///VSP///
Advanced Robotics 22 (2008) 1053–1073
ADVANCED ROBOTICS
www.brill.nl/ar

Full paper

TORSO: Development of a Telexistence Visual System Using a 6-d.o.f. Robot Head

Kouichi Watanabe [a,*], **Ichiro Kawabuchi** [b], **Naoki Kawakami** [a], **Taro Maeda** [c] and **Susumu Tachi** [a]

[a] University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan
[b] Kawabuchi Mechanical Engineering Laboratory Inc., 2-10-1 Kamata, Ohta-ku, Tokyo, Japan
[c] Osaka University, 2-1 Yamadaoka, Suita, Osaka, Japan

Received 26 February 2008; accepted 14 April 2008

Abstract
In telexistence master–slave systems, it is important to transmit visual information from remote places to the operator. Conventional imaging devices in head-mounted displays (HMDs) can only express the three-axis rotation of the neck. However, humans can obtain broader visual fields and motion parallax information from the translational motion of their necks. We have proposed a system that can acquire natural and comfortable visual information, and can accurately track the head motion of a person. Our proposed device can express the head motion and the translation movements of the neck. We have developed a robot, called 'TORSO', and constructed a telexistence visual system with a display device, HMD. In this paper, by means of a broader field of view achieved by motion involving looking around, we demonstrate the advantage and novelty of our proposed system. In addition, we suggest the evolution of the TORSO–HMD system.

Keywords
Telexistence, master–slave, 6 d.o.f. robot head, torso, motion parallax

1. Introduction

Telexistence [1], which has been the focus of our research efforts, refers to technology that allows one to experience advanced, realistic sensations of being in a remote environment by using remote operation and remote communication. In telexistence, there is a master–slave relationship between the operator and remote robot. Information from the operator is transmitted to the remote robot and information about the remote environment is returned by the robot to the operator. In

* To whom correspondence should be addressed. E-mail: Kouichi_Watanabe@ipc.i.u-tokyo.ac.jp

 DOI:10.1163/156855308X324767

order to achieve these goals, it is necessary to accurately reconstruct the information of the operator's position and posture in the remote environment. Meanwhile, it is also necessary to accurately portray the information of the remote environment for the operator; visual information, in particular, constitutes a major part of this information. Therefore, it is indispensable to accurately acquire and display this information in order to develop a telexistence system.

At present, most imaging devices [2] mounted on remote robots for binocular head-mounted displays (HMDs) [3] comprise two stereo cameras with 3 rotational d.o.f.; they are positioned at the same level as that of the human neck. In addition, an imaging device for a projection-based virtual reality system such as CAVE [4] was proposed in Ref. [5]. According to this method, the orientation of the cameras must be maintained, regardless of the orientation of the operator's head. The device — called a 'constant-oriented link' — proposed in the cited work can solve the problems involved in constructing an exact head-tracked stereoscopic display using fixed-screen display systems. This is due to the changes in the shape of the viewing volume and the rotation of the camera direction in relation to the operator's head motion. However, the goal of our project is to create the sensation of being in a remote place. Therefore, it is difficult to achieve this goal with a projection-based system wherein the operator can still see his/her own hands and legs.

In contrast, there are many head systems that have 3 rotational d.o.f. at the neck. For example, HRP [6], ASIMO and TELESAR [7], which was developed in our laboratory, are archetypal humanoid slave robots that have robot heads with 3 rotational d.o.f. Baier *et al.* put forward a multi-modal teleoperation system with HMD-based stereo vision [8, 9]. In the cited works, they carefully treated the problem of network communication and suggested that stereo vision is effective in tasks involving manipulation. Shiratsuchi *et al.* put forward a 4-d.o.f. robot head system for teleoperation in Ref. [10]. They defined a new measure of visual error and estimated the minimal necessary d.o.f.

Translational motion of the human neck broadens the visual field and provides motion parallax information. However, conventional imaging devices for HMDs cannot express this translation motion, which causes the operator to experience image mismatch. In fact, in order to acquire motion parallax information at the same level, the movement of the entire robot is needed. This is not practical with large humanoid robots and will serve no practical use in everyday working scenarios. In addition, the human neck can translate even when the lower limbs are stationary. This motion is more efficient than moving the entire body, including the lower limbs, because, in this case, only the head is moved. In addition, complex manipulation tasks might be required in cramped locations. Thus, if the robot head can be independently moved parallel to the ground, objects can be observed from a broader range of positions, consequently making manipulation tasks easier. The HRP solves this problem by using an additional camera that captures images of the surroundings, although the operator still cannot experience motion parallax.

We have developed a system that can acquire visual information in a natural and comfortable manner, and accurately mimic the head motion of a person. We have developed the neck and torso of an all-in-one robot that not only has 3 rotational d.o.f. of the neck, but also has 2 rotational d.o.f. and 1 translational d.o.f. of the waist. This robot is called 'TORSO'. TORSO can almost completely express the upper-body motions of humans, including the translational motion of the neck. The motion of looking around an object is one of the most important motions achieved by the translational motion of the neck; without moving the robot, the flanks of an object placed in front of the robot can be recognized. In Fig. 1, we show the motion of looking around an object.

In this paper, we propose a telexistence visual system that consists of TORSO and an HMD. We developed the Mutual Telexistence system with TORSO in 2004 [11]. The purpose of the present study is to represent a broader visual field for motion involving looking around and to demonstrate the effectiveness of this system.

Even though many telexistence systems have been developed, the realization of 'realistic sensation' and 'presence' has not yet been achieved. Our system can achieve these features with regard to movement. By using a remote robot that can accurately reflect the head motion of the operator, the operator obtains more realistic visual information. Moreover, in the remote environment, since the head motion of the operator is accurately represented by the remote robot, the robot can achieve the presence of this movement. The remainder of this paper is organized as follows.

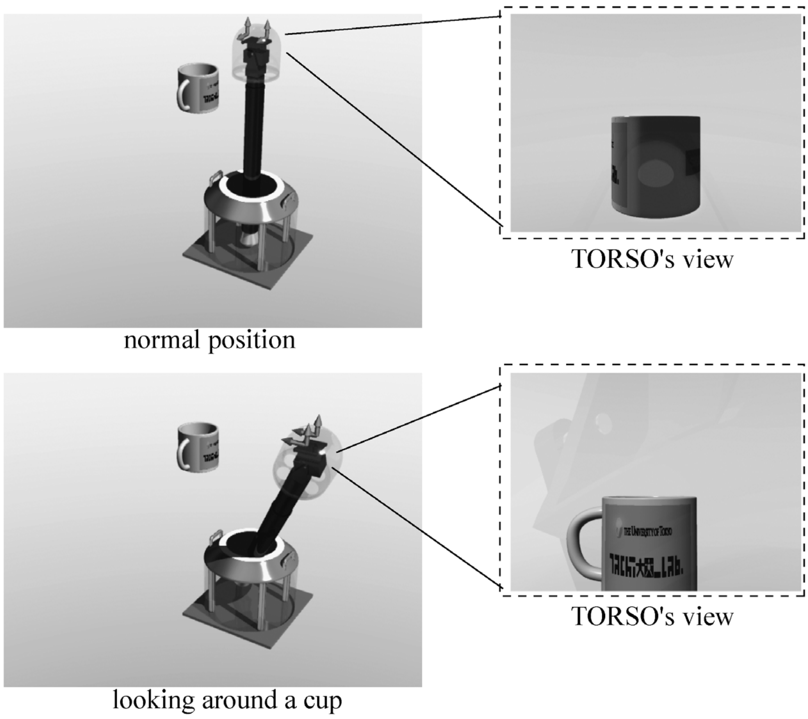


Figure 1. Motion involving looking around an object.

In Section 2, we present the specifications of the proposed TORSO system and the details of its components. In Section 3, we evaluate the task performance of the proposed system. In Section 4, we present the results of the implementation of this system. Then, in Section 5, we discuss the evolution path for the TORSO system in the near future. Finally, in Section 6, we present our conclusions.

2. TORSO–HMD System

2.1. Overview of the Proposed TORSO–HMD System

Figure 2 shows an overview of our telexistence visual system including TORSO. This system has a master–slave configuration. At the master side, the operator's head position and posture are measured by a sensor, e.g., a potentiometer. Further, the operator wears the HMD to view images of the remote environment. The position and posture of the operator's head are calculated by kinematics using the data captured by the sensors, which are transmitted to the slave side. At the slave side, the reference angle of each joint is calculated by inverse kinematics using the data from the master side. At the same time, the present angle of each joint is captured by a sensor (e.g., encoders), and the position and posture of the slave head follow those of the operator using these data and motor controls. In addition, the images of the remote environment captured by the two cameras are transmitted to the master side.

2.2. TORSO Configuration

A general view of TORSO is shown in Fig. 3. The TORSO robot mechanism consists of two main parts: a 3-d.o.f. robot head that includes a binocular camera and

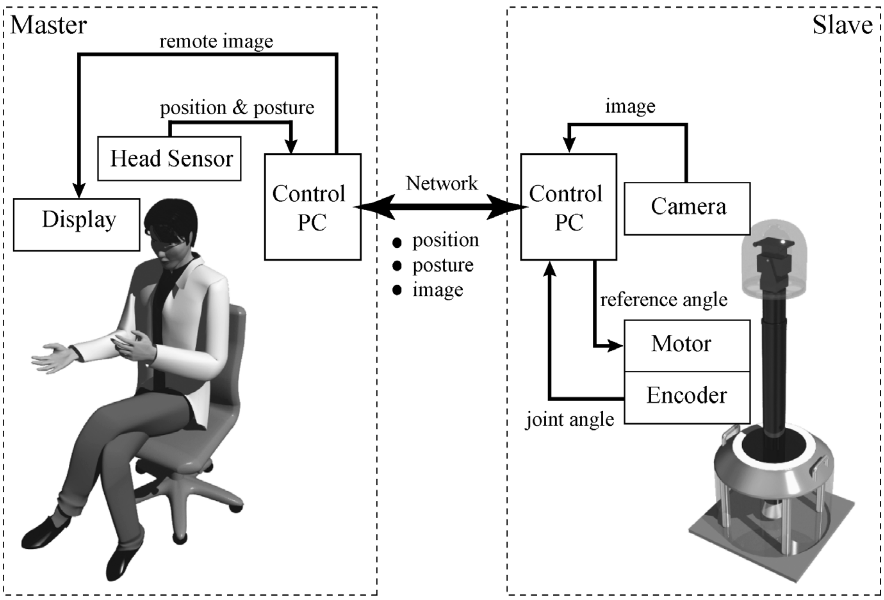


Figure 2. TORSO system diagram.

Figure 3. General view of TORSO.

a 3-d.o.f. robot body that suspends the robot head. In the design of the overall appearance of the robot, we have considered many points: adequate range of motion in spatial positioning of the head, sufficient agility for tracking the master motion of the operator's head, slenderness and lightness for high portability, and a small footprint and good mass balance for location independence. Conventional imaging devices in HMDs can achieve only three axes of rotation of the neck. Our device cannot only express the 3 rotational d.o.f. of the neck, but also its translational motion. Therefore, it is possible to achieve nearly human-like head motion in a remote environment.

For enhanced realism and improved workability in teleoperation, it is very important that the three-dimensional vision transmitted to the operator includes motion parallax in addition to binocular parallax. The most novel feature of this robot is that it can acquire motion parallax information without moving its entire body. That is, TORSO can express the upper-body motion of humans in a more natural manner and more accurately reflect the characteristics of humans than conventional robots. The dimensions and available ranges of each joint of our device are shown in Fig. 4.

The 3 rotational d.o.f. of the head are arranged such that the axes intersect at one point. The 2 rotational d.o.f. of the waist are also arranged in the same manner. This

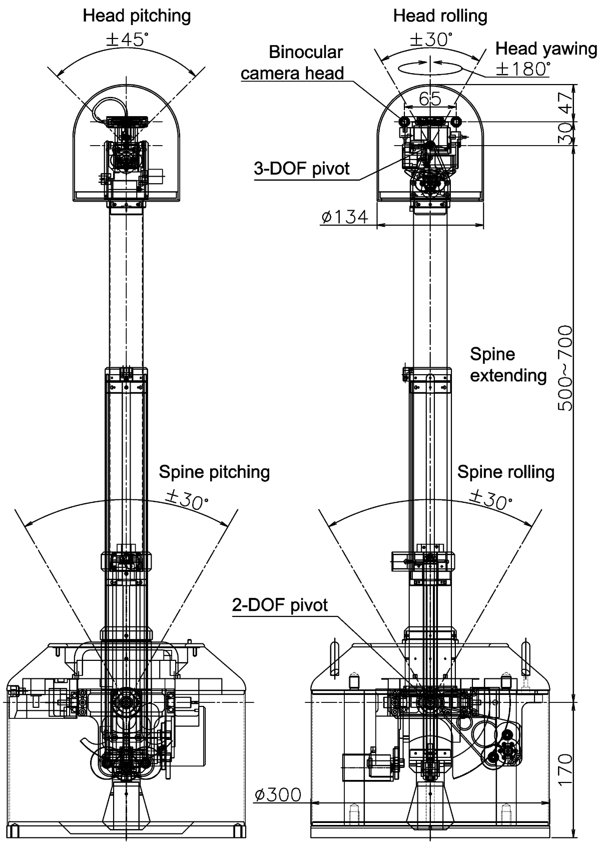


Figure 4. Schemtic of TORSO.

allocation of d.o.f. is almost equivalent to that in the human body. The translation d.o.f. of the waist not only enables stretching motion, but also resolves the difference in body height among individuals. Therefore, anyone can use the TORSO system and employ the robot sight as if it were his/her own. The range of motion of each joint of TORSO is slightly narrower than that in humans; however, it would be sufficient if the operator simply performed natural movements. Note that the word 'natural' means that the expression of 6-d.o.f. motion is possible across almost all ranges of the upper-limb motions of humans, i.e., the operator is free to assume any general posture for desk work while sitting on a chair. This, in turn, implies that the operator's upper-limb motion is expressed accurately across almost all motion ranges as long as the operator does not perform stiff movement.

TORSO has a specific arrangement of its 6 d.o.f. As shown in Fig. 5, the pan axis is allocated at the end of the d.o.f. As mentioned earlier, we assume that motion parallax is a very important factor in an active visual system for teleoperation and, therefore, focus on it. Our aim is to make panning easier than the other motions because it dominates motion parallax.

Each joint has a DC motor, an encoder and a photo switch. The photo switch determines the zero position of the encoder. Two types of DC motors (20 and 6 W) are used. The former are attached to the joints of the waist and the latter to the

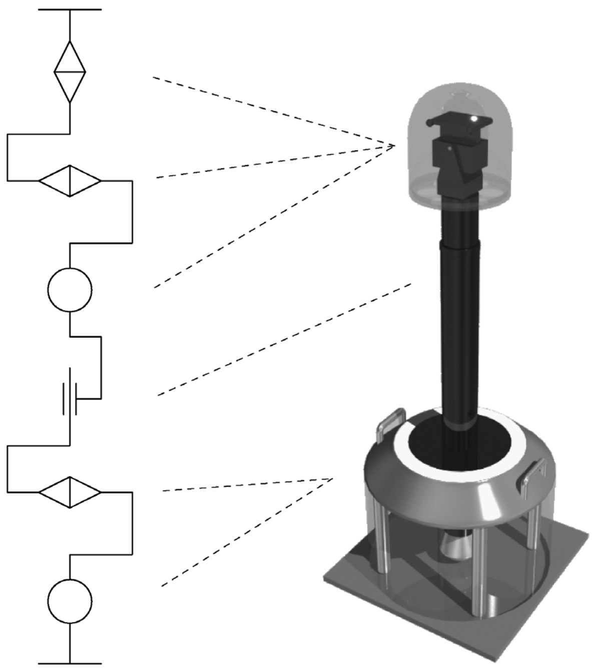


Figure 5. The d.o.f. arrangement of TORSO.

other joints. The driving performance of each joint is adequate because each joint can move at a speed higher than the maximum speed of motion of the human head. Furthermore, the weight of the head section, about 0.5 kg, is much lower than that of a conventional robot head. Two CCD cameras (diameter 7 mm) are installed in the head with a separation distance almost equal to that between the human eyes (65 mm). The angle of view of the cameras is 46° along the horizontal axis, a resolution of 768×494 pixels and a shutter speed of 1/60 s. Moreover, not only the weight of the head part but also the total weight of TORSO is less as compared to that of conventional robot head systems (less than 25 kg).

Figure 6 compares the motion range of human joints with those of our device. The expansion of the field of view achieved by the additional d.o.f. of TORSO is shown in Fig. 7. The spine has a maximum length of 0.7 m, and roll and pitch ranges of 30°; therefore, the head can reach a point at a horizontal distance of 0.35 m from the central line of TORSO. As a result, TORSO has a broader field of view than conventional robot heads because it can be positioned anywhere within an area defined by a circle of radius 0.35 m, as shown in Fig. 7 (top view). Furthermore, the field of view of it can be extended vertically. The extension region in this direction is also shown in Fig. 7 (front view). This region can support not only differences in operator height, but also differences in their stretching motions.

2.3. HMD Configuration

HMDs or head-mounted projectors (HMPs) [12] and other projection-based virtual reality systems such as CAVE have been developed. Among these devices, the HMD is the most appropriate for the robot head. When constructing a telexistence

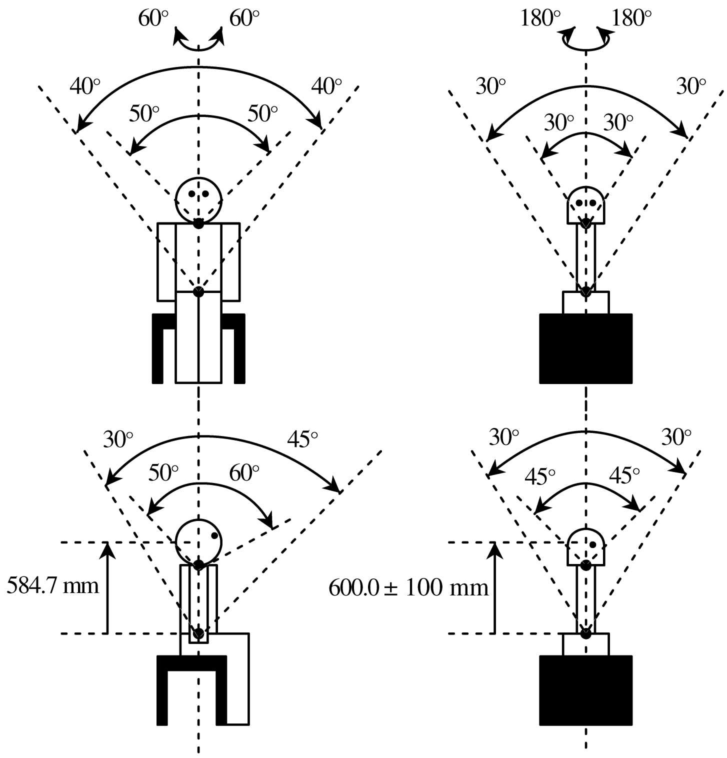


Figure 6. Motion ranges of the upper body of humans and TORSO.

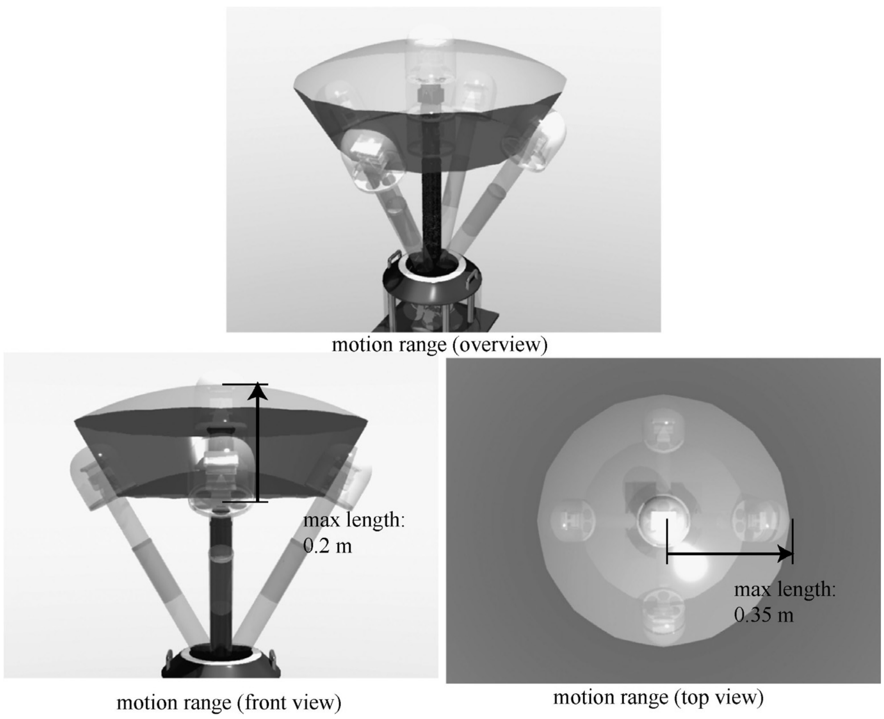


Figure 7. Extension of the field of view.

Figure 8. General view of the HMD.

visual system, the display device should be designed adequately such that it can naturally represent the remote environment. We have designed and developed the most appropriate HMD for TORSO according to the theory in Refs [13, 14].

Figure 8 shows an overview of the HMD that we have developed. It is not suspended and weighs approximately 2 kg. Further, it has two 7.1-inch liquid crystal displays on its left- and right-hand sides, each with a resolution of 1024 × 768 pixels and a response time of 30 ms. Images from the cameras mounted on TORSO are projected onto the operator's eyes through two mirrors. Two lenses are set in front of the operator's eyes. The following parameters are set on the basis of Ref. [14]. The distance between the operator's eyes and the LCD is 170 mm, and the focal length of the lenses is 200 mm (5.0 diopter), from these parameters, the accommodation depth cue is fixed at 1 m. Although the body of the HMD is lightweight because it is made from ABS resin, its center of gravity is somewhat anterior because of the weights of the LCD, mirrors and lenses. Therefore, we adjusted the center of gravity to the center of the head by putting the control board at the back of the head.

The distance between the lenses is almost the same as that between the human eyes (65 mm). Although this distance is different for each person, it is difficult to adjust it for each individual; besides, Ref. [14] shows that the influence of binocular perception is slight if the distance between the lenses of the HMD corresponds to the distance between the cameras on the robot head.

3. Evaluation by Simulation

3.1. Simulation Environment

As mentioned in Section 2, the proposed TORSO–HMD system has a broader visual field than a robot head with 3 rotational d.o.f. In this section, we evaluate the task performance of the proposed TORSO–HMD system and highlight the benefits of this system. We compare the task performance of the proposed system with that of a conventional 3-d.o.f. robot head system. For a fair evaluation, we assume a 3-d.o.f. robot head system on *XYZ* linear stages that act as the whole robot.

The aim of this evaluation is to prove that the task performance of the proposed 6-d.o.f. robot head system is higher than that of the conventional 3-d.o.f. robot head system. Therefore, we expect the task execution time of the proposed system to be less than that of the conventional system.

We constructed a robot simulator. The simulation environment is shown in Fig. 9. We used the OpenGL engine to create a CG model of TORSO and the environment, and used the Open Dynamics Engine (ODE) to simulate the dynamics of the robot and the environment. Our assessment shows that the simulation is suited to evaluating our system in the following respects: (i) in real system evaluation, we cannot distinguish between the mechanistic effectiveness of the arrangement of d.o.f. and the effect of system delay, and (ii) to develop the real system with a small delay about 1–2 frames is possible in the near future; therefore, there is some advantage in evaluating the system without system delay.

An HMD is placed on the subject's head, whose position and posture are captured by a head tracking system; this 6-d.o.f. data is then transmitted to a PC. A virtual environment is created on the PC; the space alignment of virtual objects is set beforehand. The virtual robot head reflects the position and posture of the subject's head as well as the positions of the obstacles and the target point created in the virtual environment. In the virtual environment, the position and posture of the virtual robot head are calculated from the data obtained from the subject's head to create the view images, which are displayed by the HMD, placing the subject in the virtual environment. The subjects are then given the task of locating the target point.

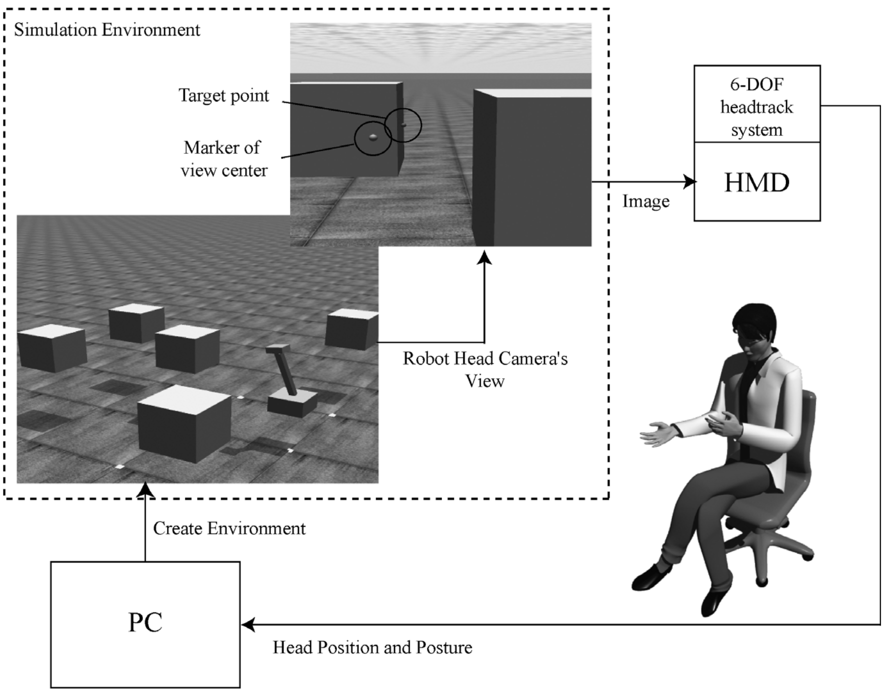


Figure 9. Experimental environment.

The virtual robot head has two operation modes, the 6-d.o.f. TORSO mode and the 3-d.o.f. + *XYZ* mode. In the 3-d.o.f. + *XYZ* mode, the robot head has 3 rotational d.o.f. at the neck and three translational d.o.f. at the base. In this mode, the translational motion afforded by the 3 d.o.f. of the base is controlled by an external controller, e.g., a pad controller. Other operation modes such as a 6-d.o.f. mode controlled by a 3-d.o.f. head tracking system with a pad controller and a 3-d.o.f. + *XYZ* mode controlled by a 6-d.o.f. head tracking system have been considered. However, the evaluation of these modes is not necessary due to the following: (i) the role of the 3 d.o.f. of the waist is to allow motion parallax without moving the entire robot. Thus, parallel motion of the entire robot can be realized by additional mechanisms. Therefore, system configurations such as those having the 3 d.o.f. of the waist for motion parallax being controlled by a pad controller have little meaning. (ii) Conventional robots have 3-d.o.f. heads and their *XYZ* movement is generally controlled by an additional input device. We think that it is important for the initial evaluation to compare such conventional configurations to those of our system. In addition, it is difficult to realistically control the *XYZ* movement mechanism supporting the entire robot with a 6-d.o.f. head tracking system.

Four patterns were created, each having the same spatial alignment of obstacles, but with the target point at different locations; one of these patterns was selected randomly. Figure 10 shows the space alignment of the obstacles and the target. The obstacles are five rectangles (dimensions of each rectangle $0.8 \times 0.8 \times 0.8$), with fixed positions. These obstacles are set around the virtual robot head and the target is set on a certain face of one of the obstacles.

The subjects were asked to begin searching for the target point as soon as they entered the virtual environment. A marker was placed at the center of the subject's

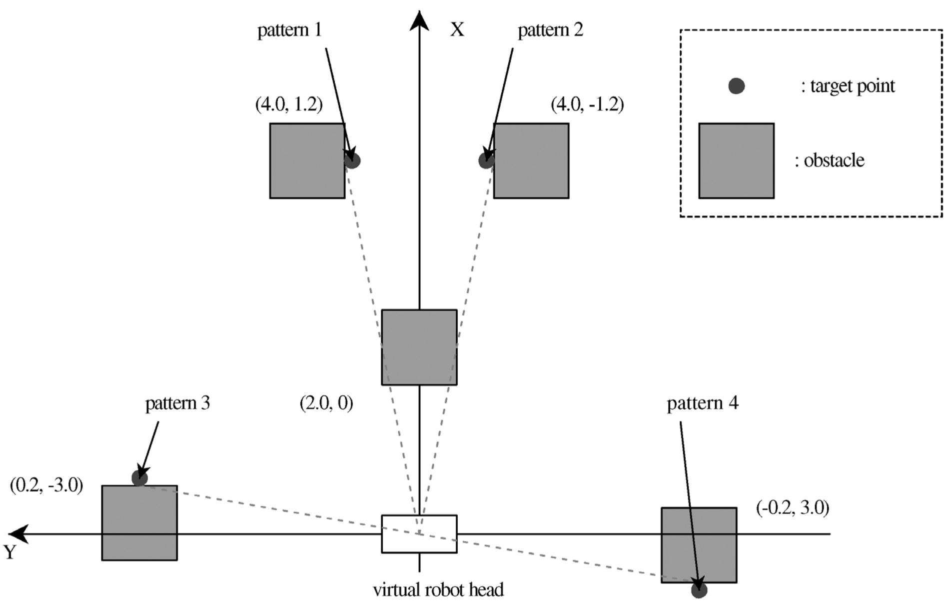


Figure 10. Experimental setting.

view and the subject was asked to move the marker over the target point to complete the task. Then, the time from the start to the completion of the task was measured.

The task execution times for 16 trials were collected: four trials of each of the four patterns. The above-mentioned trials were each run thrice and the results of only the last two runs were considered. In addition, we asked the subjects the following three questions. (i) Which modes were easy to work with? (ii) Did you experience motion sickness in any mode? (iii) Was the motion range in each mode sufficient?

The experimental data were collected from four subjects aged 20–30 years. The evaluation procedures were as follows:

(i) The subject runs the training module. The two operation modes of the virtual robot head and the four patterns of the virtual environment are consecutively displayed on the HMD.

(ii) The subject performs 16 trials in the 6-d.o.f. TORSO mode.

(iii) The subject performs 16 trials in the 3-d.o.f. + *XYZ* conventional mode using a pad controller.

(iv) The subject returns to Step (i) and repeats the procedure two more times.

The clock frequency of the control PC was 2.8 GHz and it had 1.0 GB of memory. A position sensor with 6 d.o.f. (ADL-1) was mounted on the HMD. The robot head was controlled by velocity control. The weight of the entire virtual robot head was about 25 kg, with the head section alone weighing about 1 kg. One loop step of the pad controller input was estimated to be 0.03 m/input from the measured velocity of human head motion. The parameters of the dynamics are the same in both modes.

3.2. Evaluation Results

The evaluation results are shown in Tables 1 and 2.

As seen Table 1, the average times of task execution in the 6-d.o.f. mode is obviously smaller than that in the 3-d.o.f. + *XYZ* mode for all subjects. In addition, from Table 2, it is observed that the standard deviation of task execution in the 6-d.o.f. mode is smaller than that in the 3-d.o.f. + *XYZ* mode for all subjects. Therefore, we conclude that it is possible to perform location tasks rapidly and stably by using the

Table 1.
Evaluation result (average time (s))

	Subject A	Subject B	Subject C	Subject D
6-d.o.f.	2.75	3.26	4.27	3.15
3-d.o.f. + *XYZ*	3.79	4.58	6.18	4.2

Table 2.
Evaluation result (standard deviation)

	Subject A	Subject B	Subject C	Subject D
6-d.o.f.	1.36	1.45	2.1	1.67
3-d.o.f. + *XYZ*	2.04	3.33	4.73	2.28

6-d.o.f. robot head system. All subjects answered that task execution in the 6-d.o.f. mode was relatively easier than that in the 3-d.o.f. + *XYZ* mode.

Further, all subjects answered that the motion ranges of the virtual robot head were sufficient in both modes. However, three subjects felt that the motion ranges in the 3-d.o.f. + *XYZ* mode were relatively larger than those in the 6-d.o.f. mode. Thus, the sensory motion ranges in the proposed head system may have seemed narrower than those of a conventional robot head system. One half of the subjects reported motion sickness in the 3-d.o.f. + *XYZ* mode. However, in the present case, where the virtual CG environment greatly differs from real environments, motion sickness occurs when the field of view is destroyed. In this evaluation, we use an HMD that has a large response time; therefore, it is possible to experience motion sickness in both modes. Further evaluation with respect to motion sickness needs to be carried out.

From the above results, it can be concluded that the proposed 6-d.o.f. robot head system is effective in performing task of locating a target point by using head motion parallax.

4. Implementation

In this section, we elaborate upon the broader visual field of TORSO. The positional data from the HMD are sent to the PC through a serial connection, while control of TORSO is achieved by a DA board that controls the motors, a DIO board that acquires the signal from the photo switch and a counter board that reads the encoder value. Images from the cameras mounted on TORSO are directly transmitted to the HMD without involving the PC. Therefore, the network delays in receiving the images are eliminated. Consequently, the synchronization between the operation and the camera is not taken into account by this system. In the following section, the cycle time of the control loop depends on the camera response time or the LCD refresh rate because the motion speed is very high. Therefore, an improvement in hardware performance is expected.

Further the measured position and posture data of the operator's head are used in the inverse kinematics calculation of the reference angles for each joint. Each joint torque is generated by a PID controller by using the reference and present angles. The loop time of the entire process is approximately 0.3 ms; this is very small as compared to the time of the control loop of a general robot system (in general, the required loop time is 1.0 ms for robot control systems). Therefore, very-high-speed tracking motion is possible because there is a little software overhead.

As seen in Fig. 11, TORSO can capture the side of an object in front of it when the 6-d.o.f. mode is enabled. In contrast, when only the 3-d.o.f. mode is enabled, TORSO cannot capture the side of the object. From the overview image of TORSO,

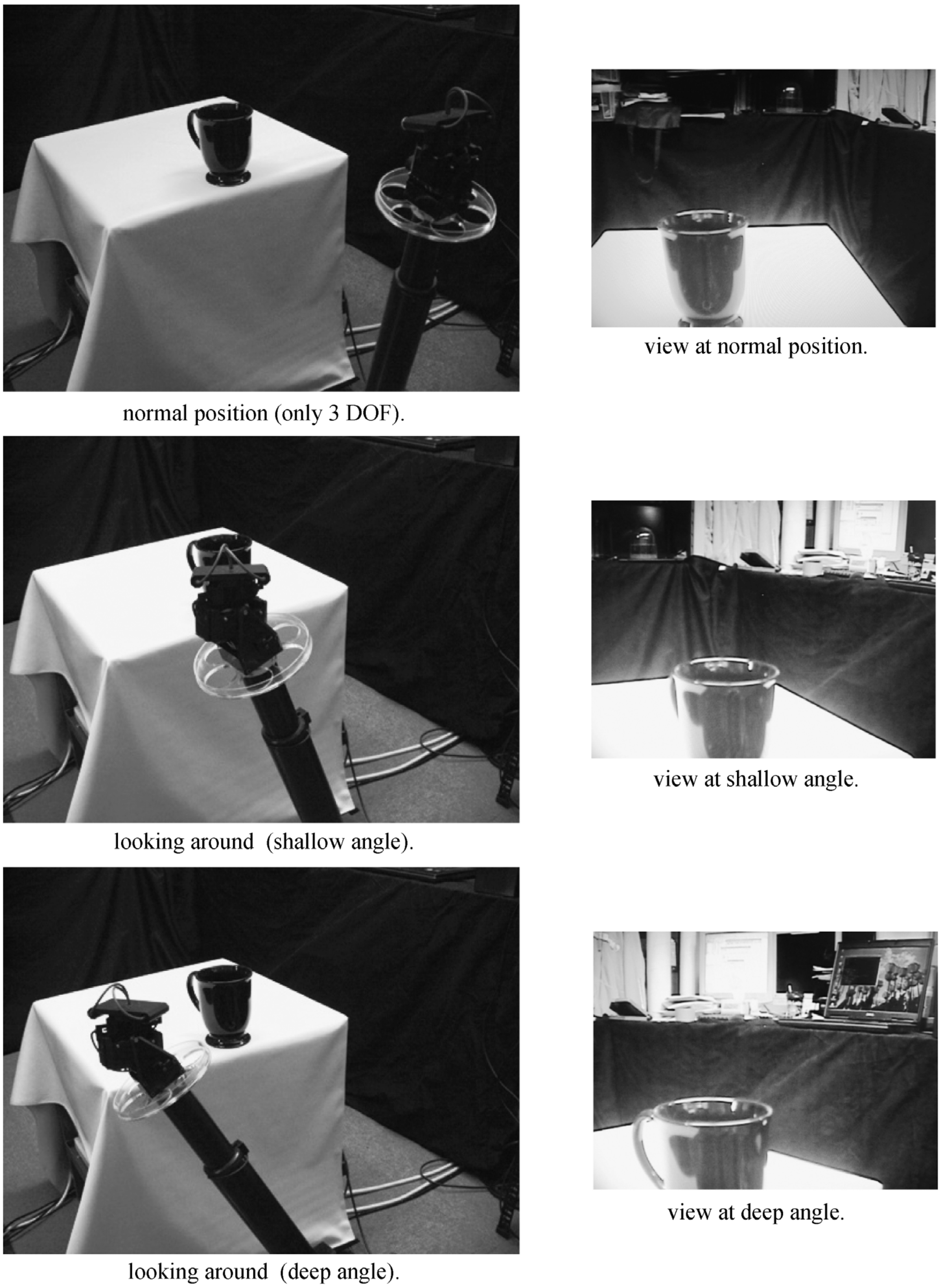


Figure 11. Actual motion and views of TORSO (looking around motion).

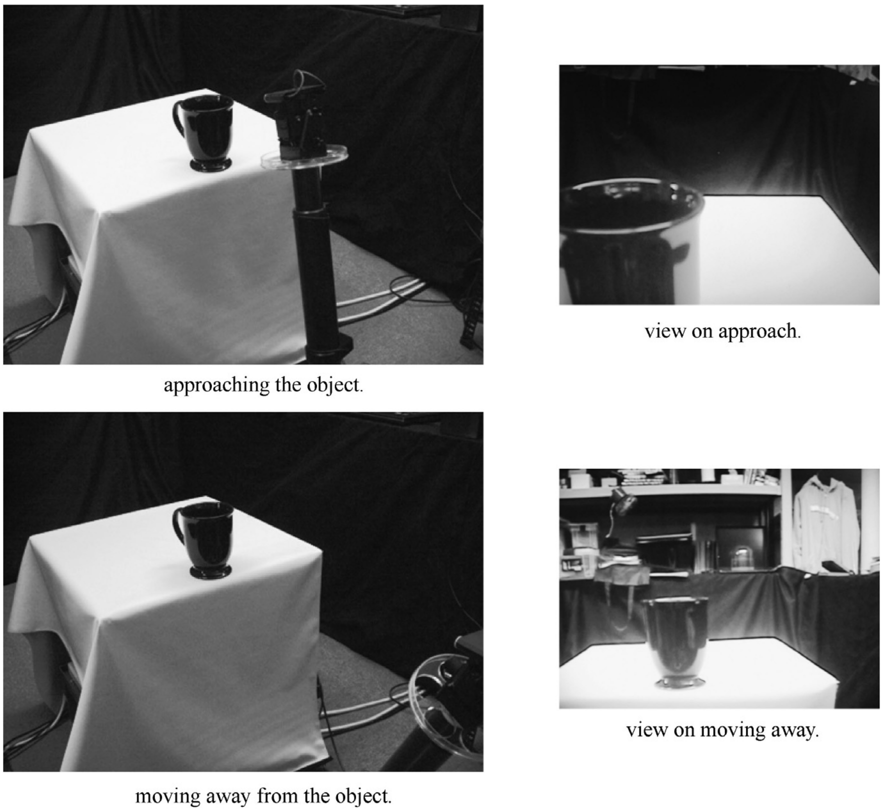


Figure 12. Actual motion and views of TORSO (back-and-forth motion).

it can be seen that the obstacle is located in the immediate vicinity of TORSO. TORSO can obtain the side information on the object without coming under the influence of such an obstacle. Furthermore, as shown in Fig. 12, TORSO can obtain the details of the object and an overall image the object and its surroundings by moving forward and backward, respectively.

At the 34th International Conference and Exhibition on Computer Graphics and Interactive Techniques (SIGGRAPH 2007), we demonstrated and exhibited our system [15]. Figures 13 and 14 are photographs of our exhibit at this conference. During the 5 days from 5 to 9 August, over 700 people experienced our system. From the feedback of many of the attendees, it was confirmed that this system would have numerous applications. In addition, we felt that our system should also have robot arms and hands. In the next section, we will suggest an application of our system with robot arms.

5. Evolution of the TORSO System

In the preceding sections, we presented the details of each component of the proposed visual system and elaborated upon the broader visual field of the TORSO robot. In this section, we will present a case of effective usage of this system. First, we consider that our system can apply to a communication system between humans

Figure 13. Exhibition at SIGGRAPH 2007 (demonstration).

Figure 14. Exhibition at SIGGRAPH 2007 (attendees).

in the sitting posture. As mentioned Section 1, we have proposed a communication system that uses two robot head systems with 6 d.o.f. each and HMDs. In this communication system, the 6-d.o.f. robot head is employed to acquire the feature information of the person facing it and to reflect the person's upper-body motion to the remote robot. TORSO is better suited for imaging devices from the human viewpoint because it is compact and lightweight and has a fast response. Further, communication that matches the visual axis of the users is easy to achieve by using the TORSO system.

Furthermore, the salient feature of the proposed robot is that the task of looking around an object can be achieved without moving the entire robot. This feature is effective for tasks involving manipulation. In the near future, we may be able to create a surrogate anthropomorphic robot that has a very high degree of realistic sensation and presence by attaching an arm that is light and easily maneuvered. The robot will execute manipulation tasks by using the attached arms, which will be easy because TORSO can perform looking around objects. If the target object for manipulation is fixed, then, in conventional robots, the operator can only see the front face of the object because he/she cannot move the target object. However, by using TORSO along with its arms, the operator can view the sides of the target object by the translational motion of the neck. Although the operator could also view the side of the target object with a conventional robot by translating the entire robot, the position and posture of the robot arms would also change. As a result, the operator will have to reconsider the manipulation procedure. The development and evaluation of TORSO with arms is an important future work.

The advantage of performing manipulation tasks using TORSO with arms is shown in Fig. 15. In the normal position, TORSO cannot precisely touch and grip the handle of a cup because it can only see the front. In this situation, the arms of the robot could collide with the cup or other objects. However, because of its ability to look around the cup, TORSO can view not only the front, but also the back of the cup, which allows it to precisely touch and grip the cup.

Several other applications of the TORSO–HMD system are considered. This device provides a large amount of information to telecommunication systems. For example, although remote meeting systems now primarily comprise video telephones, we will be able to include the feeling of actually participating in a meeting in a remote meeting room by using this system. Consequently, the meeting can involve invigorating debates due to the new aspects and features that could not be seen in conventional remote meetings. Moreover, in present shopping networks, we obtain information about commercial products only from pictures or statements. However, with our device, we will be able to obtain this information about commercial products by actually viewing them.

Furthermore, TORSO can be used to view oneself. TORSO is positioned at a distance from the operator, facing the operator. The operator can then see himself/herself from the point of view of a second party. The operator can experience a point-symmetric image with spatial information that is different from a mirror image. Although most remote-operation visual systems can achieve this task, TORSO achieves a more realistic sensation than conventional systems. Therefore, the proposed system will express the minute motions that cannot be expressed by conventional systems, becoming a new system that gives persons the chance to try to recognize themselves philosophically.

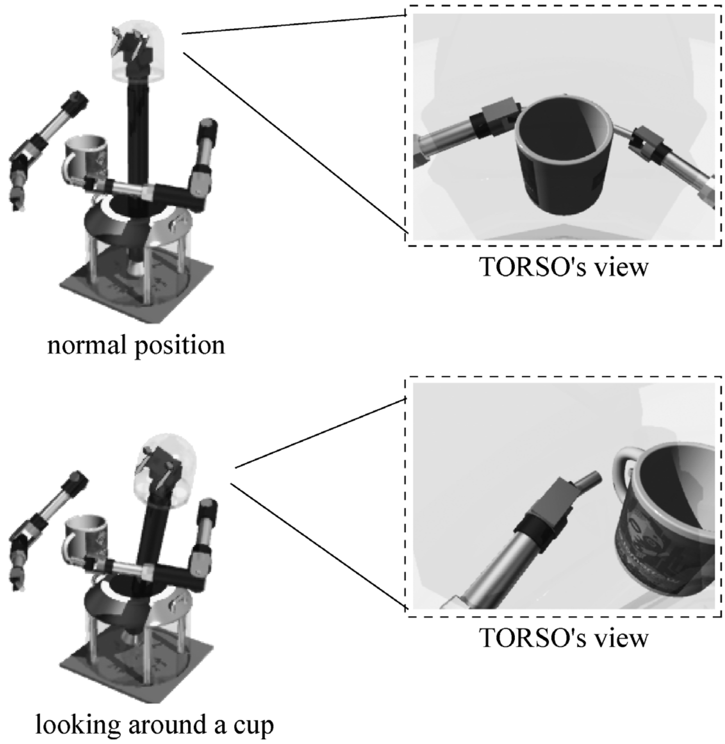


Figure 15. Enhancement of working efficiency.

The future goal of this project is the completion of an entire telexistence system. In the future, we should be able to treat robots just like ourselves; this will eventually lead to robot–human coexistence in society. Thus far, we have focused on achieving accurate head motion; our proposed device is focused on the acquisition and display of visual information that accurately reflects the motion of the human head.

6. Conclusions

Most imaging devices that are mounted on remote robots for binocular HMDs have two stereo cameras with 3 rotational d.o.f.; they are positioned at the same level as that of the human neck. However, humans can broaden their visual fields and obtain motion parallax information by the translational motion of the neck. In conventional slave robots with 3-d.o.f. robot heads, in order to obtain motion parallax information at the same level, the entire robot body must be moved. However, the human neck can translate even when the rest of the body is stationary. This motion is more efficient than moving the entire body because only the head is moved.

We proposed a system that can acquire visual information in a more natural and comfortable manner by accurately tracking the head motion of a person. We developed the neck and torso of an all-in-one robot that not only has 3 rotational d.o.f. of the neck, but also has 2 rotational d.o.f. and 1 translational d.o.f. of the waist.

We implemented the TORSO–HMD system and acquired images with broader fields of view captured by two cameras. We demonstrated and exhibited our system at SIGGRAPH 2007. From the feedback of many of the attendees, it was confirmed that this system would have numerous applications.

We presented some possible applications of this system, e.g., a communication system between humans in the sitting posture, a surrogate anthropomorphic robot with a lightweight arm that has a very high degree of realistic sensation and presence, remote meeting systems, shopping networks, devices for seeing ourselves, etc. The next step in our research will be to develop the slave head and arms, and to evaluate tasks involving manipulation.

The future goal of this project is the completion of an entire telexistence system. In the future, we should be able to treat robots just like ourselves; this would result in beneficial robot–human coexistence in society. We have focused on achieving accurate head motion, and our proposed device concentrates on the acquisition and display of natural visual information.

References

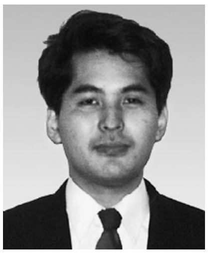
1. S. Tachi, K. Tanie, K. Komoriya and M. Kaneko, Tele-existence (I): design and evaluation of a visual display with sensation of presence, in: *Proc. RoManSy'84: The 5th CISM-IFToMM Symp.*, Udine, pp. 245–254 (1984).
2. C. DiSalvo, F. Gemperle, J. Forlizzi and S. Kiesler, All robots are not created equal: the design and perception of humanoid robot heads, in: *Proc. Designing Interactive Systems 2002 Conf.*, London, pp. 321–326 (2002).
3. I. Sutherland, A head-mounted three dimensional display, in: *Proc. Fal Joint Computer Conf. (AFIPS Conf. Proc.)*, San Francisco, CA, vol. 33, pp. 757–764 (1968).
4. C. Cruz-Neira, D. Dandin and T. Defanty, Surround-screen projection-based virtual reality: the design and implementation of the CAVE, in: *Proc. SIGGRAPH93*, Anahim, CA, pp. 135–142 (1993).
5. Y. Yanagida, A. Mitsuhashi, T. Maeda and S. Tachi, Implementation of fixed-screen-based telexistence visual system, in: *Proc. 9th Int. Conf. on Artificial Reality and Telexistence*, Tokyo, pp. 123–130 (1999).
6. S. Tachi, K. Komoriya, K. Sawada, T. Nishiyama, T. Itoko, M. Kobayashi and K. Inoue, Telexistence cockpit for humanoid robot control, *Adv. Robotics* **17**, 199–217 (2003).
7. S. Tachi, H. Arai and T. Maeda, Development of an anthropomorphic tele-existence slave robot, in: *Proc. Int. Conf. Advanced Mechatronics*, Tokyo, pp. 385–390 (1989).
8. H. Baier, M. Buss, F. Freyberger and G. Schmidt, Benefits of combined active stereo vision and haptic telepresence, in: *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Kagawa, pp. 702–707 (2000).
9. H. Baier, M. Buss, F. Freyberger and G. Schmidt, Interactive stereo vision telepresence for correct communication of spatial geometry, *Adv. Robotics* **17**, 219–233 (2003).
10. K. Shiratsuchi, K. Kawata, E. V. Poorten and Y. Yokokohji, Design and evaluation of a telepresence vision system for manipulation tasks, in: *Proc. IEEE Int. Conf. on Robotics and Automation*, Roma, pp. 4313–4318 (2007).
11. S. Tachi, N. Kawakami, M. Inami and Y. Zaitsu, Mutual telexistence system using retro-reflective projection technology, *Int. J. Humanoid Robotics* **1**, 45–64 (2004).

12. M. Inami, N. Kawakami, D. Sekiguchi, T. Maeda and S. Tachi, Head-mounted projector, in: *Conf. Abstract and Applications of SIGGRAPH99*, Los Angeles, CA (1999).
13. S. Tachi and K. Yasuda, Evaluation experiments of a telexistence manipulation system, *Presence* **3**, 35–44 (1994).
14. T. Maeda and S. Tachi, Development of light-weight binocular head-mounted displays, in: *Proc. 2nd Int. Symp. on Measurement and Control in Robotics*, Tsukuba, pp. 281–288 (1992).
15. K. Watanabe, I. Kawabuchi, N. Kawakami and S. Tachi, TORSO — completion of egocentric telegnosis system, in: *Conf. Abstract and Applications of SIGGRAPH07*, San Diego, CA (2007).

About the Authors

Kouichi Watanabe received the BE degree in Information and Image Sciences from Chiba University, in 2004, and the ME degree in Information Physics and Computing from the University of Tokyo, in 2006. Currently, he is a candidate for the PhD in Information Physics and Computing at the University of Tokyo. His research interests include virtual reality and robotics.

Ichiro Kawabuchi received the PhD degree in Mechanical Engineering from Tokyo Institute of Technology in 1999. From 2001 to 2003, he was jointly appointed as a Research Staff of the Japan Science and Technology Agency (JST) at Tachi Lab., University of Tokyo, where he was engaged in the design of telexistence humanoid robots. He is currently working as a Chief Engineer and a CEO of Kawabuchi Mechanical Engineering Laboratory, Inc. and is occupied in developing specialized experimental mechanics for many well-known institutes. His current research interests include the design of compact robots with many d.o.f. and development of mechanisms. He is a member of the JSME, RSJ and VRSJ.



Naoki Kawakami received the PhD degree in advanced Interdisciplinary Studies from the University of Tokyo. He is a Lecturer in the Department of Information Physics and Computing at the University of Tokyo. His research interest is virtual reality.

Taro Maeda received the BE degree from the University of Tokyo, Tokyo, in 1987. Until 1992, he was a Researcher in the Mechanical Engineering Laboratory, Ministry of International Trade and Industry, Tsukuba. Until 1997, he was a Staff Researcher of the University of Tokyo. Until 2003, he was an Assistant Professor of the University of Tokyo. Until 2007, he was a Senior Research Scientist, Supervisor of NTT Communication Science Laboratories, Atsugi-shi. Since 2007, he has been a Professor in the Department of Bioinformatic Engineering Graduate School of Information Science & Technology at Osaka University. He received the PhD (Engineering) from the University of Tokyo, in 1994. His research interests include psychophysics, modeling of brain by neural networks, man–machine interface for teleoperation and virtual reality. He was awarded the 1990 Society of Instrument and Control Engineers prize for his paper on modeling of a visual horopter.

Susumu Tachi received the PhD degree in Mathematical Engineering and Information Physics from the University of Tokyo. He is a Professor in the Department of Information Physics and Computing at the University of Tokyo. He is a Founding Director of the Robotics Society of Japan, a Fellow of the Society of Instrument and Control Engineers, and is the Founding President of the Virtual Reality Society of Japan. His research interests include virtual reality, telexistence, augmented reality and robotics.