genlock input, and each camera is rotated 90 [deg], so that the scanning lines of a camera become direction of the perpendicular direction. This arrangement enables us to select line images from analog signals in a scanning line unit.

Since the output of 12 cameras has the huge amount of data and it is difficult to deal it with a single PC, we select required NTSC scanning lines in an early stage using switching circuits.

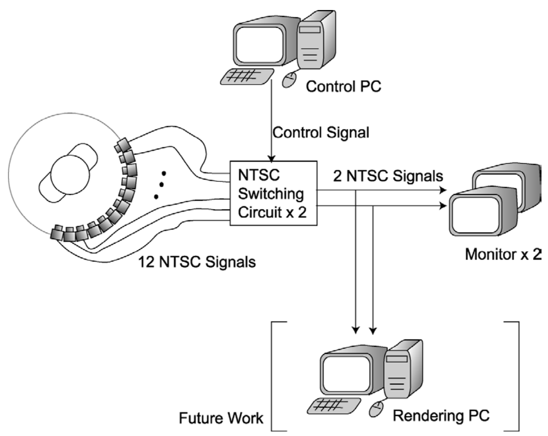


Fig. 10 Image Capture System

5-2. Virtual Screen

In this experiment, we set the virtual screen as a 30-sided prism using the method described in Section 4-4.

5-3. Result

The picture obtained in the NTSC selection stage is shown in Fig. 11. These images represent the necessary NTSC scanning lines for reconstructing the image of the object seen from a virtual viewpoint, which position is 1800mm distant from the center of the booth, using the output of 5 cameras of 12 sets. In addition, since there exists some overlap areas, where two adjacent cameras require the same number of scanning line, we arranged two switching circuit and divided the output of the even-number cameras and the odd-number cameras.

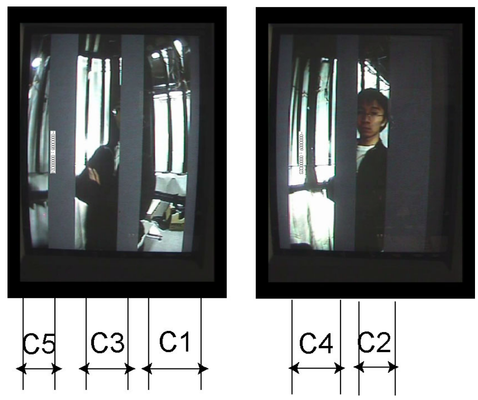


Fig. 11 Output Images Selected by Scanning Lines

As a next step, we plan to input these obtained images to the rendering PC, texture-map on a 30-sided prism virtual screen, and generate an object's image seen from the virtual viewpoint in real time (Fig. 12).

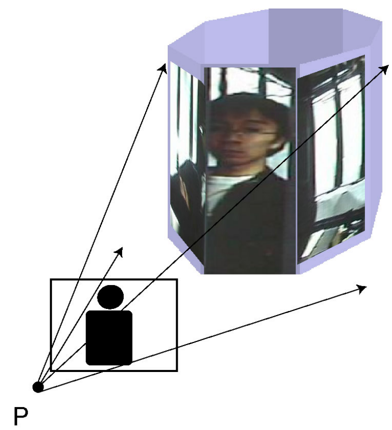


Fig. 12 Texture mapping and rendering (Future work)

6. Future Works

6-1. Object Clipping from a Background

In TWISTER, since only the object is clipped from obtained images and arranged to the virtual environment, an actual background does not have a meaning. Hence, we should consider the method of clipping the object image from the background image. Basically, we plan to use luminance key. Only the object is lighted, while surroundings are made dark, and the object is clipped from black background.

Since TWISTER system contains not only image capture component but also display component, we have to delete the image of bright LED. There are some methods to delete the image of LED.

- Synchronize the image capture cycle with LED's flashing cycle.
- Calculate the position of LEDs and overwrite black there

Much more consideration is necessary about this issue.

6-2. Rotation of Cameras

As evaluated in Section 4-3, it is difficult to obtain the accurate images by using only 30 cameras or so. We considers image-capture component to rotate around the object with display component to increase the number of cameras virtually,

By rotating high-speed camera around the user while getting many images in high time frequency, we can obtain images from much more positions than the number of cameras (We call this camera "Virtual Camera"). Furthermore, since high-speed movements of cameras are hard to recognize for the user, the cameras can capture the user's image without interrupting user's viewing action.

7. Summary

We have descried the theory of image-capture system of TWISTER, the outline of the system design and the experiment for it. From this consideration, it turns out that how to choose the camera reconstructing each light information and the setting method of a virtual screen can be considered independently. Furthermore, we have evaluated the error between virtual screen and the position of the object, and suggested the N-sided prism virtual screen for TWISTER.

We will advance the experiment. We plan to generate the object's image from arbitrary virtual positions using this system, and display the images on TWISTER's display component or another devices in the future.

References

1. F. Kishino: Human Communication -the Virtual Space Teleconferencing system-: Proceedings of The Institute of Image Information and Television Engineers, Vol. 46, No. 6, pp. 698-702, 1992.
2. C. Cruz-Neira, D. J. Sandin and T. A. DeFanti: Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE, Proceedings of ACM SIGGRAPH '93, pp. 135-142 (1993)
3. S. Tachi, T. Maeda, Y. Yanagida, M. Koyanagi and Y. Yokoyama: A method of mutual tele-existence in a virtual environment, Proceedings of ICAT '96, pp. 9--18, 1996.
4. Jason Leigh, Thomas A. DeFanti, Andrew E. Johnson, Maxine D. Brown, and Daniel J. Sandin: Global tele-immersion: Better than being there, In ICAT97, pp. 10-17, 1997
5. M. Hirose, T. Ogi, S. Ishikawa and T. Yamada: A Development of Immersive Multiscreen Display (CABIN), Proceedings of the Virtual Reality Society of Japan Second Annual Conference, pp. 137-140 (1997)
6. F. Okuno, H. Hoshino, J. Arai and I. Yuyama: Real-time Pickup Method for a Three-dimensional Image based on Integral Photography, Applied Optics, Vol.36, No.7, pp. 1598-1603 (1997)
7. Y. Kunita, M. Inami, T. Maeda, S. Tachi: Design of Visual Display for Mutual Tele-Existence, Proceedings of the Virtual Reality Society of Japan Second Annual Conference, pp. 352-353 (1997)
8. Y. Kunita, M. Inami, T. Maeda and S. Tachi: Prototype System of Mutual Tele-existence, ACM SIGGRAPH '99 Conference Abstracts and Applications, p. 267 (1999)
9. Y. Kunita, M. Inami, T. Maeda and S. Tachi: Real-time rendering system of moving objects, Proceedings of the 1999 IEEE Workshop on Multi-View Modeling \& Analysis of Visual Scenes (MVIEW'99), pp. 81-88 (1999)
10. Y. Kunita, M. Inami, T. Maeda and S. Tachi: Real-Time Rendering System of 3D Moving Objects, Proceedings of the 3D Image Conference '99, p. 13-18 (1999)
11. Y. Kunita, M. Inami, T. Maeda and S. Tachi: A Real-time Projection System of Photo-Realistic Human Figures into Three-Dimensional Virtual Environments, Proceedings of the 5th Symposium on Intelligent Information Media (IIM 1999), pp. 207-212 (1999)
12. Y. Kunita, N. Ogawa, A. Sakuma, M. Inami, T. Maeda and S. Tachi: Immersive Autostereoscopic Display for Mutual Telexistence: TWISTER I (Telexistence Wide-angle Immersive STEReoscope model I), Proceedings of the 3D Image Conference 2000, p. 111-114 (2000)
13. Y. Kunita, M. Inami, T. Maeda and S. Tachi: Immersive Autostereoscopic Display for Mutual Telexistence (II): Design and Development of

TWISTER II, Proceedings of the Virtual Reality Society of Japan 5th Annual Conference, p. 369-370 (2000)

14. T. Endo, A. Katayama, H. Tamura, M. Hirose: An Omnidirectional Image Construction Method from Multi-viewpoint Images, Proceedings of the Virtual Reality Society of Japan 5th Annual Conference, pp. 415-418 (2000).

15. Y. Kunita, N. Ogawa, A. Sakuma, M. Inami, T. Maeda and Susumu Tachi: Immersive Autostereoscopic Display for Mutual Telexistence: TWISTER I (Telexistence Wide-angle Immersive STEReoscope model I), Proceedings of the IEEE VR 2001, pp.31-36 (2001)

16. J. Tago, T. Nitta, T. Naemura, and H. Harashima: Video-Based Rendering with Dynamic Layered Representation, Proceedings of the 3D Image Conference ‘01, pp.33-36 (2001).

17. K.Tanaka, J. Hayashi, Y. Kunita, M. Inami, T. Maeda, S. Tachi: The Design and Development of TWISTER II: Immersive Full-color Autostereoscopic Display, Proceedings of ICAT '01, pp. 56-63 (2001).

18. J. Hayashi, K. Tanaka, M. Inami, and S. Tachi: An Immersive Full Color Autostereoscopic Display, Proceedings of the 3D Image Conference '02, p. 93-96 (2002)

19. K. Tanaka, J. Hayashi, I. Kawabuchi, D. Sekiguchi, M. Inami, N. Kawakami and S. Tachi: The Design and Development of TWISTER III, an immersive full-color autostereoscopic display for motion pictures, Proceedings of the Virtual Reality Society of Japan 7th Annual Conference, p. 411-412 (2002)