Non-verbal Human Communication Using Avatars in a Virtual Space

Arita, Daisaku
Department of Intelligent Systems, Kyushu University

Taniguchi, Rin-ichiro
Department of Intelligent Systems, Kyushu University

出版情報：Lecture notes in computer science. 2774, pp.1077-1084, 2003-09. Springer Berlin, Heidelberg
バージョン：

Non-Verbal Human Communication Using Avatars in a Virtual Space

Daisaku Arita and Rin-ichiro Taniguchi

Department of Intelligent Systems, Kyushu University
6-1, Kasuga-koen, Kasuga, Fukuoka, 816-8580, JAPAN
{arita,rin}@limu.is.kyushu-u.ac.jp

Abstract. Distance learning such as TV-based learning, Internet learning, etc. has become popular in these days. In particular, distance learning on Internet is a promising approach to decrease the geographical barrier of education and to provide a personalized learning framework. However, such distance learning has a problem that participants of the distance learning can not acquire the feeling of "reality", or realistic atmosphere of learning, though multi-modal information, or visual and audio information is transferred between learning sites. To solve the problem, or to provide more realistic learning environment for distance learning, we are researching for a virtual environment for distance learning with high reality. Particularly, in this research, we are aiming at construction of a virtual class room, which a teacher and geographically scattered students can attend via the network.

1 Introduction

Distance learning such as TV-based learning, Internet learning, etc. has become popular in these days. In particular, distance learning on Internet is a promising approach to decrease the geographical barrier of education[1]. It enables individuals to have their favorite course wherever they are. It can provide a personalized learning framework, which is becoming an important feature of learning in the future. On the other hand, it is no doubt that a classroom, or classmates, have an important role in learning. Students learn a lot from other students' activity. Group learning is often an effective method of learning. Therefore, we have to establish a distance learning framework having both features of individual-based learning and group-based learning, and to construct a lecture space virtually on the network with geographically dispersed participants.

The biggest problem of distance learning is that participants can not acquire the feeling of "reality", or realistic atmosphere of learning, even though multi-modal information, or visual and audio information is transferred between learning sites. For example, they can not always get views which they want to see. To solve the problem, or to provide more realistic learning environment for distance learning, we are researching for a virtual environment for distance learning with high reality, which we call Virtual Environment for Immersive Distance

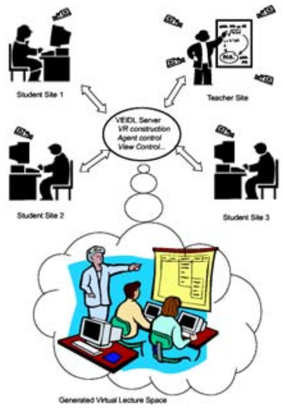


Fig. 1. Concept of VEIDL

Learning (VEIDL in short). Particularly, in this research, we are aiming at construction of a virtual classroom, where a teacher and geographically dispersed students can attend via the network.

2 Overview of VEIDL

In the virtual class room, which provides realistic situation of lecture space, participants (a teacher and students) can not only communicate with one another, but observe other participants activity. We assume that each participant including a teacher is on a separated site. At each participant site, there are multiple cameras and a microphone for acquiring participants' information. The acquired participants' information is transferred between sites. This information exchange is mediated by a server called VEIDL server. According to the transferred participants' information, a virtual 3-D lecture space is constructed in real-time at each site. Each participant equips a head-mounted display (or another personalized 3D display system) and a headphone. Through these equipments, each participant can see and hear the other participants in a virtual lecture space which is established on each site as if he/she were in a real, traditional lecture space. In other words, each participant at a separated location in the real world can share a single lecture space in a virtual world. This research realizes not only an immersive distance learning environment but also a learning environment unable to realize in a real environment, for example using a virtual screen, a virtualized animal walking on desks, etc.

However, it is almost impossible to acquire all the information about each participant in the real world and to present it to the others because of limitations of observation ability of the system and network bandwidth. Then, we propose a system which acquires and transfers only minimum information to present participants' intentions, and which compensates other information for participant avatars to behave naturally by means that each participant agent has a model of each participant. In other words, through constructing a virtual environment for distance learning, we will realize a system which understands important information of participants including participants' intentions referring to multi-modal or audio-visual input and which makes avatars behave naturally according to the information in order to communicate participants' intentions with each other.

Here, the most important issue is the balance between observed information from sensors and pre-defined knowledge, i.e., a behavior model of participants in lecture scenes. If we have an adequate behavior model, not only we can correct errors of the observed information but we can reduce the amount of information to be transferred by transferring only meaningful information.

3 Research Issues of VEIDL

Since verbal information is transferred without processing, it is important how non-verbal information representing participant's intentions is acquired and how a virtual lecture space is visualized.

3.1 Acquisition of Participant's Information

The 3D position and the pose of a teacher are acquired by a vision-based full body motion capture system using multiple cameras[2]. Since non-verbal information about the teacher is usually important, enough number of cameras and computers must be provided to observe the information precisely. On the other hand, the 3D position and the pose of each student are acquired by a desktop motion capture system using a stereo camera[3]. Since the variation of the students' non-verbal information is rather restricted, and since expensive hardware can not be used for the students, the system for each student must be as simple as possible.

3.2 Visualization of a Virtual Lecture Space

A virtual 3-D lecture space is constructed in real-time on each site according to the participants' information, which is generated from the acquired information and pre-defined knowledge. Prepared teaching materials such as slides are also adequately presented in the virtual lecture space. Each participant can see the virtual lecture space through a head-mounted display (or another personalized 3D display system). The visual information is changed in response to the face direction of the participant.

An agent of each participant must be constructed in advance and distributed to all sites. In order that each participant avatar behaves naturally based on the information supplied by the VEIDL server, each participant agent has such information as

For Deciding Action state transition diagrams and state transition probabilities on them

For Displaying Motion a 3D shape-color model and motion models

as the pre-defined knowledge. Since the server does not precisely control avatars' action, the actions of avatars which represent the same agent, i.e., the same participant, are slightly different in each site. However, this is not serious problem since the important information about participants is reflected in their avatars.

State Transition Diagram and State Transition Probability. Each human part (left arm, right arm, left leg, right leg) has a state transition diagram. Each state in the diagram means a single action such as hanging arm, finger pointing and folding for an arm and standing and walking for a leg, so it is guaranteed that each part can do only one action at a time.

The state transition probability is conditional probability whose conditions are as follows:

Action Information from an Acquiring Site When an action information about a participant is received, in order that an avatar do the action, a state transition probability for transiting to the state of the action becomes 1. This means the current state immediately transits to the state of the action. In case that the current state can not transit directly to the state of the action, the current state moves to the state of the action via other states. Using state transition diagram in this way guarantees that an avatar changes its action without any unnatural or impossible motions.

States of Other Human Parts State transition probabilities may be changed in response to the states of the other parts. For example, if the current state of a left arm transits to "folding", then the state transition probabilities of a right arm are changed for the current state to transit to "folding". In addition, it is possible to make state transition probabilities for transiting to "putting a left hand to the waist" higher when the current state of a right arm is "stick pointing". Changing state transition probabilities in this way may establish coordinated actions of arms and legs.

After transiting to a state and finishing the action of the state, a decision of state transition is given according to state transition probabilities. However, avatar's behavior becomes unnatural if avatar starts the next action as soon as finishing the previous action. Accordingly, the total of state transition probabilities for transiting from a state to all other states is set less than 1 and a decision of state transition is given at regular intervals (usually video-frame intervals) after finishing the previous action until transiting the next state. This means that time between two actions is indefinite and avatar's behavior becomes more natural.

3D Shape-Color Model and Motion Model. In our research, we employ the 3D human model of a computer graphics software, POSER, as a 3D shape-color model of each avatar.

In addition, we construct a motion model for each action in advance. Some of the motion models rotate human parts toward an instructed angle. The others move the tip of human parts such as a hand and a foot toward an instructed position by means of inverse kinematics.

4 Transferred Information

4.1 Information about a Teacher

We think that the following information should be acquired in order that teacher's avatar represents his/her intentions.

Walking and the Position The avatar walks around. Though its standing position in a virtual space reflects the teacher's standing position in a real space, motion parameters of its legs and feet are generated based on the pre-defined walking motion model.

Stick Pointing and the Position The avatar points on a virtual screen where the teacher's pointing stick points in a real space. The correspondence between the virtual space and the real one is also calculated.

Finger Pointing and the Direction The avatar points its finger to the direction that the teacher points to in the real world. If the teacher points to a student in the real world, the avatar of the teacher points the avatar of the student in the virtual lecture space.

Head Direction The avatar moves its head in order that students have the feeling of being watched.

Body Direction The avatar turn its body to the direction.

4.2 Information about a Student

The 3D position and the pose of each student are acquired by a desktop motion capture system using a stereo camera[3]. We think that the following information should be acquired in order to make the avatar's behavior similar to the student;

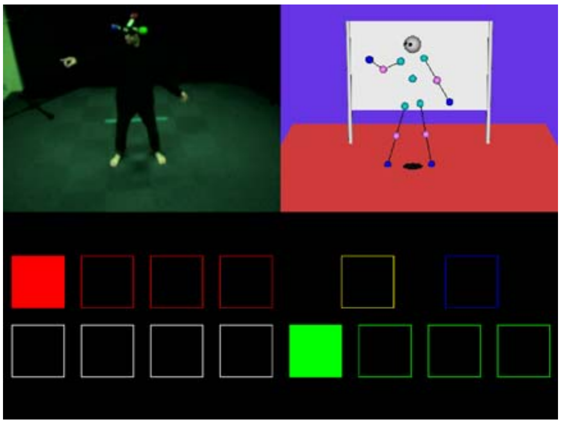
Head Direction The avatar moves its head in order that other participant including a teacher have the feeling of being watched.

Hand Raising In case that the student raises his/her hand, the avatar raises its hand.

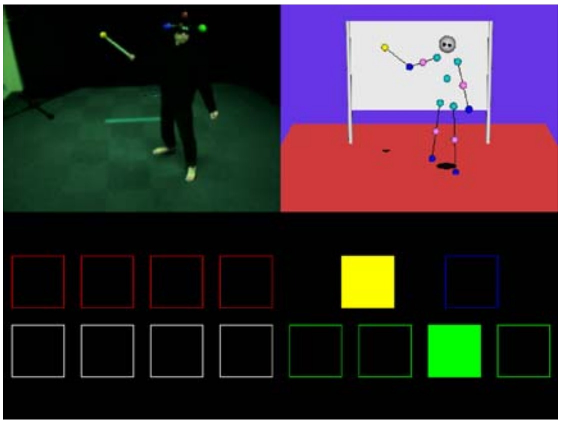
5 Preliminary Experiment

In this paper, we show results of a preliminary experiment, which is the first step for acquiring and presenting information about a teacher. Figure 2 shows experimental results of teacher's information acquisition.

Figure 3 shows that a teacher's avatar is walking and pointing its finger according to acquired information.



(a) Upper left of this figure shows an original image from a camera. Upper right shows results of motion capture system. Each square in the lower part means that the teacher do an action. In this moment, the teacher points his figure to the student who sits at the left end of a virtual lecture space. In addition, the teacher turn his/her face to the student.



(b) In this moment, the teacher points by a stick and turns his/her face to the third student from the left end of a virtual lecture space

Fig. 2. Results of teacher's information acquisition

6 Conclusion

As the first step toward the virtual environment for immersive distance learning, we construct a prototype system which automatically extracts important information from participant actions and reflects it in the behavior of avatars. In this step, we do not regard real-time performance as important but want to show our proposed framework, acquiring minimum information about par-

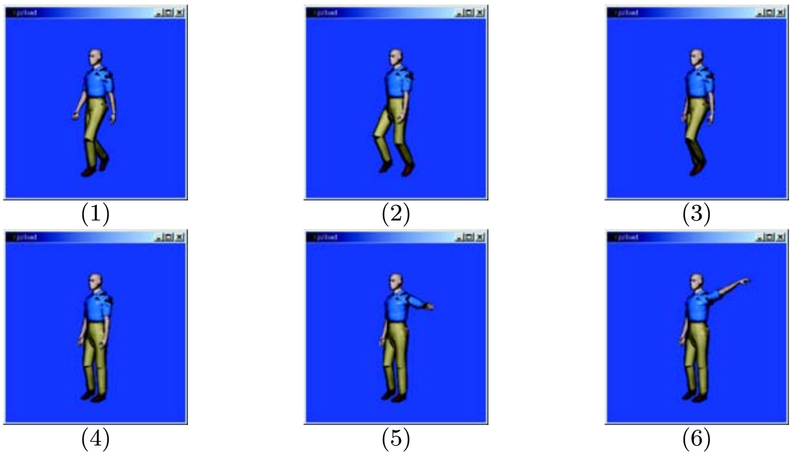


Fig. 3. Displaying teacher's avatar

ticipants and compensating other information for participant avatars to behave naturally, is reasonable and well-performed.

We have some future works as follows:

Time Consistency For constructing VEIDL with high reality and interactivity, VEIDL must guarantee time consistency, which means that all events are presented to participants in the same order they occur. Furthermore, in addition to such event level time consistency, signal level time consistency between audio and visual information, or voice and mouth synchronization, must be guaranteed. For distributed environments such as VEIDL, time consistency is one of the most difficult problems.

Scale of VEIDL We must discuss how many number of students is best for VEIDL. Maybe two factors should be considered. One is limitation of system performance, especially performance of VEIDL server. Another is educational effectiveness. Though one of strong points of VEIDL is interactivity between geographically dispersed participants, the strong point is not effective with too large and too small number of students.

References

1. Kameda, Y., Ishizuka, K., Minoh, M.: A live video imaging method for capturing presentation information in distance learning. In: Proc. of IEEE International Conference on Multimedia and Expo. Volume 3. (2000) 1237–1240
2. Taniguchi, R., Yonemoto, S., Arita, D., Hoshino, R.: Real-time human motion analysis for human-machine interface. In: Proc. of Working Confence on Advanced Visual Interfaces. (2002) 195–202
3. Yonemoto, S., Taniguchi, R.: Vision-based 3d direct manipulation interface for smart interaction. In: Proc. of 16th International Conference on Pattern Recognition. (2002) 655–658