

HUMAN MOVEMENT INSTRUCTION SYSTEM THAT UTILIZES AVATAR OVERLAYS USING STEREOSCOPIC IMAGES

Masayuki Ihara[†] Yoshihiro Shimada[†] Kenichi Kida[†]
Shinichi Shiwa[†] Satoshi Ishibashi[†] Takeshi Mizumori^{††}

[†]NTT Cyber Space Laboratories

^{††}NTT Communications Corp.

1-1 Hikarinooka Yokosuka-Shi Kanagawa 239-0847

Japan

ihara@nttvdt.hil.ntt.co.jp

ABSTRACT

In the Mixed Reality environment that combines the real world and the virtual world it is important to control the 'place' that comprises both real and virtual objects. In this paper, we study the overlay of humans and avatars in virtual space in the creation of an immersive human movement instruction system that works through a network. In this system, users wear stereoscopic glasses and motion capture devices and perform within an immersive virtual space experience system called CAVETM. A user can look at a stereoscopic image of an avatar that is displayed over his or her own body. In this project, the authors connected two CAVETM systems to a network to develop a system that enables model movement instruction and judgment of movement skills to be implemented by people in remote locations. This paper will introduce the setup and configuration of this system.

Keywords: mixed reality, control of 'place,' human movement instruction, overlay display, stereoscopic vision, avatar

1 INTRODUCTION

In the Mixed Reality (MR) environment that combines the real and virtual worlds it is important to control the 'place' that comprises both real and virtual objects. In this paper, the authors study the overlaying humans on avatars in virtual space, and introduce a human movement instruction system that uses the overlaying.

Generally, in the MR environment ob-

jects are displayed as two-dimensional images and the person viewing those images does not gain a sense of immersion into the environment. That is, in the past, MR almost always referred to the synthesis of photo-realistic images and computer graphics. Recently, see-through head mounted displays have been developed and, through the use of stereoscopic images, users can obtain a sense of immersion when viewing images. However,

no systems reflect a user's own movements in the images. The CAVETM system [Cruzn93a] was developed at the University of Illinois as a system for enabling a sense of immersion using stereoscopic vision and huge screens. Hitherto, CAVETM has mostly been used for the visualization of chemical phenomena, for designing cars from remote locations, and for simulating the operation of airplanes and cars. Recently, we have seen research projects, such as the CAVETM version InterspaceTM [Kouno99a] and CALVIN Project [Leigh96a], where humans are expressed as avatars in the virtual space. These projects attempt to achieve smooth communication between remote locations.

However, until now there has been no system in which a user's own body movements are recognized and reflected in the display of images in the virtual space, where users look at these images with such a sense of immersion that it feels as if they are looking at the real thing, and where the user, a person in the real world, can collaborate with an avatar of the virtual world. Certainly there has been much research in the past on the recognition of human movement. 'Shall We Dance?' [Ebiha98a], in which movement recognition technology was used to enable collaboration between avatars on a virtual stage, is one such example. There have also been a number of motion capture equipment marketed. One example of their utilization is 'Cyber Dance' [Miral98a] where the movements of one's own body captured through motion capture were reflected in the movement of an avatar. However, in all of these systems, the user merely looks at an avatar displayed on a flat monitor. This is grossly inadequate when considered in terms of the sight and senses with which we people look at things in the real world. On the other hand, in virtual space walk-through systems, such as the CAVETM version

InterspaceTM [Kouno99a], users can view images from their own perspective while walking around the virtual space and can feel the atmosphere of a 'place' that cannot be experienced on a flat screen. If human movements can be treated in the same way, overlaying the user onto an avatar so that the user can view both movements of himself or herself and of the avatar, then the user will obtain a greater sense of sharing the virtual space.

In this project, the authors built a movement instruction system that allows collaboration between humans and avatars. In this system, the user wears stereoscopic glasses and motion capture devices and performs within CAVETM. The user can see a stereoscopic image of an avatar that is displayed over his or her own arms and legs. The avatar is either driven by motion data files of model movements prepared in advance or by the model movements of the user (instructor user) who moves in another CAVETM on the network. The trial system used in the project consisted of two CAVETM connected onto a network. We developed a system of teaching movements by sending and receiving motion data via the network. The system is not only used to teach movement but also has a function that judges the skill of user (pupil user). This paper will introduce the setup and configuration of this system.

2 MIXED REALITY

While generally speaking MR refers to the fusion of the real and virtual worlds, the word is often used to refer to the synthesis of photo-realistic images and computer graphics. However, 'reality' can be sensed not only in response to photo-realistic images but also from a sense of immersion felt in response to life-size stereoscopic images. The space that we created in which there is not just a synthesis of images but some sort of collaboration between real

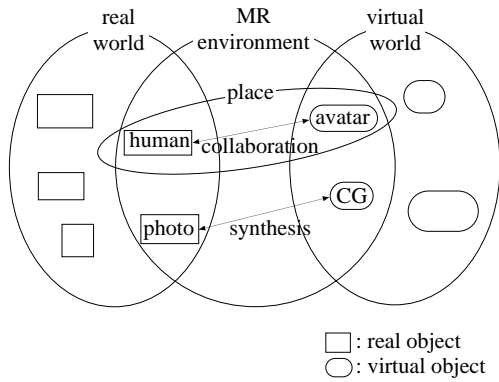


Figure 1: The MR environment and 'place'

and virtual objects is called the 'place,' in which we human beings can feel 'reality.'

Figure 1 shows a model of the MR environment, its components, and the 'place' relationship. A 'place' exists when there is some sort of collaboration (for example shared movement) between real objects in the real world, that is human, and virtual objects in a virtual world, that is avatars. An MR environment (in its broad sense) with reality can be created by quality control of this 'place' (for example by enabling a sense of overlay using stereoscopic vision). Accordingly, an important factor in achieving is the smooth collaboration between real and virtual objects, that is control of the 'place.'

3 THE CAVETM SYSTEM

The authors decided to use CAVETM as the platform for research of 'place' control technology. The reasons for the selection of CAVETM are given below.

- A user can gain a sense of reality through stereoscopic images.
- A sense of immersion can be obtained through use of large screen images.
- A user can see the image even when changing direction by moving.

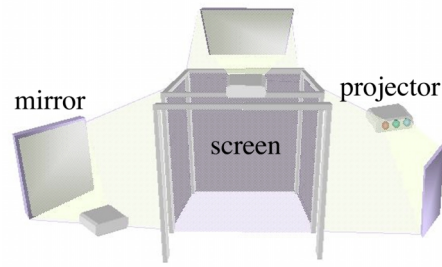


Figure 2: CAVETM used by authors

The senses of reality and immersion enabled by stereoscopic vision can also be achieved through the use of head mounted displays but, consideration of the quality of the response when a user changes direction led us to select CAVETM.

A picture of the CAVETM used by the authors is shown in Figure 2. There is a total of four screens, one at the front, one each on the right and left and one on the floor. These screens are 3 meters square. Images from each of the projectors are reflected by mirrors onto each screen. The projected images are stereoscopic images with the parallax of both eyes. By wearing special glasses for stereoscopic vision, the user can see these stereoscopic images.

4 OVERLAYING A PERSON AND AN AVATAR

As explained above, stereoscopic vision is enabled with CAVETM. Therefore, if a life-size avatar is displayed as a stereoscopic image, a person can be 'overlaid' onto that avatar. Strictly speaking, it would be more correct to speak not of 'overlaying' but of 'appearing overlaid through stereoscopic glasses.' Figure 3 shows such an overlay. The overlaying of a user's body onto an avatar allows the user to view both the gestures of himself or herself and of the avatar. Furthermore, those gestures will be given solid

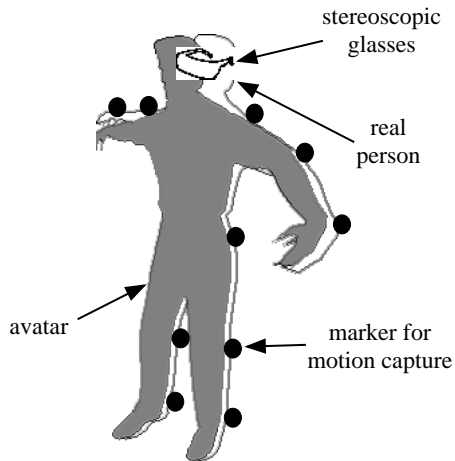


Figure 3: Overlaid display image

form in the same way as we humans sense our own gestures in the real world. In other words, through 'overlay' we can obtain a sense of reality that is close to that experienced in the real world. This can be applied when learning dance or sport moves. The application enables for users to sense if the overlay on the avatar is complete or if they are slightly out of synchronization. If the avatar moves using model movements, the user should always try to make gestures that directly overlay those of the avatar. To control the movement of an avatar, motion capture can be used. The user moves while wearing motion capture markers. The avatar is moved based on the three-dimensional coordinates of the markers as calculated through motion capture.

5 MOVEMENT INSTRUCTION

In the past, people often watched videos of dance and sports moves carefully to learn them. However, from the point of view of learning movement, it is difficult to say that watching a two-dimensional video is very effective in learning three-dimensional movements. There is also the problem that it is impossible to judge in-

stantaneously whether or not one's movements are correctly proceeding according to the model seen in the video. The authors believe that the proposed system solves these problems. In other words, a user can use stereoscopic vision to sense differences between the model movements and his or her own movements in three dimensions. The system detects these differences and can issue judgment about the skill level. These will increase the efficiency with which movements are learnt.

6 PROPOSED SYSTEM

6.1 Overview of Processing

A diagram of the configuration of the hardware used in the movement instruction system proposed and developed by the authors is given in Figure 4. In Figure 4, the user moves in CAVETM while wearing motion capture devices (Figure 5). The movements are filmed using two cameras that are placed at the entrance to the CAVETM. The cameras are VGA visible light cameras. High luminance color LED markers are used for motion capture. These markers are placed in 16 positions as shown in Figure 6. The three-dimensional measuring device used is the Real-Time Optical Meter "Quick MAG IV", developed by Japan's OKK Inc. The images filmed with the cameras are processed in the measuring device and output as a time-series of three-dimensional marker coordinates. The output coordinate data is sent to the Onyx2TM graphics workstation through a control PC and the movement of the avatar is controlled based on this coordinate data. If this coordinate data is sent to a remote system through a network, it can be reflected in the movement of the avatar in the remote system. It is also possible, of course, to use this data to control an avatar in a local system. The image of the avatar movement

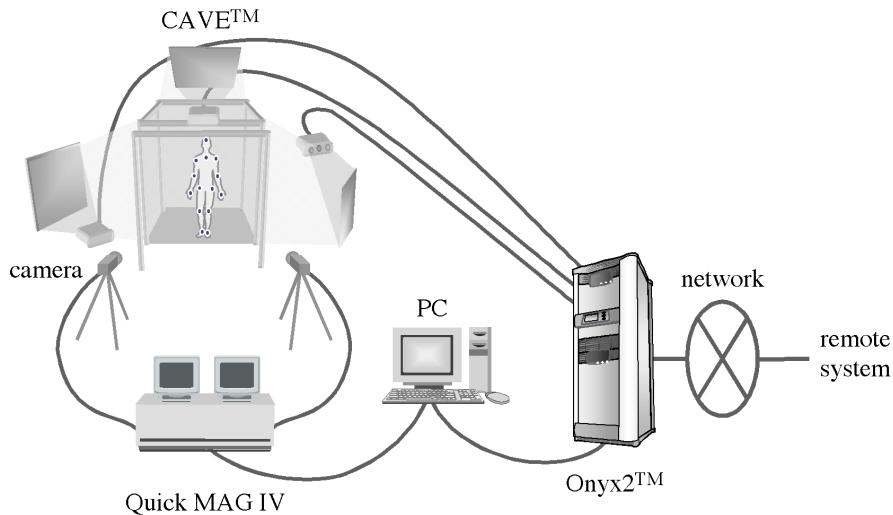


Figure 4: Proposed system



Figure 5: Movements in CAVE™

is projected onto all CAVE™ screens by projectors.

6.2 Module Configuration

The configuration of modules in the system is shown in Figure 7. The system comprises eight threads. The functions of each thread are as shown in Table 1. In Figure 7, 'ThreadCom' is the information data for communication between threads that is sent by and received from threads. 'PartnerCom' is the partner information data that is sent by and received from remote systems on a network. (A 'partner' refers to a remote user that is sharing actions.) Specifically, the identification flag that shows if the user is a pupil

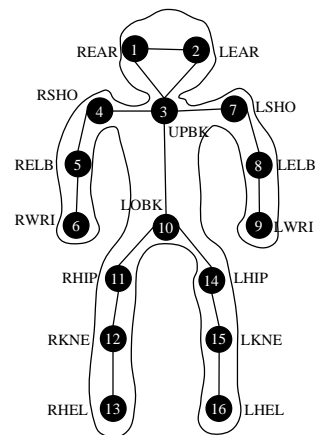


Figure 6: Marker positions

learning movements or a teacher showing model movements, and the name of the avatar file being used are recorded here. 'LineCom' is data relating to the lines that pass through the network. The libraries included are the 'CAVE™ library' provided by the CAVE™ system, the MIDI 'SOUND library,' and the 'MCAPI library' which is the data input API for motion capture. The data bases included are the 'Avatar file,' which is the CG data for the avatar, the 'MIDI file,' which is the music data for BGM, and the 'Rec file' where movement data is recorded as a file.

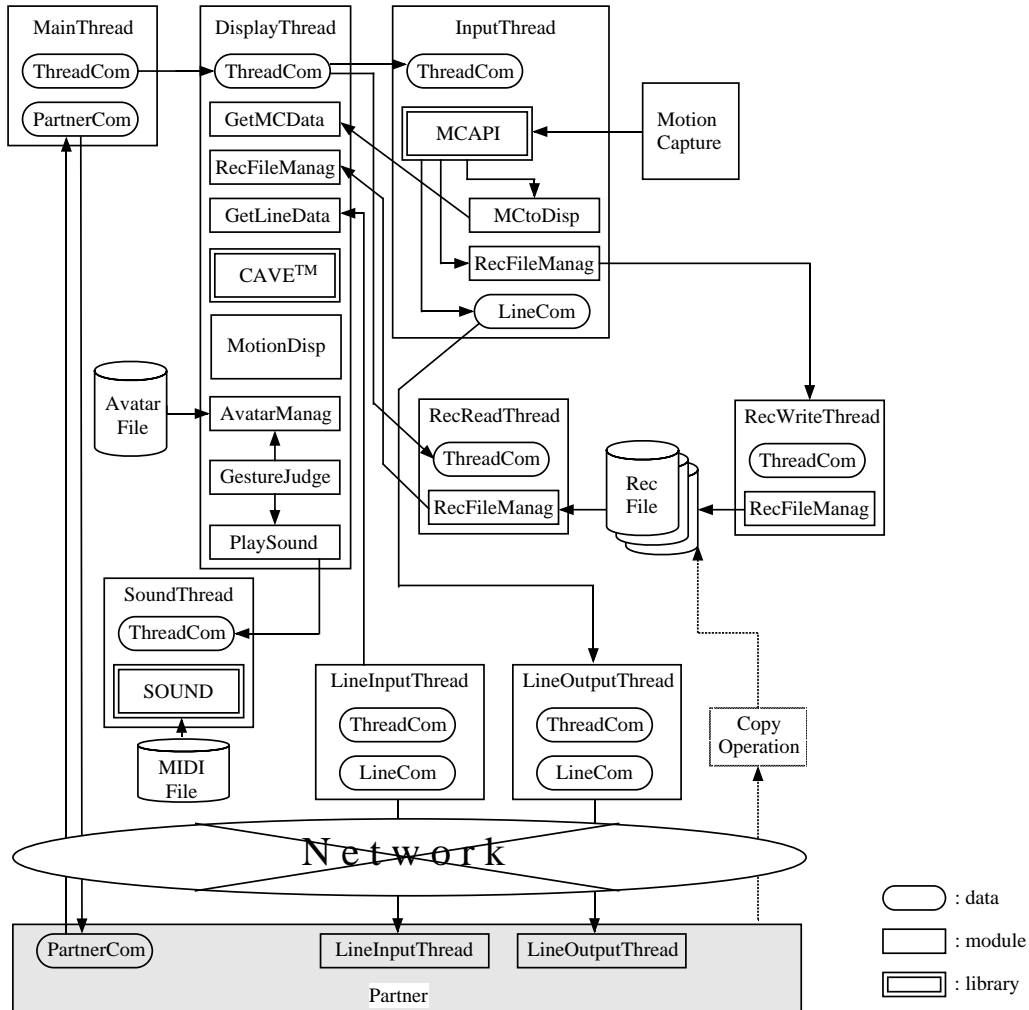


Figure 7: Module configuration

6.3 Calibration

When overlaying an avatar and a person, it is necessary to control the size and position of the displayed avatar so that it matches the size and three-dimensional position of the user. To this end, the Main thread starts calibration when use of the system starts and acquires the three-dimensional coordinates of the static state. Specifically, the user is asked to stand still and a number of coordinate values are continuously measured. The distribution of this coordinate data is calculated and threshold value processing implemented. When a distribution value less than or equal to the threshold value is

obtained, a static state check is implemented. The standards used for the check are 'the back is almost vertical' and 'both arms and both legs are basically hanging down straight.' The three-dimensional coordinates of the static state thus obtained are compared to the coordinate values of the joint nodes of the avatar. The rate of increase or decrease at each body part is determined and the size and position of the avatar controlled using those rates.

6.4 Skill Judgment

The developed system can also determine the skill with which movements are made.

Table 1: Thread functions

Thread name	Function
Main	Sets conditions at start up execution and control them during execution
Display	Receives movement data Converts coordinates Judges movement skill level Displays avatar data dynamically
Input	Enters data from motion capture Transfers data to Display thread Transfers data to LineOutput thread Outputs movement recording file Sends data to partner
LineInput	Receives data from partner Transfers data to Display thread
LineOutput	Sends data to partner
RecRead	Reads movement recording file
RecWrite	Writes to movement recording file
Sound	Outputs music

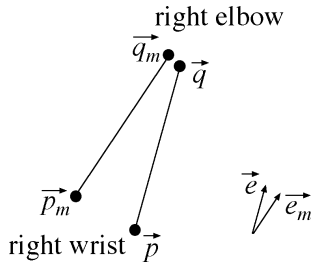


Figure 8: Joint vectors relating to lower right arm

The three-dimensional positions of the avatar and user parts are compared for each body part, for example lower right arm, upper left arm. When there is a large difference between these positions, image and noise effects are used to indicate that the user movements are wrong.

The following section will explain the specific algorithm used to determine the skill level using an example of the lower right arm. In Figure 8, \vec{p}_m and \vec{q}_m are vectors that show the joint nodes at the right wrist and right elbow of the avatar, where the model movements are being made. These are both the joint extremities relating to the lower right arm. \vec{p} and \vec{q} are show the joints of the users right wrist and right elbow, both being the joint extremities re-

lating to the lower right arm. The system uses differences in the slope of the vector that joints both joints to judge the movement skill level. If the overlay coefficients for judging differences in the right wrist and right elbow are h and k , the overlay coefficient for judging differences in the slope is l , and the threshold for judgment is T , then the skill level can be judged as shown in Equations (1) and (2). Note here that \vec{e}_m and \vec{e} are unit vectors, that is, $\vec{e}_m = (\vec{q}_m - \vec{p}_m)/|\vec{q}_m - \vec{p}_m|$, $\vec{e} = (\vec{q} - \vec{p})/|\vec{q} - \vec{p}|$.

If

$$E = h|\vec{p}_m - \vec{p}| + k|\vec{q}_m - \vec{q}| + l|\vec{e}_m - \vec{e}| < T \quad (1)$$

then right gesture else if

$$E \geq T \quad (2)$$

then wrong gesture.

7 CONSIDERATIONS

Use of the proposed system will enable sharing of movements between parties in separate locations. Furthermore, since a life-size stereoscopic avatar image can be viewed, placement of this on the remote avatar will enable collaborative operation of some three-dimensional virtual object.

Use of the avatar overlay function will enable learning of dance and sport moves, such as those used in baseball or golf. Because the proposed system adjusts enlarges and reduces a skeleton that is equivalent to body parts such as arms and legs, and because it matches the coordinate origins, even if the user is a short child, their image can be overlaid on the avatar controlled by the model movements being made by an adult teacher. Even if the user moves slightly to the left or right or backwards or forwards, the basic state is overlaid on the avatar. If the collaborative work function is used, for example

if a three-dimensional stereoscopic bat is prepared for a lesson about baseball batting moves, both the teacher and student can simultaneously swing the virtual bat to assist in learning moves. However, to achieve something close to reality where the users feels as if they are really swinging a bat, the system needs to be combined with an haptic interface.

Our system is the basic system that measures human movement, controls avatars, overlays avatars and humans and judges the movement skills. Although movements of 16 joints are filmed using two cameras in this system, the method and objects of measurement should be properly tuned in practical systems.

8 CONCLUSION

This paper has discussed the importance of the control of the 'place,' which is made up of real and virtual objects, in the Mixed Reality environment. We have focused on the 'place' through the overlay of people and avatars. As an example of the control system, we have proposed an immersive human movement instruction system. The setup and configuration of this system have been introduced in this paper. In this system, a user moves inside an immersive virtual space experience system called CAVETM while wearing stereoscopic glasses and motion capture devices. The user can see a stereoscopic image of an avatar over his or her own body. In this project, two CAVETM were connected by a network enabling instruction of model movements and judgment of movement skill levels. The hardware and module configurations have been introduced in the paper. The system has the effectiveness and future potential.

In the future, this system could be used as a tool in assisting the actual learning of movements and as a tool in evaluating the

results of such learning. We would like to look further into the problems associated with whole body movements made while using stereoscopic vision.

ACKNOWLEDGMENTS

The authors would like to thank Mr. Masahiko Hase, the project manager of the Media Communication Project at the NTT Cyber Space Laboratories who gave us the opportunity to conduct this research project.

REFERENCES

- [Cruzn93a] Cruz-Neira, C., Sandin, D. J. and Defanti, T. A.: *Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE, Proceedings of SIGGRAPH '93*, pp. 135-142, 1993
- [Ebiha98a] Ebihara, K., Davis, L. S., Kurumisawa, J., Horprasert, T., Haritaoglu, R. L., Sakaguchi, T. and Ohya, J.: *Shall we dance? -Real time 3D control of a CG puppet-, SIGGRAPH '98 Enhanced Realities, conference abstracts and applications*, pp. 124, 1998
- [Kouno99a] Kouno, T., Suzuki, Y., Yamamoto, N., Shiwa, S. and Ishibashi, S.: *Immersive virtual communication environment, Technical Report of IEICE, MVE 99-184*, pp. 1-8, July 1999 (in Japanese)
- [Leigh96a] Leigh, J. and Johnson, A.: *Supporting transcontinental collaborative work in persistent virtual environments, IEEE Computer Graphics and Application*, Vol. 16, No.4, pp. 47-51, July 1996
- [Miral98a] <http://miralabwww.unige.ch/>