A Method of Constructing a Telexistence Visual System Using Fixed Screens

Yasuyuki Yanagida, Taro Maeda, and Susumu Tachi
Dept. of Mathematical Engineering and Information Physics,
School of Engineering, The University of Tokyo
{yanagida, maeda, tachi}@star.t.u-toyo.ac.jp

Abstract

Projection-based visual display systems are expected to be effective platforms for VR applications, in which the displayed images are generated by computer graphics using three-dimensional models of virtual worlds. However, these kinds of visual displays, as well as other kinds of fixed-screen-based displays such as various head-tracked displays (HTD) and conventional CRT displays, have not been utilized to achieve precise telexistence in a real environment, which requires appropriate stereoscopic video images corresponding to the operator's head motion. We found that the time-varying, off-axis projection required in these systems has prevented fixed-screen-based displays from being used for telexistence, as ordinary cameras only have fixed and symmetric fields of view about the optical axis. After evaluating the problem, a method to realize a live-video-based telexistence system with a fixed screen is proposed, aiming to provide the operator with a natural three-dimensional sensation of presence. The key component of our method is a feature that keeps the orientation of the cameras fixed regardless of the operator's head motion. Such a feature was implemented by designing a constant-orientation link mechanism.

1. Introduction

1.1. Background

Visual display systems based on Immersive Projection Technology (IPT), such as CAVE [1], CABIN [2], Responsive Workbench [3], and others, are expected to be effective platforms for VR applications, as the visual quality provided by those systems is considerably high. It is often pointed out that the factors that afford this high quality include a large field of view obtained by a large screen and high resolution per screen by means of computers and projectors with high-resolution graphics. Moreover, it is considered that visual display systems which use screen(s) fixed to the surrounding environment have some advantages for the stability of the displayed image when the operator moves his/her head.

The angular error of the points in the displayed image, which is caused by the tracking error or the latency of the system, was analyzed to examine the stability of the image provided by a projection-based VR display [1]. In this analysis, the result was compared for Head-Mounted Displays (HMD) [4], CRT, and CAVE, and it was demonstrated that fixed-screen-based paradigms (CRT and CAVE) have significant advantages over the HMD when rotational tracking error exists. This result is inherently derived from the characteristics of fixed-screen-based displays, in which pure rotation about the observation point does not require the displayed image to be updated. A similar analysis was reported [5] focusing on the effect of inevitable time delay through the entire system [6] when using the HMD.

In spite of these merits, fixed-screen-based displays have not been utilized for precise telexistence/telepresence in a real environment. So far, only HMDs have been used for visual display systems in telexistence/telepresence systems, such as TELESAR [7–9] and TOPS [10]. IPT displays have been used only as large- or multi-screen monitors when live video images of the real world are displayed, or they have produced an incorrect stereoscopic view for the user even if his/her head motion is tracked when ordinary fixed stereo cameras are used. This is a common aspect of conventional CRT displays and projection-based head-tracked displays (HTD) with a single screen as well as for multi-screen IPT displays.

In order to provide the operator with a precise three-dimensional sensation of presence by the cue of disparity and convergence, the system must be designed so that the projection used at the remote site (camera) and that at the operator's site are kept consistent. In other words, the shape of the viewing volume at both sites should coincide. Hence, a different configuration of the camera system is required if we use different types of camera systems. We observed that there are no camera systems that provide a viewing volume corresponding to that of display systems

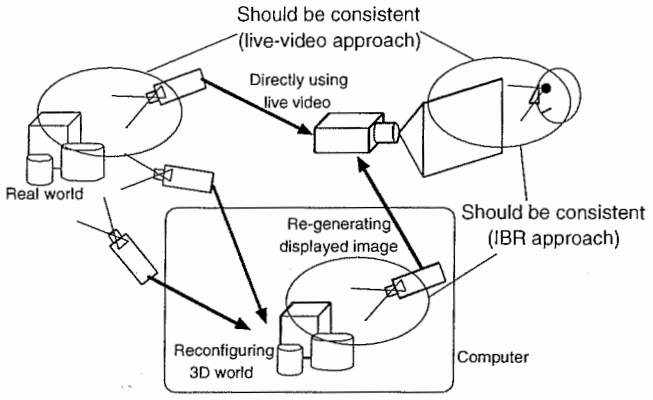


Figure 1. Live-video approach and image-based modeling/rendering approach for telexistence

using fixed screens and found that this is the key to realize a precise telexistence/telepresence visual system using fixed screens. In this paper, we propose a method to construct a precise stereoscopic telexistence system.

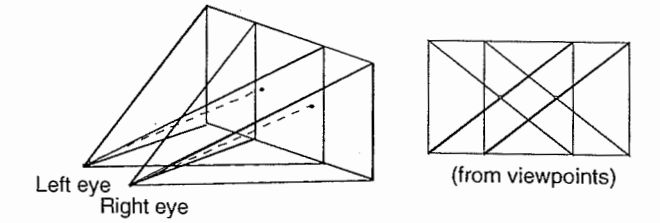
1.2. Scope of the Research

Strictly speaking, precise stereoscopic view requires consideration of miscellaneous sources of error [11], including accommodation convergence, interpupillary distance (IPD), variation in IPD, and the displacement of eye position caused by the motion of the eyeballs. Problems arising from using incorrect IPD in a head-tracked display are also reported [12]. However, we focus on problems related to head tracking and the positional relationship between the screen and cameras/eyes, as this is the most critical factor to provide stereoscopic images. Here, we assume that the IPD is correctly adjusted and that its value is constant. In other words, taking advantage of eye tracking is beyond the scope of this paper. We attempt to achieve a precise stereoscopic view as far as we can do by head tracking, just as many other HMD-based systems currently do.

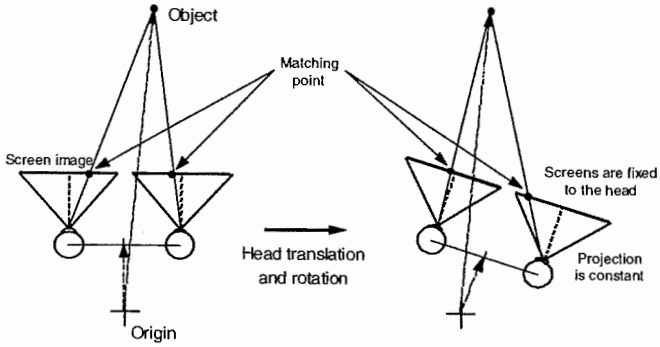
Next, it should be noticed again that we use a live video directly rather than computer-generated graphics images. There are two different approaches to realize telexistence in a real environment (Figure 1):

(1) Live-video-based approach: The image captured by the remote camera is directly sent to the display system at the operator's site.

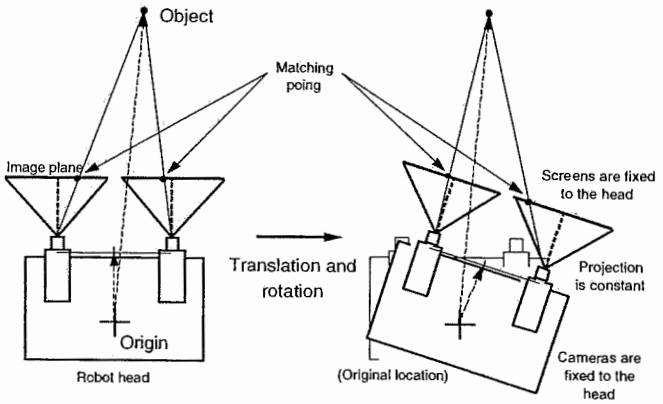
(2) Image-based modeling and rendering (IBR) approach [13, 14]: The remote environment is acquired by using one or more cameras and range-finding devices to reconstruct a three-dimensional model of the remote environment in the computer, and the displayed image is generated by computer graphics.



(a) Shape of the viewing volume



(b) Head motion at the operator's site

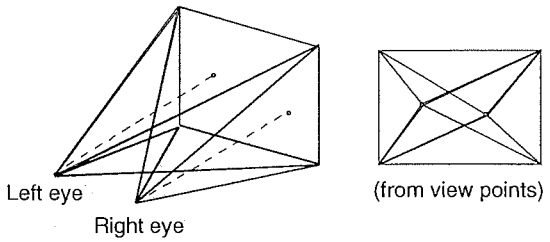


(c) Head motion at the remote site

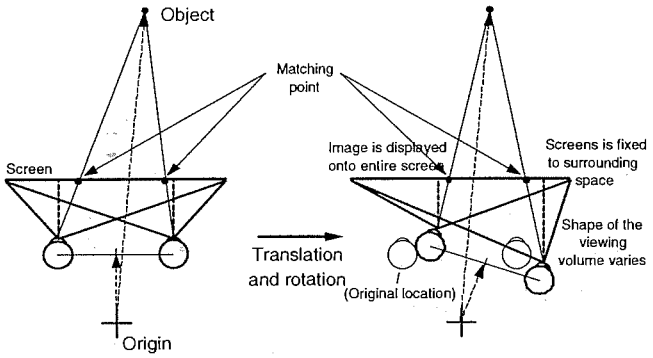
Figure 2. Projection for the HMD system

The main difference between the two approaches is that the video image is used *as is* (in the form of the sequence of two-dimensional pictures) in the live-video-based approach, whereas a three-dimensional model of the world intermediates the captured image and the displayed image (explicitly or implicitly) in the IBR approach. The IBR approach offers freedom to specify the viewpoint and the parameter of projection when generating the displayed image. In other words, the IBR approach can be regarded as a way to bring the whole remote environment to the computer at the operator's site. Thus, it allows multiple users to share a virtual space by synthesizing multiple spaces into a single virtual world.

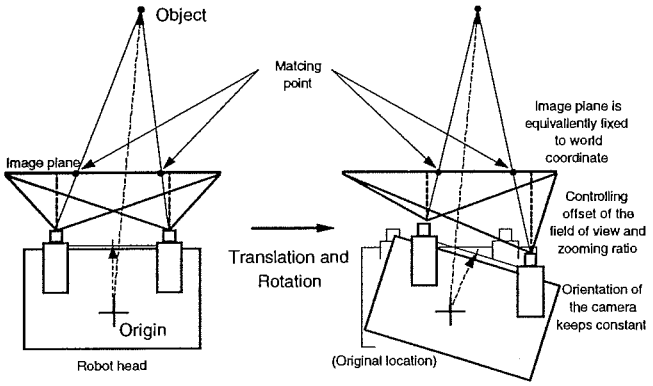
On the other hand, in the live-video-based approach, it is necessary to keep the consistency between the projection performed at the remote site and that at the op-



(a) Shape of the viewing volume



(b) Head motion at the operator's site



(c) Head motion at the remote site

Figure 3. Projection for the fixed-screen-based system (Type A: off-axis projection, used for CG)

erator's site. As a consequence, a robot is used to follow the operator's head motion, occupying physical space at the remote site. In this case, an operator is mapped to his/her slave robot, so that this approach can be regarded as a way to take the operator to the remote environment. Multiple users can share a space *physically* by using a robot existing in a remote environment for each.

In the following chapter, we discuss the problems inherent in displaying stereoscopic video images to a fixed screen in depth.

2. Analysis of the Problem

2.1. Projection for HMD and Fixed Screen

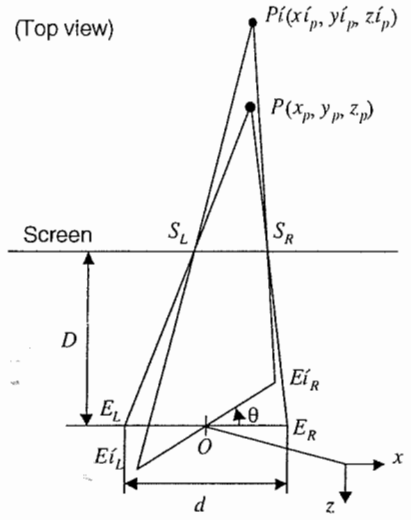
As we mentioned before, the projection used to generate (capture) images at the remote site and to display the image at the operator's site should coincide. This projection depends on the type of display device/system. To clarify this, let us compare the projection used for an HMD and that for a fixed-screen-based display.

The shape of the viewing volume for a typical HMD is shown in Figure 2 (a), and the treatment when the operator's head moves is shown in (b). Most typical HMDs are designed so that the image of the screen is parallel to the operator's forehead and the field of view is symmetric about the optical axis. In this case, the shape of the viewing volume is a regular pyramid and is constant, as the screens of HMD are fixed relative to the operator's head. Considering these characteristics, we can use ordinary cameras fixed to the robot's head, arranged so that the distance between the two cameras is identical to that of the operator's eyes (Figure 2 (c)). The operator can see a natural three-dimensional world around him/her if the position and orientation of the robot head are controlled to follow the operator's head.

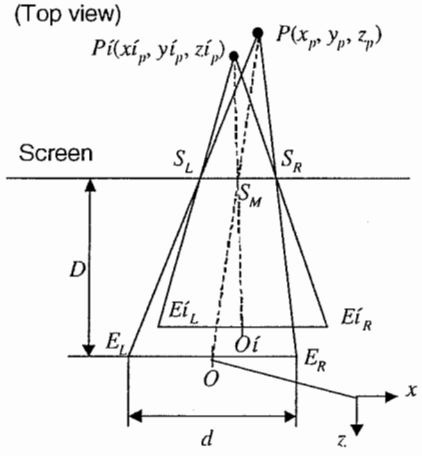
On the other hand, off-axis projection [15], also known as the function `glFrustum()` in OpenGL API [16], is used to generate computer graphics images for fixed-screen-based systems, as shown in Figure 3 (a). The shape of the viewing volume is a general frustum, and the optical axis does not point to the center of the screen. When the operator moves his/her head, the shape of this frustum varies according to his/her motion in real time, as shown in Figure 3 (b). It is easy to control the shape of this frustum when generating computer graphics images, as we can specify the parameters of this frustum whenever we start drawing the image for each frame. However, we cannot apply this scheme to obtain real-time video images because such off-axis projection is not popular for cameras and controlling the shape of the viewing volume in real time by the camera optics is rather difficult.

2.2. Need for Head Tracking

Even though a fixed-screen-based display can provide stable images corresponding to the operator's head motion, head tracking is still important. Let us consider a situation in which a stereo image that does not reflect the operator's head motion is displayed on the screen. In this case, the point on the screen does not move according to the operator's head motion. Then, the perceived point moves according to the operator's motion. Figure 4 (a) shows the case with the operator's rotational head motion,



(a) Effect of head rotation



(b) Effect of head translation

Figure 4. Distortion of the world caused byusing non-head-tracked images

and (b) with translational motion. In each figure, the left and right eyes are located at the points E_R and E_L, respectively, and the image of the point P is projected at the point S_R and S_L on the screen. The operator's head is directed to the screen at the initial state. The interpupillary distance of the operator is denoted as d, and the distance between the screen and the center of the eyes is denoted as D.

If the operator rotates his/her head by θ, as shown in Figure 4 (a), the position of each eye moves to E'_R and E'_L, respectively. As the projected point does not move, the point $P(x_p, y_p, z_p)$ is perceived to move to P'. If we set the origin O at the center of the eyes, the x and z coordinates of the point P' are calculated as follows:

$$x = D\frac{\alpha\cos\theta + (\alpha^2 - \gamma^2\beta^2)\sin\theta + \gamma^2\beta\sin\theta\cos\theta}{\beta - (\cos\theta + \alpha\sin\theta)},$$

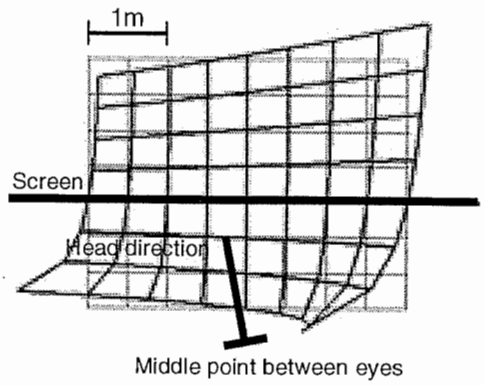


Figure 5. Distortion of the perceived world by using non-head-tracked stereo images (top view)

$$z = D\frac{\cos\theta + \alpha\sin\theta - \gamma^2\beta\sin^2\theta}{\beta - (\cos\theta + \alpha\sin\theta)},$$

where

$$\alpha = x_p / z_p,$$

$$\beta = 1 + \frac{D}{z_p}, \text{ and}$$

$$\gamma = \frac{d}{2D}.$$

This relationship was derived from the condition that the set of points P', S_L, E'_L and P', S_R, E'_R are located on the same line, respectively.

The calculated distortion is shown in Figure 5. Here, we used the value $\theta = 10$ [deg] and $D = 2.0$ [m]. The size of each mesh is 0.5 [m], and the calculated points distribute in the range of –2 to 2 [m] along the x-axis and –0.5 to –4 [m] along the z-axis. The result shows a significant distortion of space.

Next, let us consider a case of translational motion, as shown in Figure 4 (b). This situation is equivalent to the case in which there is displacement of the eye (face) against the HMD optics [17]. The displacement of eye position in fixed-screen-based systems, however, is dynamic, unlike the case of HMD, where the relative position of each eye is fixed to the screen once the operator puts on the HMD. Now let us review this situation briefly. When the center of both eyes moves from the origin O to the point $O'(\Delta x, \Delta y, \Delta z)$, the point P is perceived to move to P'. If we denote the center between S_R and S_L as S_M,

$$\overrightarrow{OS_M} = -\frac{D}{z_p}\overrightarrow{OP}.$$

As both lines, OP and $O'P'$, include the point S_M, the location of the moved point is described as

$$\overrightarrow{OP'} = \overrightarrow{OO'} + \overrightarrow{O'P'} = \overrightarrow{OO'} - \frac{z_p}{D}\overrightarrow{O'S_M}$$

$$= \overrightarrow{OO'} - \frac{z_p}{D}\left(\overrightarrow{OS_M} - \overrightarrow{OO'}\right)$$

$$= \overrightarrow{OP} + \left(1 + \frac{z_p}{D}\right)\overrightarrow{OO'}$$

This indicates that the point P moves in the same direction as the operator's motion if it is located in front of the screen ($z_p > -D$), and in the opposite direction if it is located behind the screen ($z_p < -D$), with the entire space expanding or shrinking about point S_M. No points originally located on the screen are affected by the operator's head motion.

2.3. Direction of the Camera

Another problem is that the image plane of the camera cannot be kept parallel to the screen if the camera is fixed to the robot's head. The image projected onto the plane, which is not parallel to the display screen, cannot be restored by simple two-dimensional image manipulation. It requires some time-consuming three-dimensional compensation, e.g., texture mapping to a virtual screen. Though recent progress in 3D graphics hardware enables almost real-time processing of the captured image, it is not advisable to force excessive processing on the system. Moreover, a problem arises if the camera points towards the edge of the screen, causing the effective display area on the screen to be extremely reduced.

3. Principle of a Telexistence Visual System Using a Fixed Screen

According to the discussion in the previous chapter, the problems to be solved to construct a fixed-screen-based telexistence system can be summarized as follows:

(a) Tracking the operator's head and functionally controlling the time-varying off-axis viewing volume, which consists of each of the operator's eye and the display screen.

(b) Keeping the image plane of the camera parallel to the supposed screen at the remote site, which is a copy of the screen at the operator's site. This is a requirement to avoid complex three-dimensional image processing and to use both the screen area and the captured image as much as possible.

It is possible to construct a precise telexistence system using fixed-screen-based display systems as long as both of above conditions are fulfilled in real time.

When implementing these two functions, several possible choices can be taken into account. First, a method to realize the time-varying off-axis projection by a camera can be considered:

Type A: A special camera is used whose optical axis and zooming ratio can be controlled in real time to enable the time-varying off-axis projection at the stage of image generation. In this case, the job at the display system (operator's site) is completely identical to that of the systems for displaying computer graphics images (Figure 3).

Even though this is a certain solution for the problem, we still attempt to solve this problem when using an ordinary camera whose field of view is symmetric about the optical axis. Concerning problem (a), there is an alternative approach to substituting symmetric projection for off-axis projection, as long as we allow a part of the screen or the captured image to be discarded. In this case, the control of the time-varying off-axis projection results in the control of the size and position of the image captured by the symmetric projection.

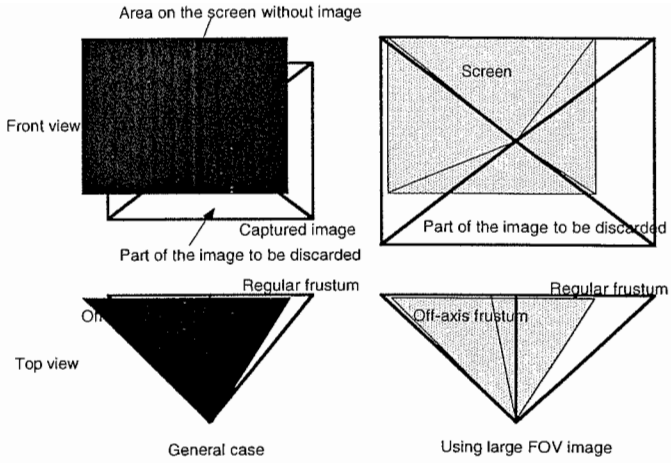
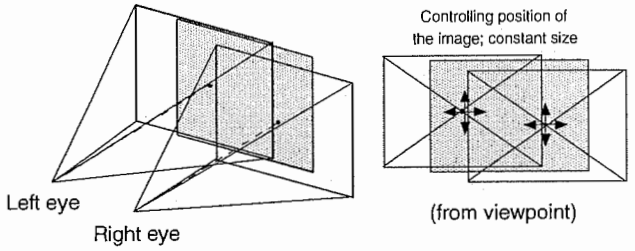


Figure 6. Substituting a regular pyramid for the off-axis frustum

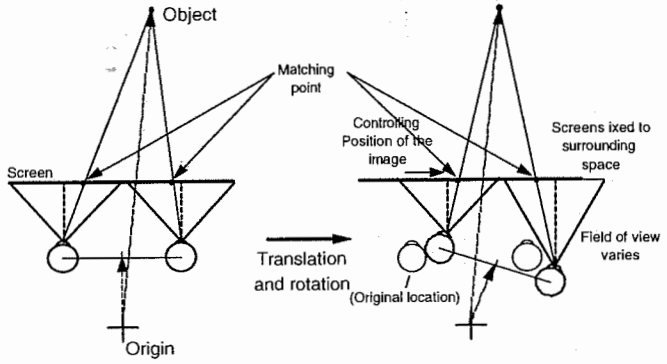
Figure 6 shows the use of the viewing volume in the shape of a regular pyramid as the substitution for the off-axis frustum. The upper side is the front view, and the lower side is the top view. In this method, the image captured by an ordinary camera with symmetric viewing volume is used, and the position and size of the image are adjusted before the image is displayed on the screen at the operator's site. This adjustment takes place so that the operator can obtain the same image as he/she would observe by a special camera with off-axis projection everywhere in the area with image. This substitution does not cause any extra distortion.

We further categorize the method into two types according to the role of the image-capturing subsystem and the display subsystem.

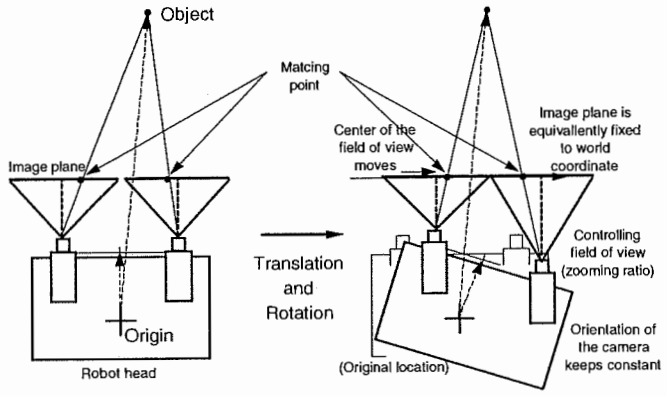
Type B: The position of the image is controlled at the display subsystem, and its size is kept constant.



(a) Shape of the viewing volume

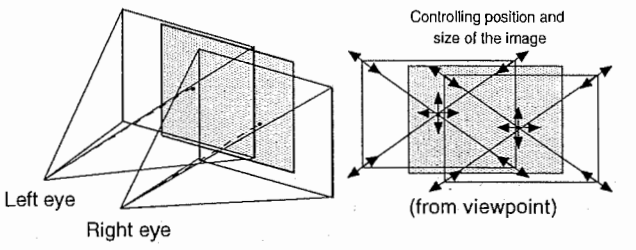


(b) Head motion at the operator's site

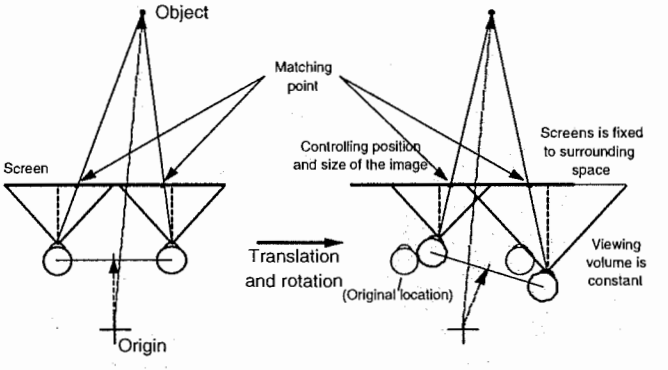


(c) Head motion at the remote site

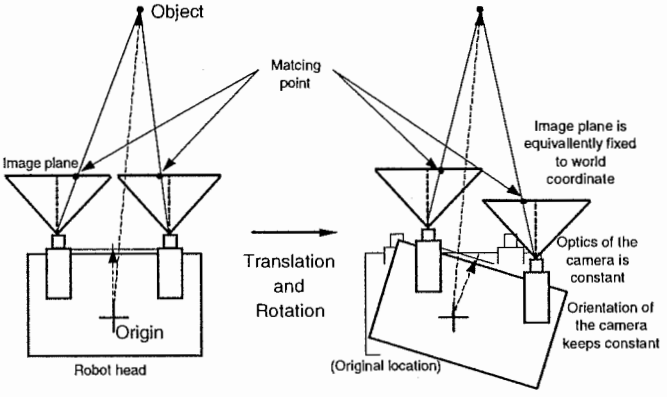
Figure 7. Projection for a fixed-screen system (Type B: Zooming at the remote site and position control at the operator's site)



(a) Shape of the viewing volume



(b) Head motion at the operator's site



(c) Head motion at the remote site

Figure 8. Projection for a fixed-screen system (Type C: Both zooming and position control at the operator's site)

In this case, the zooming ratio of the camera should be controlled in real time, as the field of view of the image on the screen varies according to the operator's head motion (Figure 7).

Type C: The optics of the camera is completely fixed. Both the position and size of the displayed image are controlled at the display subsystem so that the field of view at the operator's site remains constant (Figure 7).

In Type B, the display subsystem is simple and easy to configure. It still requires real-time control of the zooming ratio, which could limit the characteristics of the system response. In Type C, the image-capturing subsystem will be the simplest of all these methods. The position and size of the displayed image can be controlled by electronic, electric, or optical/mechanical methods.

4. Implementation of the System

4.1. System Overview

The block diagram of the entire system is shown in Figure 9. The operator's head motion is measured by a mechanical tracker with 6 degrees of freedom (Shooting Star ADL-1), and the position of each camera is controlled to follow the operator's head motion. The video image obtained by the camera is sent to the image ma-

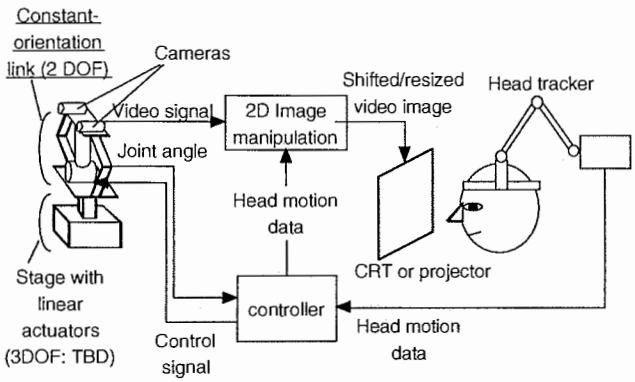


Figure 9. Prototype configuration of a telexistence visual system using a fixed screen

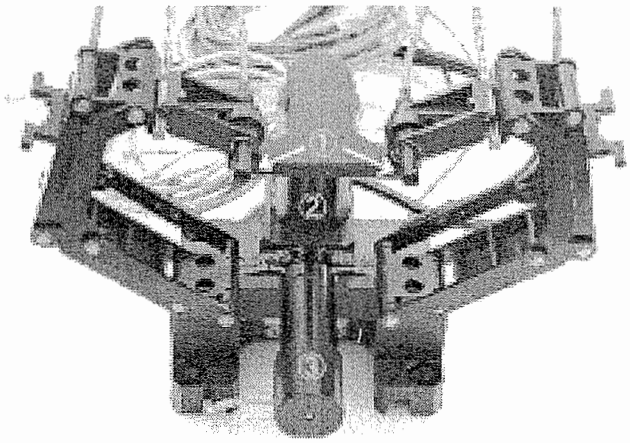


Figure 10. A Prototype of a constant-orientation link: (1) CCD cameras, (2) motor for the yaw axis, (3) motor for the roll axis.

nipulation subsystem and processed before being displayed on the screen. We used ordinary PCs as the image manipulation subsystem and the camera position controller in this first prototype system. Among the elements that compose the system, a mechanism to control the camera position and orientation is specific for fixed-screen-based telexistence system so that a prototype of the mechanism is originally designed.

4.2. Constant-Orientation Link

The technical element commonly required by the three methods described in the previous chapter is a mechanism to keep the orientation of the camera constant, regardless of how the operator moves his/her head. This requirement is derived from "problem (b)" in Chapter 3. To implement this function, it is better to compose a mechanical link with constraints than to provide an excessively independent joint on the top of the robot head.

The function required for the link is to follow the operator's yawing motion and rolling motion, whereas the orientation of the camera is kept constant. A pitching motion is not necessary, as the two cameras move in the same way. Finally, a complete system using a fixed screen can be constructed if this link mechanism is carried on the stage, which can translate itself with 3 DOF.

Figure 10 shows the prototype model of the constant-orientation link. Actually, this link mechanism consists of two parts: 2 DOF serial links to move the camera ("neck" part) and parallel links and a sliding mechanism to keep the orientation of the camera ("wing" part). The link is designed so that it has less moment about the yaw axis, considering the human operator's characteristics for head rotation. The movable range of the joints is 70 [deg] for the yaw axis and 30 [deg] for the roll axis, respectively. A DC motor drives each joint (70W for the yaw axis and 20W for the roll axis), and a rotary encoder (3000 [ppr]) measures the angle of each joint. The link is equipped with a small CCD camera (7mm in diameter, 0.41M pixels) to obtain the image corresponding to the position of the operator's right/left eye.

A PC measures the value of the joint angle through the up/down counter board, and the control process runs on a PC. Two DC motors for the roll and yaw axes are driven by a PWM circuit. The waveform of the pulse fed to the motor is generated from a 1MHz clock with a resolution of 8 bits (256) for duty ratio, which results in a frequency of pulse at 3.9kHz. The target value is sent from the PC to the driver circuit at the rate of 1kHz. We configured a master-slave system using a mechanical tracker (ADL-1), and the prototype of constant-orientation link could follow the operator's natural rotational head motion.

4.3. Image Manipulation Subsystem

The image manipulation subsystem was implemented based on the "Type C" method, i.e., the captured image was shifted and resized before being displayed to the screen. A graphics board with NTSC video input/output interface (Canopus Spectra 2500, nVIDIA RIVA TNT chip, AGP port) was used for two-dimensional image manipulation. This graphics board was installed on a PC (Intel Pentium-II 450MHz CPU). The captured video image was stored to the local memory on the graphics board and then shifted and expanded/shrunk using a feature of Microsoft DirectDraw. The shift amount and the size of the image could be controlled at the rate of 1/60 [sec].

5. Conclusion

This paper discusses the problems inherent in constructing a precise head-tracked stereoscopic display system using live-video and fixed-screen lines, i.e., that the shape of the viewing volume changes and that the direc-

tion of the camera is rotated if the operator moves his/her head. Based on this analysis, we propose a method to construct a telexistence visual system using fixed screens, i.e.,

(1) Introducing the mechanism to maintain the orientation of the camera while the motion of the operator's head is tracked and the position of the cameras is controlled to follow the position of the operator's eyes.

(2) Controlling the shape of the off-axis time-varying viewing volume, which can be equivalently realized by shifting and resizing the image obtained by an ordinary camera with symmetric projection.

Based on the proposed principle, a prototype system was constructed to show the feasibility of this method. By using this method, we can fully exploit the advantageous characteristics of CAVE and other fixed-screen-based systems for telexistence in real environment, which have been available only for displaying computer graphics images so far.

It can be noticed that our proposed method is a generic technology applicable for any type of fixed-screen-based visual display systems, which means that the method can be applied to simple and easy-to-construct systems such as systems using CRT. Future research will include the application of this method to fields requiring stable and consistent three-dimensional sensation, such as tele-surgery systems. Furthermore, the evaluation of the entire system is an important goal for the future.

Acknowledgement

This study was supported in part by a Grant for "Research for the Future Program # 97I00401" from the Japan Society for the Promotion of Science.

References

[1] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti, "Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE," *Computer Graphics (Proc. SIGGRAPH '93)*, pp. 135–142, 1993.

[2] M. Hirose, T. Ogi, S. Ishiwata, 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.

[3] W. Krüger, C-A. Bohn, B. Fröhlich, H. Schüth, W. Strauss, and G. Wesche, "The Responsive Workbench," *IEEE Computer*, pp. 42–48, July 1995.

[4] I. E. Sutherland, "A Head-Mounted Three-Dimensional Display," *Proc. Fall Joint Computer Conference, AFIPS Conf. Proc.*, Vol. 33, pp.757–764, 1968.

[5] Y. Yanagida, M. Inami, and S. Tachi, "Improvement of Temporal Quality of HMD for Rotational Motion," *Proceedings of The 7th IEEE International Workshop on Robot and Human Communication (RO-MAN '98)*, pp. 121–126, 1998.

[6] M. R. Mine, "Characterization of End-to-End Delays in Head-Mounted Display Systems," *Tech. Rep. TR93-001*, Dept. of Computer Science, University of North Carolina at Chapel Hill, 1993.

[7] S. Tachi, K. Tanie, K. Komoriya, and M. Kaneko, "Tele-existence (I): Design and Evaluation of a Visual Display with Sensation of Presence," *Proc. 5th Symp. On Theory and Practice of Robots and Manipulators (RoManSy '84)*, pp. 245–254, 1984.

[8] S. Tachi, H. Arai, and T. Maeda, "Tele-Existence Master-Slave System for Remote Manipulation," *Proc. Int'l Workshop on Intelligent Robots and Systems (IROS '90)*, pp. 343–348, 1990.

[9] S. Tachi and K. Yasuda, "Evaluation Experiments of a Tele-Existence Manipulation System," *Presence*, Vol. 3, No. 1, pp. 35–44, 1994.

[10] M. S. Shimamoto, "TeleOperator/TelePresence System (TOPS) Concept Verification Model (CVM) Development," *Recent Advances in Marine Science and Technology '92*, HI, USA, pp. 97–104, 1992.

[11] W. Robinett and J. P. Rolland, "A Computational Model for the Stereoscopic Optics of a Head-Mounted Display," *Presence*, Vol. 1, No. 1, pp. 45–62, 1992.

[12] Z. Wartell, L. F. Hodges, and W. Ribarsky, "Balancing Fusion, Image Depth, and Distortion in Stereoscopic Head-Tracked Displays," *Computer Graphics (Proc. SIGGRAPH '99)*, pp. 351–358, 1999.

[13] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs, "The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays," *Computer Graphics (Proc. SIGGRAPH '98)*, pp. 179–188, 1998.

[14] R. Raskar, "Oblique Projector Rendering on Planar Surfaces for a Tracked User," *SIGGRAPH '99 Conference Abstracts and Applications*, p. 260, 1999.

[15] W. Robinett and R. Holloway, "The Visual Display Transformation for Virtual Reality," *Presence*, Vol. 4, No. 1, pp. 1–23, 1995.

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

[17] J. P. Rolland and W. Gibson, "Towards Quantifying Depth and Size Perception in Virtual Environments," *Presence*, Vol. 4, No.1, pp. 24–49, 1995.