ICAT '99

Development of Surrounded Audio-Visual Display System for Humanoid Robot Control

Takashi Nishiyama[1)], Hiroshi Hoshino[1)], Kenshi Suzuki[1)], Ryoji Nakajima[1)], Kazuya Sawada[1)] and Susumu Tachi[2)]

1) Advanced Technology Research Laboratory,
Matsushita Electric Works, Ltd.
1048, Kadoma, Osaka, 571-8686, JAPAN
{takashi, hoshino, kenshi, ryoji, sawada}@ai.mewco.jp
2) Department of Mathematical Engineering and Information Physics, The University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo, 113, JAPAN
tachi@star.t.u-tokyo.ac.jp

Abstract

This paper presents an audio-visual display system that is embedded in a cockpit where a human operator controls a tele-operated humanoid robot as if he was on the robot. The developed audio-visual display system consists of a surrounded visual display subsystem, a HMD with a head-tracking function, a surrounded audio display subsystem and a man-machine interface subsystem. Design principles, architectures and implementations of those subsystems are precisely described. These subsystems should be connected to a master-slave controlling subsystem in the humanoid robotics system.

Key words: humanoid robotics system, surrounded visual display, tele-existence, stereoscopic camera system, augmented reality

1. Introduction

Nowadays, automatization in the following fields is in greater needs; the maintenance of plants and power stations, the working of construction work, providing aid in case of emergency and disaster and caring the elderly in the coming aging society. In order to meet these demands, the national 5-year-project of HRP, Humanoid Robotics Project, was established in 1998 by MITI, Ministry of International Trade and Industry[1]. This report focuses on the development of an audio-visual display system for a humanoid robotics system that has been researched and developed in the project.

Fig.1 shows an overview of the humanoid robotics system. In applying the humanoid robot, shown in the figure on the right, to the above field, it is desirable that the robot moves and operates in an autonomous way on behalf of a human operator. However, it happens that the robot is not able to decide its own behavior by itself and it is requisite that the human operator decides the robot's behavior in the end. That is, the human operator remotely controls the movement of the robot. In the field of teleoperation and telerobotics[2], it is required to introduce tele-existence technology where the human operator is made to feel that he is located at the robot site, when the operator is provided with as much information about the site as possible in the form of audio-visual and/or force feedback[3],[4]. This tele-existence technology is embedded in the cockpit for the tele-operated humanoid robot shown in the Fig.1 on the left. Therefore, the human operator in the telexistence cockpit controls the humanoid robot as if he was on the robot and expands his own audio, visual and tactile function from the viewpoint of time and space. We describe the audio-visual display system that consists of a surrounded visual display subsystem, a HMD, head-mounted display, with a head-tracking function, a surrounded audio display subsystem and a man-machine interface subsystem, except for a force feedback subsystem and a communication subsystem between the cockpit and the robot.

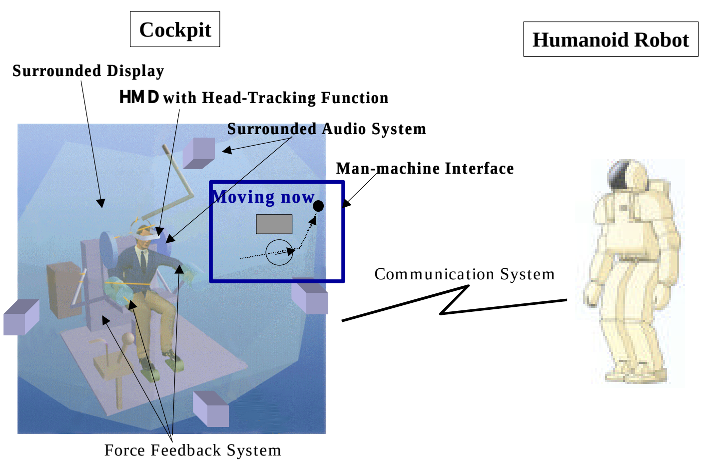

Fig.1 An Overview of the Humanoid Robotics System

2. Surrounded Visual Display System[5]

2.1 Design Principles of the Surrounded Visual Display System

The surrounded visual display system shown in Fig.1 displays stereoscopic real images captured by cameras mounted on the robot to the human operator with a sense of high realty. Thus, the surrounded visual display should be designed to satisfy the following four conditions; 1) a wide field of view, 2) high resolution, 3) binocular stereoscopic displaying, and 4) real time processing. Here, the condition of "real time processing" means that it takes little time for preparing a series of images presented on the display system. Also, a speed of updating the series of images is high enough for the human operator to look at the displayed images to control a tele-operated robot. It is self-evident that the other three conditions of 1), 2) and 3) should be satisfied. Design principles on the four conditions are as follows.

1) A wide field of view
The wider a view field of a robot site is given to a human operator, the more he gets the feeling of being surrounded by the robot site [6]. In our research, we aim to realize more than 100-degree field of horizontal view. Because it is generally acknowledged that the effect of a wide field of view becomes saturated when the view exceeds so-called "induction field of view" around 100-degree [7].

2) High resolution
It is requisite that a resolution of displaying images is high. However, keeping the resolution high is incompatible with keeping a field of view wide, which causes a trade-off between a resolution and a field of view to occur. Because the surrounded visual display system will be mainly utilized for a man-machine interface in the phase of "locomotion control", we have designed to make the display resolution correspond to the eyesight of 0.3, which is the lowest eyesight level for people to drive a car without glasses.

3) Binocular stereoscopic displaying
It is self-evident that a binocular stereoscopic display system allows us to perceive a three-dimensional space in a natural way.

4) Real-time processing
In order to present a human operator with stereoscopic images in a wide field of view, we adopted an approach to make multi-cameras placed in a panoramic form, and to make these cameras capture real images presented on a certain display. A shape of the display can be a curved surface, either domed [8] or spherical [9]. However, we have chosen such a multi-screen type as is used in the CAVE system [10]. Presenting real images on a curved display leads to a barrel or spool-shaped warp. Thus, the warped images need revising to be presented on the curve-shaped display in a natural way. However, it takes too much time to revise the warped images. This does not match the condition of "real-time processing" as is mentioned above. Thus, it is necessary to take an alternative approach, using a couple of multi-cameras and multi-screen type display in order to satisfy the "real-time processing". If we set each screen corresponding to each camera and present real images captured by each camera on each corresponding screen, any wrap will not be caused in displayed images. Therefore, the revision is not necessary.

2.2 Structure of the Developed System

The developed surrounded display system is composed of 9 pieces of screens as is shown in **Fig. 2**. Each screen has 60 inches for the diagonal distance. On a backside of each screen, two projectors are allocated for displaying the stereo-images, which is to be realized by polarizing the right-eye and left-eye images. The shape of the surrounded display is approximated along a spherical surface. Thus, each normal line attached to each screen crosses with each other on one point that corresponds to a central point of the approximated sphere. The distance between the central point and each screen is about 1.3 meter. The human operator looks at, away from the distance of 1.3 meter, the stereo images projected on the surrounded display. Therefore, the angle of the vision field to each screen is 50 ° (H)× 39 ° (V), whereas the angle of the vision field to the whole screen is 150 ° (H) × 117 ° (V). **Table 1** summarizes a specification of the developed surrounded display system, and a global view of the system is shown in **Fig.3**.

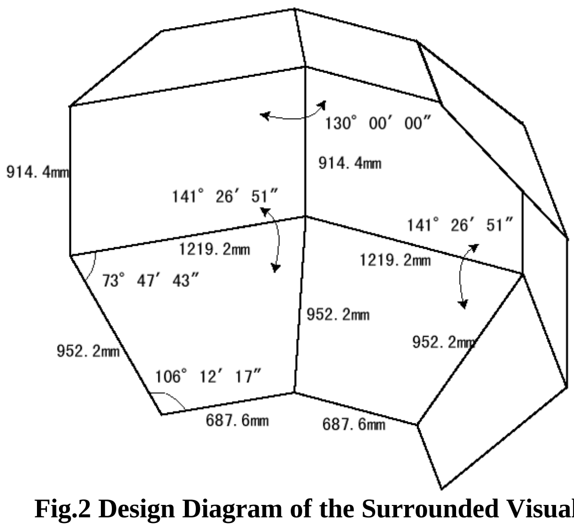

Fig.2 Design Diagram of the Surrounded Visual Display System

Table 1. Specification of the Surrounded Visual Display

Field of View	150 ° (H) ×117 ° (V)
Maximum Resolution	Eyesight 0.3

Stereoscopic Method	Polarized Method
Feature	9 pieces of screens
Distance to each screen	About 1.3 meter

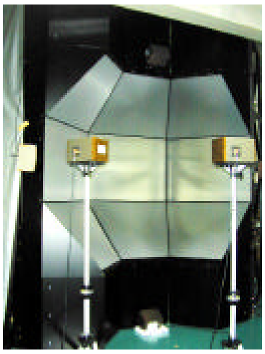

Fig.3 A Global View of the System

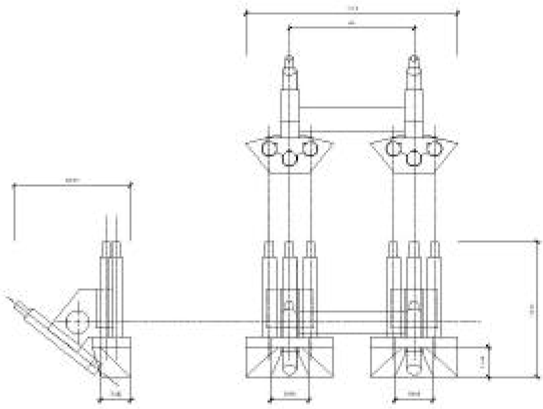

Fig.4 Design Diagram of Stereo-multi-camera System for a Wide Field of View

2.3 Stereoscopic Camera System for the Surrounded Visual System

In HRP, a stereo multi-camera system for a wide field of view is mounted on the humanoid robot as is shown in Fig.1on the right. The real stereo images captured by the multi-camera system are displayed on such 4 portions of the surrounded screen as the left, right, center and bottom. **Fig.4** shows a design diagram of the stereo multi-camera system for a wide field of view. There are two sets of 4 small cameras: each set corresponds to each eye, allocated the distance of 65mm from each other. Each camera corresponds to each screen. The vision field of the camera system is set 150 ° in the horizontal, 19 ° in the upper vertical and 58 ° in the lower vertical directions. **Fig.5** shows a photo of the developed multi-camera system.

3. HMD System with Head-Tracking Function

3.1 The Reason the HMD System with Head-Tracking Function is needed

In the cockpit, the human operator controls the movement of two legs, arms and hands of the humanoid robot. The movement of the two legs makes a body of the robot walk around, whereas those of the two arms and hands allow the robot to manipulate objects in the robot site. The surrounded visual display system provides the human operator with a wide field of view enough to control the movement of the two legs. On the other hand, a HMD with a head-tracking function allows the human operator to easily manipulate the objects in the robot site. When the human operator moves his head, a camera platform mounted on the robot pans and tilts in real time to track the operator's head motion. Stereo images at which the operator's head faces are displayed on the HMD. Thus, the human operator can easily control the two arms and hands of the robot to manipulate the objects in the robot site.

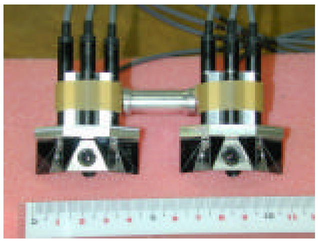

Fig.5 Photograph of the Developed Stereo-multi-camera System for a Wide Field of View

3.2 Specification of the Developed System

We have paid attention to the following items in realizing a HMD system with a head-tracking function. According to measuring principles, there are three methods to implement a head-tracking function, 1) mechanical link method, 2) remote-sensing method, and 3) internal-sensing method. The features of those methods are described as follows:

1) Mechanical link method
Suppose we design a link mechanism that has more than 6-degree of freedom. A top portion of the link mechanisms is set on a head of a human operator. If the operator moves his head, then encoders or potentiometers mounted on joints of the link mechanism track the movement to count values of joint angles. An algorithm of kinematics enables us to calculate the values to determine a position and a direction of the operator's head in a 3-dimensional space. The advantage of this method is both speed and accuracy of measurement is high and an approach itself is robust to a noise around a measurement environment. On the other hand, the disadvantage of this method is a region capable of measurement is so limited that the movement

of the human operator is constrained.

2) Remote-sensing method

Magnetic, infrared or sonar sensors are used to measure a distance to the head of the human operator to locate the position. The advantage of this method is the movement of the operator is not constrained. This enables us to locate the position of the operator in a broader region. Its disadvantage, however, is one approach using infrared or sonar sensors is weak in occlusion phenomena, and the approach using magnetic sensors is not robust to a noisy environment where metals or sources of electromagnetic waves exist.

3) Internal-sensing method

This method utilizes giro sensors and accelerometers to measure values of angle velocities and accelerations, and integrates the values to determine a position and a direction of the operator's head. The advantage of this method is a speed of measurement is high and the movement of the operator is not constrained. Thus, the region capable of measuring is not limited. On the contrary, the disadvantage of this method is errors of measurement are easy to be accumulated due to the integral calculation of this method.

On the cockpit we are now developing,

- A movement of a human operator is mainly limited to the region of his head or upper part of his body.
- A cycle of controlling a camera platform mounted on the robot is at the high speed of 5 milliseconds.
- As is shown in Fig.1 on the left, there are sources of electromagnetic waves such as master arms in the cockpit.

Considering those three items, we adopt the mechanical link method noted by 1). **Table 2** shows the performance list of a head-tracking system we adopted.

Table 2. Specification of the Head-tracking System

Measurement Cycle	300Hz (Maximum)
Time Delay	Lower than 2msec (average)
Spatial Resolution	About 0.63mm
Measurement Precision	0.5cm
Measurement Region	91cm (diameter) × 45cm (hight)
Signal Output	RS-232C

On the other hand, a **HMD**, a head mounted display gives us the feeling of being surrounded in an actual site, when a field of view is wider and a resolution is higher. Such a HMD with wide field of view and high resolution will make the HMD increase in weight. Therefore, the HMD forces a weight burden on a user of the HMD. Considering the following two points in this study,

- A horizontal field of view, effective to a human binocular parallax, should be at more than 45 degrees and a horizontal field of view, effective to a human mobile parallax, is set at about 60 degrees [11].
- The maximum view angle of each camera mounted on the robot is at about 50〜60 degrees [12].

we adopt a HMD system that has 50〜60 degrees in the horizontal field of view. Furthermore, in the cockpit, a human operator faces at two phases: One is called "**locomotion control phase**" where the operator controls the two legs of the slave-robot to walk. And the other is called "**manipulation control phase**" where the operator controls master-arms and hands to manipulate an object in a slave-robot site with slave-arms and hands of the robot. In the locomotion control phase, the operator looks at the real images presented on the surrounded visual display. In the manipulation control phase, on the other hand, the operator sees the real images presented on the HMD. Considering the operation of both phases sifting from each other, it is desirable that an HMD system should install a mechanical device, which leaps up and down a glass portion shown in **Fig.6**. **Table 3** specifies the **HMD** we adopted.

Fig.6 HMD with the glasses Up and Down

Table 3. Specification of the HMD

Display Method	LCD (Active matrix TFT)
Resolution	921,600
Image Input	VGA
Field of View	48° (H) ×36° (V)
Weight	About 900g
Others	Glass portion leaped up and down, Distance between eyes can be changed

Integrating the HMD with the head-tracking device, we have developed a HMD system with a head-tracking function as is shown in **Fig.7**. The developed system has a counter-balancing mechanism so that a user will not be bothered to feel a weight of the HMD he is wearing. As is shown in the figure, three axes are installed in the top portion of a head, the left and right portions of ears, and the top portion of the head, respectively. These axes enable the user to move his head in pan, tilt, and roll directions.

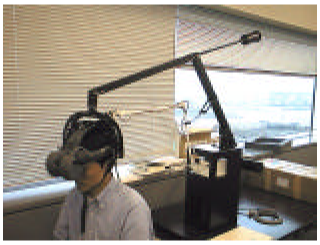

Fig.7 The Developed HMD System with Head-tracking Function

4. Surrounded Audio Display System

A surrounded audio display system consists of 8 speakers and a headphone the human operator is wearing. A 3-dimensional microphone system mounted on the robot detects a sound signal around the robot. The sound signal is displayed on the 8 speakers and the headphone. **Fig.8** shows an overview of the developed stereo audio display system.

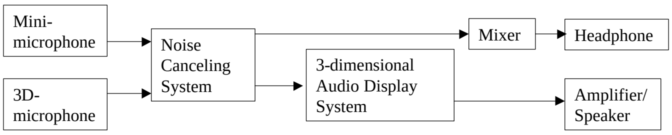

Fig.8 A Block-diagram of the Stereo Audio Display System

5. Man-Machine Interface System with Augmented Reality

5.1 The Needs of the Augmented Reality Function

When the human operator tries to assess situation around the robot, he keeps looking at the stereo images presented on the surrounded visual display system or the HMD with head-tracking function. If he is given such invisible information as geometric size, mass etc. of objects in the robot site, he can easily recognize the situation in order to control the robot. Thus, the augmented reality technique [13] is employed in the visual display system, which enhances the usage of the man-machine interface in the cockpit. If the man-machine interface can provide the operator with information he wants to have at the right time and in a way he can easily understand, the man-machine interface becomes much more friendly than ever.

5.2 Design Principle of Man-Machine Interface

We classify the behaviors of the human operator into three categories; “situation assessment”, “planning”, and “execution of a plan”. These categories correspond to primitive units of human cognitive behaviors. Also, we analyze transitions among these primitive units to specify a man-machine interface for the human operator. Specifications are analyzed and drawn by the approach of developing software called UML, Unified Modeling Language[14].

Fig.9 shows the state transition diagram of a man-machine interface that is now being developed. There are the two modes of a locomotion control mode and a manipulation control mode. Each has three main processing-states; a state where the human operator assesses situation around the robot, a state where he requests a simulation before actual control of the robot, and a state where he actually controls the robot.

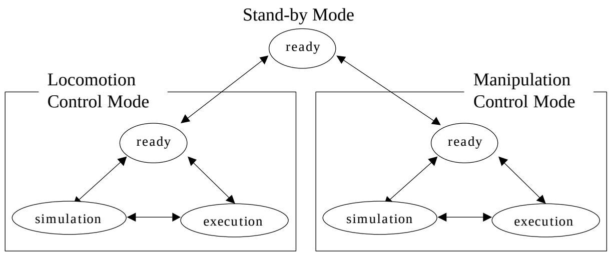

Fig.9 A State Transition Diagram of a Man-machine Interface

5.3 Total Hardware Architecture of Man-Machine System

Fig.10 shows total hardware architecture of a man-machine interface we have developed. Real images captured by multi-camera systems mounted on the robot are sent into an augmented reality system, discussed in the following section, through multi-analog wireless modules. The augmented reality system has the input images digital-processed, and sends digitized images into the surrounded visual display and the HMD. There are two types of multi-camera systems; “a multi-camera system with a wide field of view” for the surrounded visual display system and “a binocular stereoscopic multi-camera system” for the HMD. The binocular stereoscopic multi-camera system has installed a camera platform capable of controlling a camera direction with 2 degrees of freedom, pan and tilt rotations. Thus, the camera direction is remotely controlled according to the head movement of the human operator.

Software system, which controls a man-machine interface as is mentioned in **5.2**, is embedded in a “controlling PC (Kayak Windows PC)” presented in Fig.10. The controlling PC is in charge of controlling the HMD system with head-tracking function. Also, it is in charge of communicating with the augmented reality system and the surrounded audio display system through TCP/IP protocols. Moreover, this controlling

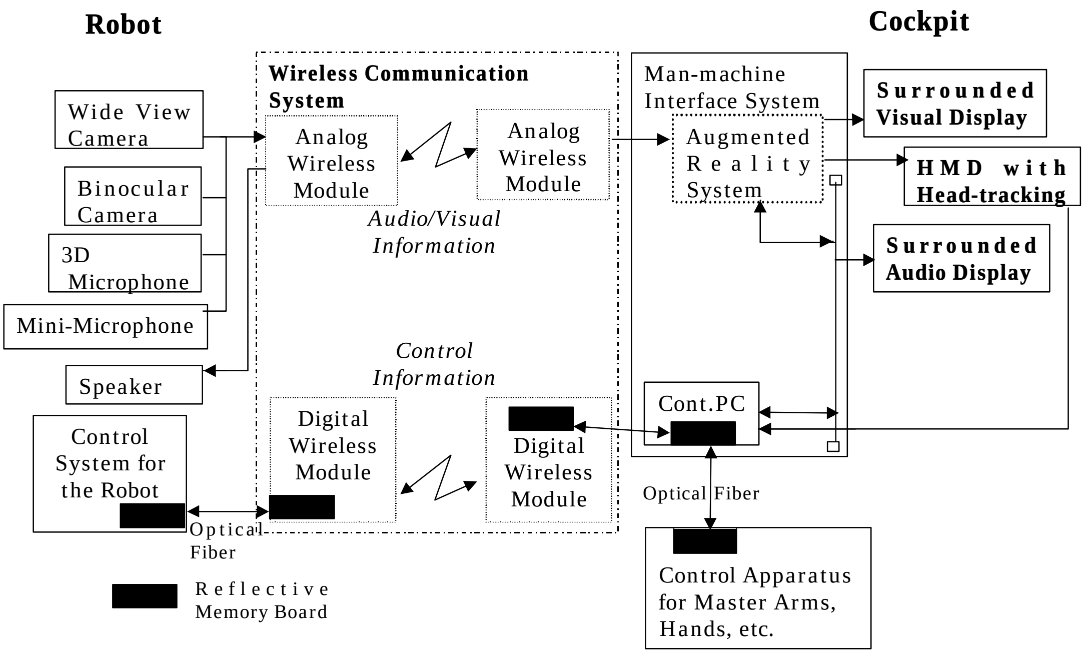

Fig.10 Total Hardware Architecture of the Man-machine Interface

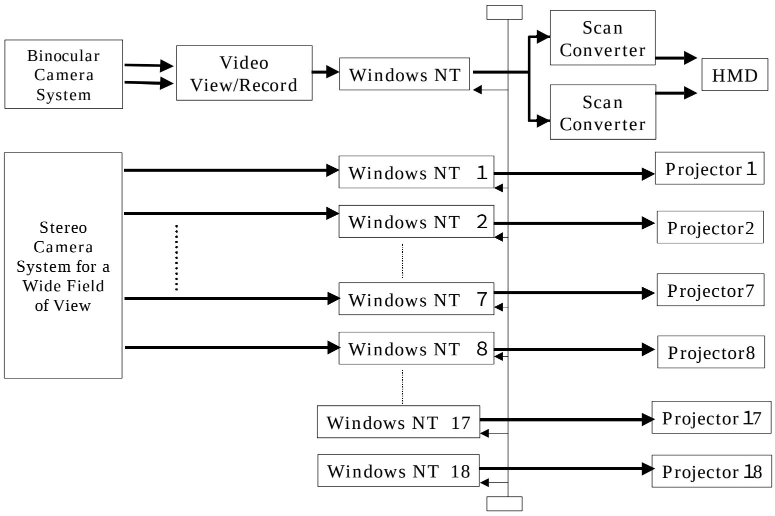

Fig.11 Architecture of the Augmented Reality System

PC sends information on controlling the robot or master-arms and hands through a reflective memory module that is installed in this controlling PC.

Fig.11 shows an overview of hardware architecture, which realizes an augmented reality system. The stereo images captured by the binocular stereoscopic camera system on the robot are sequentially given to a Crystal Eyes Video View/Record unit. This unit integrates both left and right images into one field, and sends a series of field data into a graphical computer (SGI 320 Windows NT). This graphical computer integrates the input field data with other CG data sent by the controlling PC, and inputs the integrated field data into two scan-converters. These two scan-converters cut both left and right field data, respectively, to send each field data to the HMD (ProView 60). On the other hand, the stereo real images captured by the multi-camera system for the wide field of view are also sent to a group of graphical computers (SGI 320 Windows NT). These computers also integrate the input images with other CG information sent by the controlling PC, and present the integrated images to corresponding projectors.

5.4 Implementation of Man-Machine Interface

Fig.12 illustrates an example of the man-machine interface when the operator is in charge of locomotion control phase. As is previously mentioned, the real images captured by the multi-camera system for the wide field of view are presented on such 4 screens of the surrounded visual displays put to the left, right, center and bottom sides. An operational menu for the operator is presented on a left-bottom screen of the display. On a right-bottom screen, a software robot that partially simulates the real slave-robot is placed. The operator manipulates a 3-dimensional mouse device (Magellan Plus), to control the operational menu as is shown in **Fig.13**. In short, he sets a goal position and direction of the robot and requests a simple path planning to the goal point. And also he commands a simulation of locomotion to reach the goal point, and has the real robot to execute locomotion. When the simulation is carried out, the software robot presents itself and starts walking on the associated screen.

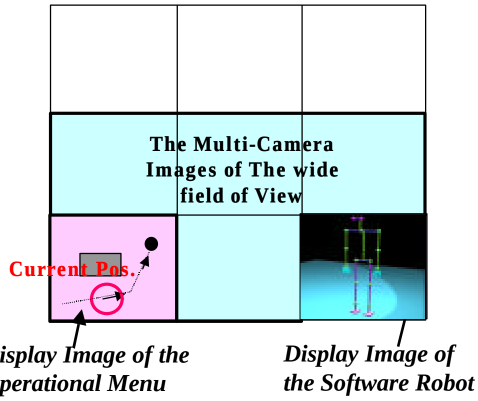

Fig.12 Illustration of the Man-machine Interface —Locomotion Control Phase—

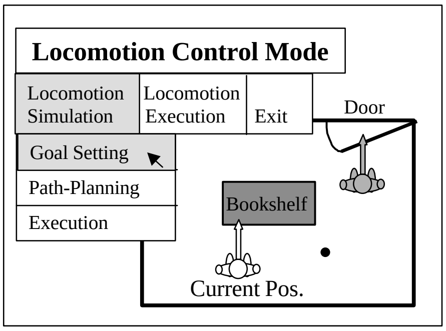

Fig.13 An Example of the Man-machine Interface —Locomotion Control Phase—

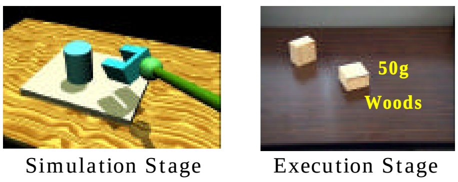

Fig.14 An Example of the Man-machine Interface —Manipulation Control Phase—

On the other hand, when the operator is in charge of manipulation control phase, he looks at the images presented on the HMD. **Fig.14** illustrates the example of the man-machine interface in the manipulation control phase. In the simulation stage, the operator manipulates both left and right master-arms and hands to control VR arms and hands presented on the HMD. Also, in the actual stage of its execution, the operator looks at the real images captured by the binocular stereoscopic camera system as is shown in the figure on the right. The operator manipulates the master-arms and hands to control the real slave-arms and hands of the robot. In this execution stage, invisible information is super-imposed on the real images as is shown in the figure on the right.

6. Conclusion

In this report, we describe the audio-visual display system embedded in the humanoid robotics system. This audio-visual display system, even though under development so far, will complete the whole humanoid robotics system in our further researches.

Acknowledgement

This study is being conducted by MSTC, Manufacturing Science and Technology Center as part of the “Humanoid Robotics Projects”, for the NEDO, New Energy and Industrial Technology Development Organization, under the ISTF, Industrial Science and Technology Frontier Program of AIST, Agency of Industrial Science and Technology in MITI, Ministry of International Trade and Industry.

References

[1] Research and Development of Humanoid Robotics System, Annual Report in 1998, MSTC, (1999) (in Japanese).

[2] Akin, D.L. Minskey, M.L. et al., Space Application of Automation: Robotics and Machine Intelligence Systems (ARAMIS)- Phase II, *Telepesence Technology Base Development*, NASA Contact Report 3734 (1983).

[3] Tachi, S., Tanie, K., Komoriya, K. and Kaneto, M., Tele-Existence (I): Design and Evaluation of a Visual Display with Sensation of Presence, *Proc. of the 5th Symposium on Theory and Practice of Robots and Manipulators (RomanSy ’84)* (1984) 245-254.

[4] Tachi, S., Real-time Remote Robotics –Toward Networked Telexistence, *IEEE Computer Graphics and Applications*, Vol.18, No.6 (1998) 6-9.

[5] Hoshino, H., Nishiyama, T., Nakajima, R., Sawada, K., Development of Surrounded Projection Based Display, *Proc. of the Virtual Reality Society of Japan, 4th Annual Conference*, (1999) 147-148 (in Japanese).

[6] Takeda, T. and Kaneko, T., Effect of Body Sway by Using Visual Wide-Field Images, *The Journal of the Institute of Television Engineers of Japan*, Vol.50, No.12 (1996) 1935-1940 (in Japanese).

[7] Hatada, T., Artificial Reality with Visual Effects, *Bulletin of the Japan Soc. Prec. Eng.*, Vol.57, No.8 (1991) 1330-1334 (in Japanese).

[8] Shibano, N., Hatanaka, T., Nakanishi, H., Hoshino, H., Nagahama, R., Sawada, K., and Nomura, J., Development of VR Presentation System with Spherical Screen for Urban Environment Human Media, *Transactions of the Virtual Reality Society of Japan*, Vol.4, No.3 (1998) 549-554 (in Japanese).

[9] Nozawa, H., Hashimoto, W., Iwata, H., A Study about a Production Method of the Spherical Surface Display, *Proc. of the Virtual Reality Society of Japan, 4th Annual Conference*, (1999) 157-160 (in Japanese).

[10] Cruz-Neira, C. et al., Surrounded-Screen Projection-Based Virtual Reality, *Proc. of SIGGRAPH'93*, (1993) 135-142.

[11] Maeda, T., Head-mounted Display for the Study on Tele-existence, Japanese Journal of Optics, Vol.25, No.1 (1996) 14-20 (in Japanese).

[12] http://www.sony.co.jp/soj/BizPartnaers/ISP/lineup/Color-j.html#EVI370

[13] Rekimoto, J. A Tutorial Introduction to Computer Augmented Environments, *Computer Software*, Vol.13, No.3 (1996) 4-18 (in Japanese).

[14] Brue Powel Douglass, Real-time UML, Addison-Wesley (1998)