Physical Instructional Support System Using Virtual Avatars

Tomoaki Ogawa
Department of Computer and Information Engineering,
Nippon Institute of Technology 4-1 gakuendai,
Miyashiro-cho, Minamisaitama-gun, Saitama 345-8501
Japan
c1065201@cstu.nit.ac.jp

Yasushi Kambayashi
Department of Computer and Information Engineering,
Nippon Institute of Technology 4-1 gakuendai,
Miyashiro-cho, Minamisaitama-gun, Saitama 345-8501
Japan
yasushi@nit.ac.jp

***Abstract*—Certain sports such as martial arts and dance have sets of good typical motion types. These motion types were abstracted from the physical movements of excellent practitioners. They are devised for instruction purpose, and are optimal movements of the target sports. It is extremely important for learners to learn these typical motion types. Traditionally the transfer of such typical motion types is done by in-person instructions. Therefore it is considered that it is not suitable for distance learning environment. We have developed a support system to convey this motion types by way of communication networks. In this paper, we propose a real-time physical instructional support system. The instructor and the learners communicate with each other by the virtual humanoid 3D-CG avatars through the Internet. By using this system, it is possible for the instructor to demonstrate his motions, and for learners to obtain the instructor's movements at distant places. The obtained information is beyond simple camera images. Because the system provides the three-dimensional perspectives and it superimposes the instructor's movements on the learner's avatars as well as provides a means of real-time direct communication.**

Keywords-education; sport; Kinect; MMDAgentt; remote instruction

I. INTRODUCTION

Certain sports such as martial arts and dance have good sets of typical motion types. These motion types were abstracted from the physical movement of excellent practitioners. They are devised for instructional purpose, and are optimal movements of the target sports. It is extremely important for learners to learn these typical motion types.

Traditionally learning these motions is done by imitation learning. In imitation learning, the learner mimics the motions of the instructor. Even imitation learning, not-in-person learning is possible. Various methods for transferring physical movements without direct instructions are developed. Self-learning through books and recorded video viewing are two popular ways to learn. None of them are as effective as direct-learning through actual experiences with instructors, though.

In order to mitigate the problems of self-learning, Chua et al developed a distance learning system that feedbacks the learner's movements to him in order to highlight differences from the recorded instructor's movements [1]. With the distance learning system, the user can observe the instructor's demonstrations as well as his own recorded movements as many times as he wants. When the learner encounters some confusing situation, however, he has no way to communicate with the instructor to get immediate advices. It is also hard for the leaner to understand the instructor's intention behind a certain physical motion.

We have devised a way of solving these problems. In this paper, we propose a real-time physical instructional support system. The instructor and learners communicate with each other by the virtual humanoid 3D-CG avatars through the Internet. By using this system, it is possible for the instructor to demonstrate his motions, and for learners to obtain the instructor's movements at distant places. The obtained information is beyond simple camera images. Because the system provides three-dimensional perspectives and it superimposes the instructor's movements on the learner's avatars as well as provides means of real-time direct communication.

In this system, the virtual avatars of the instructor and the learners share the same virtual space. Both the instructor and the learners exercise their performances in front of the sensor called Kinect. Kinect is a Microsoft product and widely available [2]. Since Kinect tracks the performers' movements in real-time, the virtual avatars in the shared virtual space move precisely as the owner of the avatar. Since Kinect features an RGB camera and a depth sensor, it can provide the depth values of the scene in real-time. Using these depth values, the system constructs the humanoid 3D avatars in the virtual space. Since each virtual avatar maintains 3D information from Kinect, the system can display the movements of the avatar from any angles by rotating the virtual camera.

Also, since the avatars of the instructor and the learners share the same virtual space, this system can express the differences between the model movement and the learners' not-so-good real movements directly. Those differences are hard to express in verbal description and camera images. In addition, the system allows resizing the avatars of learners and instructor so that they have the same size. This can compensate for the difference in physique between the two. In addition to these features, we have integrated a gesture and voice recognition and speech synthesis subsystems into our physical instructional support system so that the instructor and the learners can verbally communicate with each other as well as they can control the virtual camera and their avatars.

Figure 1 shows the overview of the physical instructional support system we are currently constructing. In this figure, the Kinect sensor obtains 3D motions of the instructor and the learners. The local PC connected to the Kinect extracts the skeleton information of the data sent from the Kinect, and transmits it to the server. There is one central server that provides communications between local PC's so that all the participating local PC's share one virtual space where all the avatars that represent the instructor and the learners reside. The local PC constructs avatars from the skeleton information sent from the server. Each avatar corresponds to the skeleton information sent from each remote PC, which manipulates the avatar so that each avatar reflects its owner's movements in the virtual space.

The rest of this paper is organized as follows. In the next section, we describe the related works. In section III, we describe the leaning process of physical body movement. In section IV we describe the physical instructional support system we are constructing. We conclude our discussion in section V.

II. RELATED RESEARCH

There are quite a few research works for distant learning systems that support obtaining certain physical motion types. The most notable self-learning system that is closely related to our system was developed by Chua et al.

Chua et al developed a self-learning system that uses a virtual avatar in order to learn Tai Chi. [1]. The user learns typical motions by observing instructor's pre-recorded demonstration. Then the user tries to imitate the movements. The system gives feedback to the learner's movement by using avatar and superimposes it on the instructor's model movement so that it highlights the differences between them. The learner is supposed to correct his movement by observing the difference. Unfortunately, it lacks direct communication between the instructor and the learners. It confines its effectiveness in self-study environment. Unlike their work, we have integrated a gesture and voice recognition and speech synthesis subsystems into our physical instructional support system so that the instructor and the learners can verbally communicate each other as well as they can control the virtual camera and their avatars.

Usui et al developed a dance entertainment system that consists of a recording studio and a mobile movie viewer [3]. The user can mimic a dance movie and record his movement, and then review his dance technique with other dancers' recorded movements on the mobile viewer. This system provides a convenient way of comparison learning by using a handy mobile viewer.

Honjou et al developed a self-learning system for golf swings. In the system, the user mounts a semi-transparent head-mounted display (HMD) that shows the instructor's swing concurrently when the user swing his club [4]. These studies pay attention to how to show the instructor's motion, at the same time the user performs his motion so that the user can recognize the difference between the instructor's model movement and his own movement. They all lack, however, interactive features so that the user has no way to contact the instructors as he practices.

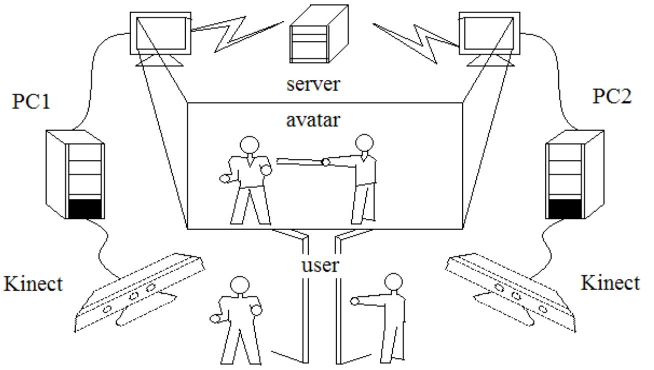

Figure 1. System overview

For real-time interaction, Nakazawa et al developed a high-definition HyperMirror system utilizing satellite communication for interactive distance learning [5]. The system provides extremely good videoconference experiences by using the real image of the attenders so that all the participants can feel as if they were in the same room. The video image, however, is hard to manipulate so that the system cannot rotate the image of people. The user cannot observe counter-part's image at different angles. Even though providing cordial feeling to the participants as if they are sharing the same space, video image is not ideal for physical instructions. In addition, Morikawa et al have reported that there exists some aversion of self-image [6].

Kimura et al studied how to display visual information for the remote learning of various body movements over network. In the system, the instructor and the learner equipped various sensors on the joint parts of the arm, and they can view the movement of the arm in the HMD screen [7]. The instructor and the learner see the virtual avatar in the shape of the arm, so that the instructor's arm position can be superimposed over the learner's. The instructor can correct the learner's movement by verbal instruction over IP telephone. They have succeeded in constructing an instructional system, but it is limited to only arm position due to the requirement of too many sensors. Our system requires no attached sensors so that the system should be able to adapt for whole movements.

So far we are not aware of any support system of interactive physical movement learning which uses the speech and gesture recognition so that the instructors and the learners can communicate as naturally as real face-to-face instruction. Such a system is what we are presenting in this paper.

III. LEARNING PROCESS OF MODEL MOVEMENTS

In this section we discuss the learning process of certain sports that have typical physical movements. We assume that the instructor and the learners pursue the following process to teach and learn the typical movements.

1. Presentation: the instructor demonstrates his model motions to the learner.

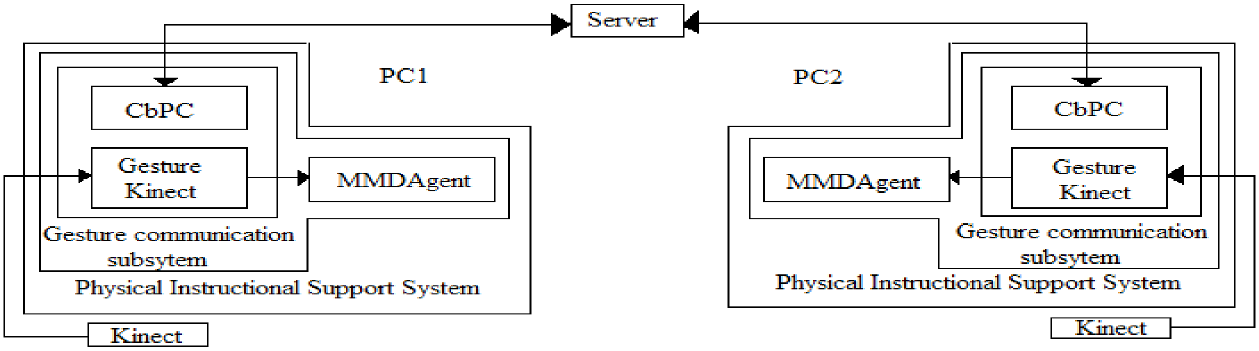
Server
PC1
PC2
CbPC
Gesture Kinect
MMDAgent
Gesture communication subsytem
Physical Instructional Support System
Kinect
CbPC
Gesture Kinect
MMDAgent
Gesture communication subsytem
Physical Instructional Support System
Kinect

Figure 2. System configuration

2. Comprehension: the learners observe the instructor's demonstration and try to comprehend how to imitate the model motions.
3. Practice: the learner imitates the motions based on his comprehension.
4. Assessment: the instructor assesses the movements of the learner who mimics his demonstration. The instructor makes corrections on the learner's exercises with gestures and verbal advices.
5. Confirmation: the learner repeats his exercises based on his comprehension to confirm the movements and understanding of the intention of the instructor.

The most important part of the learning process is the assessment of leaner's physical motions. In this assessment, the instructor measures the progress of the learner's understanding, and gives effective instructions to correct the leaner's movements.

For effective instruction, the instructional support system put the first priority to the smooth sharing of information and understanding between the instructor and the learners. Our instructional support system satisfies this requirement through smooth communication between them based on the virtual avatars.

IV. INSTRUCTION-SUPPORT-SYSTEM

A. *Functional requirements and solutions*

In order to construct the physical instructional support system that endorses the learning process we have just described in the previous section, we set the following functional requirements and solutions.

1) The system must keep tracking of the performer's movement and reflects it to the corresponding avatar.

2) The system must recognize the speech and gesture of the performer.

3) The system must provide real-time interactions between performers in distant places.

We discuss them some details below:

1) Tracking control of a virtual avatar

The system keeps tracking of the performer's movement and reproduces the instructor's and learners' movements on the virtual avatars. Each performer, either the instructor or the learner, has its own avatar. All the avatars share the same virtual space, and all the participants at various distant places can observe all the avatars in that virtual space. In other words, they see the same virtual space from their own remote display terminals.

In order to achieve this requirement, we utilize a sensor called Kinect to keep tracking of motion of the instructors and the learners. The instructors and the learners are displaying their physical movements in front of the Kinect and the system reflects the motions on the virtual avatars in the same virtual space. We use MMDAgent for modeling of the virtual avatars [8]. MMDAgent is a Tool Kit that constructs humanoid 3D models and allows the models to communicate with the human users. The Speech Processing Laboratory of Nagoya Institute of Technology has developed MMDAgent, and the institute provides it freely. We construct the avatars by using MMDAgent.

2) Speech recognition and gesture recognition

The system must rotate and move the virtual camera as well as the avatars. The system must provide a means to get the attention of instructor. The instructor and the learners should be able to verbally communicate with the system and instruct the system to rotate the virtual camera as well as the avatars. Also the learners should be able to verbally request the system to get attention of the instructor. The communication between the human users, i.e. the instructor and the learners, and the system are done through speech recognition subsystem so that they can concentrate their physical movement practices.

In order to achieve this requirement, we utilize MMDAgent for speech recognition. Also, we have integrated a gesture recognition subsystem into MMDAgent so that the users can communicate with the system via not only voice but also gestures. Figure 2 shows the system configuration. A detailed explanation about the gesture communication subsystem is given in the next subsection.

3) Distant communication

The system can display avatars in distant places so that all the participants in distant places can see the same virtual avatars at the same time. The participants should be able to perform voice-chat. The voice chat between users and the verbal commands to the system are differentiated by means of the “mode” so that the users can give commands to the system as well as can talk to each other.

In order to achieve this specification, the system has been connected with UDP in socket communication to realize high-speed connection. The gesture communication subsystem system has a means to switch between voice-chat and speech recognition for commanding the system.

B. *Gesture communication subsystem*

We have developed a gesture communication system to amend MMDAgent. This subsystem recognizes the gestures of human performers that are sent from the Kinect. The different gestures indicate whether the user wants to command the system, such as to rotate the virtual cameras and the avatars or the user wants to communicate to one of the users in a distant place. This subsystem consists of the following two parts (see Figure 2).

Part 1, Gesture Kinect, is in charge of the gesture recognition that obtains the skeleton information sent from the Kinect. The Kinect senses 3D motion of the user and constructs skeleton information as shown in Figure 3. Gesture Kinect recognizes certain gestures of the user and sends commands to MMDAgent to control the virtual camera and avatars. In order to point precise position of learner’s body, the instructor can create and use a virtual tact through gesture command. (Figure 1 shows the avatar of the instructor uses the tact to pint the learner’s body.) Table I shows typical gesture commands.

Part 2, Communications between the PC (CbPC), controls the voice-chat and transmitting the skeleton information for avatars.

V. CONCLUSION AND DISCUSSION

In this paper, we proposed a real-time physical instructional support system through virtual avatars. The instructor and learners who are in remote places can share the same virtual space by their avatars. This system can transmit 3D movements that are unable to be realized by the video image. Thus the instructor can give the learners precise guidance. The instructor can point the precise position of a certain learner, and show the model movement by superimposing his avatar. Since the instructor and the learners share the same virtual space, it is possible for them to keep mutual understanding. As performing in front of Kinet sensor, the need for HMD and other sensors attached to the bodies is completely eliminated. Also the speech and gesture recognition subsystem eliminates the use of keyboard and mouse, so that they can perform, demonstrate, and learn in more natural and effective environment than all the previous similar systems. In addition, communication with avatars drastically decreases the amount of transmission data because only skeleton information just enough to construct avatars’ movement is sent and received between distant places.

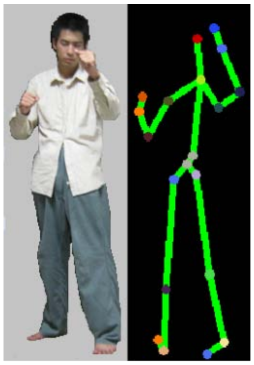

Figure 3. Skelton information

REFERENCES

[1] P. T. Chua, R. Crivella, B. Daly, N. Hu, R. Schaaf, D. Ventura, T. Camill, J. Hodgins, and R Pausch, “Training for physical tasks in virtual reality environments: Tai chi,” Proc. IEEE Virtual Reality. IEEE Press, 2003, pp. 87-94.

[2] http:// research.microsoft.com/ en-us/ um/ redmond/ projects/ kinectsdk/ default.aspx, last access: Dec. 6, 2011.

[3] J. Usui, H. Hatayama, T. Sato, Y. Furuoka, and N. Okuda, “Paravie: dance entertainment system for everyone to express oneself with movement,” Proc. 2006 ACM SIGCHI International Conf. Advances in Computer Entertainment Technology, 2006, pp. 14-16.

[4] N. Honjou, T. Isaka, T. Mitsuda, and S. Kawamura, “Proposal of Method of Sports Skill Learning using HMD,” Trans. Virtual Reality Society of Japan, Vol. 10, No.1, 2005, pp. 63-70.

[5] A.Nakazawa T, Okubayashi, H Mori, T Maesako, O Morikawa, M Nakao, N Tomii, T Sato, T Kawasugi, and G Hashimoto, “Applying High-Definition ”HyperMirror” to Distance Learning Utilizing “KIZUNA” ”, 27th International Symp. Space Technology and Science, 2009-j-19p, pp. 1-6.

[6] O. Morikawa and T. Maesako,“HyperMirror: Toward Pleasant-to-Use Video Mediated Communication System,” Proc. ACM Conf. Computer-Supported Cooperative Work (CSCW98), ACM Press, 1998, pp. 149-158.

[7] A. Kimura, T. Kuroda, Y. Manabe, and K. Chihara. “A study of display of visualization of motion instruction supporting,” Japan Journal of Educational Technology research, Vol.30, 2007, pp. 45-51.

[8] MMDAgent: Toolkit for building voice interaction systems http://www.mmdagent.jp/, last access: Dec. 6, 2011.

TABLE I. GESTURE INDEX

Gesture Index		
Command	**Effect**	**Pose**
JMenu	display menu	Raise the right hand, specify by the left hand
JRotate	Rotate the camera view	Raise the left hand, specify by the right hand
JMove	Move the camera view	Raise the left hand, specify by the right hand
JPopItem	Create the tact	Put hands together